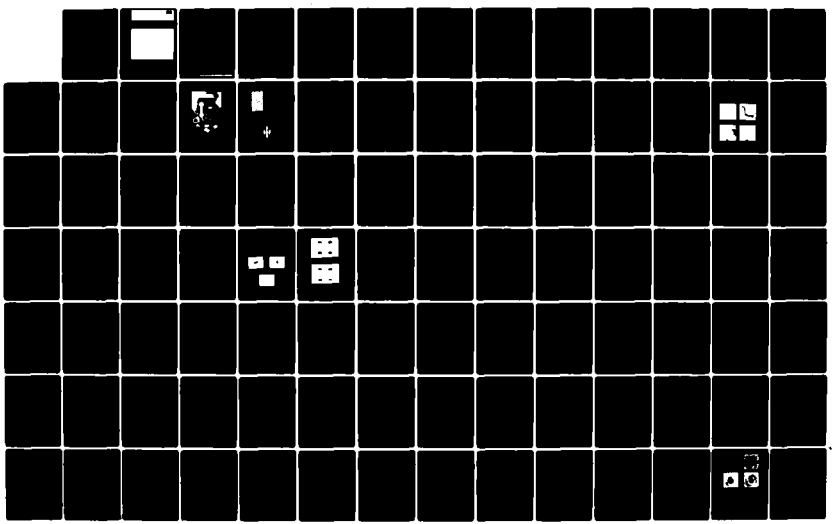
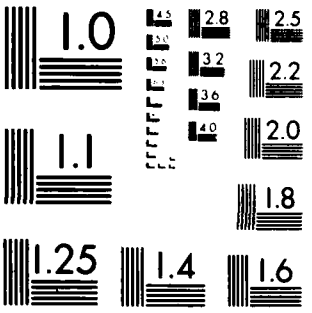
AD-A135 629 TOPICAL MEETING ON SIGNAL RECOVERY AND SYNTHESIS WITH INCOMPLETE INFORMAT..(U) OPTICAL SOCIETY OF AMERICA WASHINGTON D C J W QUINN 31 AUG 83 AFOSR-TR-83-1094 1/3

UNCLASSIFIED AFOSR-83-0026 F/G 20/6 NL

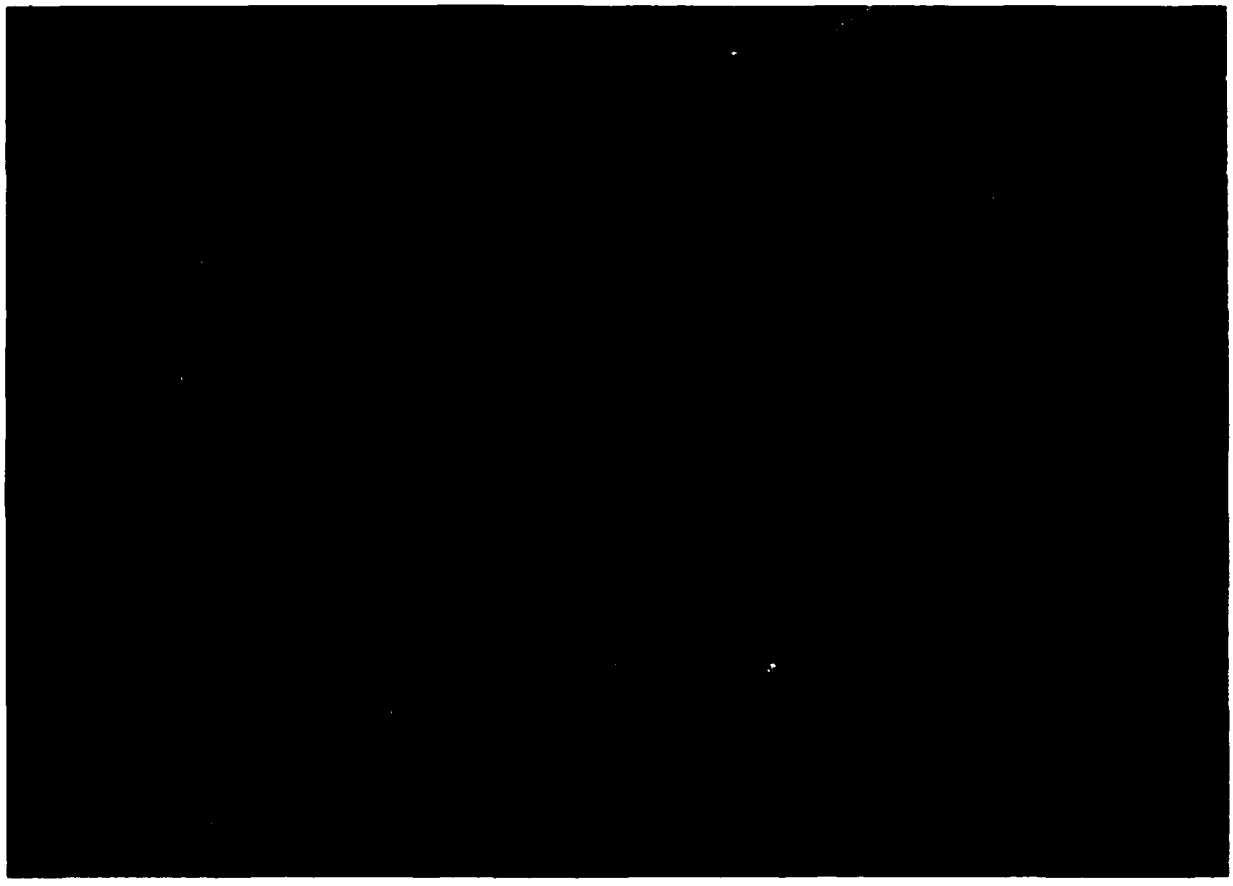




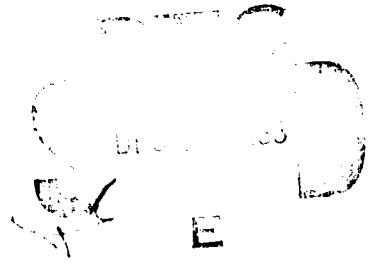
(2)

SIGNAL RECOVERY AND SYNTHESIS WITH INCOMPLETE INFORMATION AND PARTIAL CONSTRAINTS

AFOSR-TR- 83-1094



TECHNICAL DIGEST



WINTER '83
January 12-14, 1983
Incline Village, Nevada

TECHNICAL PROGRAM COMMITTEE

William T. Rhodes, *(General Cochairman)*
Georgia Institute of Technology

Alexander A. Sawchuk, *(General Cochairman)*
University of Southern California

James R. Fienup, *(Technical Program Chairman)*
ERIM

Pierre H. Chavel
Institut d'Optique, France

Hedzer A. Ferwerda
State University of Groningen, The Netherlands

Neal Gallagher
Purdue University

Monson Hayes
Georgia Institute of Technology

Norman Hurt
MRJ, Inc.

Craig Rushforth
University of Utah

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-1094	2. GOVT ACCESSION NO. AD-A135629	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) RECOVERY AND SYNTHESIS WITH INCOMPLETE INFORMATION AND PARTIAL CONSTRAINTS		5. TYPE OF REPORT & PERIOD COVERED FINAL Dec. 1, 1982-Aug.31, 1983 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. Jarus W. Quinn		8. CONTRACT OR GRANT NUMBER(s) AFOSR -83-0026
9. PERFORMING ORGANIZATION NAME AND ADDRESS OPTICAL SOCIETY OF AMERICA 1816 Jefferson Place, N.W. Washington, D.C. 20036		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2305/B1
11. CONTROLLING OFFICE NAME AND ADDRESS UNITED STATES AIR FORCE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH/NE BUILDING 410, BOLLING AFB, WASHINGTON DC 20332		12. REPORT DATE Aug.31, 1983 13. NUMBER OF PAGES five
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Topical Meeting on Signal Recovery and Synthesis with Incomplete Information and Partial Constraints was held in Lake Tahoe, Nevada on January 12-14, 1983. The meeting was open to all interested scientists and engineers and was interdisciplinary in scope. The program consisted of both invited and contributed papers. The two and one-half day meeting dealt with such diverse fields as image processing, crystallography, astronomy, geophysical signal processing, electron microscopy, optical information processing, and remote sensing having encountered the problem of image formation with missing information or unknown

20. (Continued)

(or only partially known) measurement system parameters. Specific examples include phase retrieval from intensity measurements, tomographic reconstruction with missing projections, blind deconvolution, and multidimensional spectral extrapolation and interpolation. A related set of problems occurs in multidimensional signal processing when system constraints are only partially specified. Examples include computer holography and pupil synthesis for incoherent optical processing. A common mathematical structure is shared by all these application areas. The purpose of this meeting was to bring together specialists with interests in these diverse areas and to stimulate interchange of ideas.

TOPICAL MEETING ON SIGNAL RECOVERY AND SYNTHESIS WITH INCOMPLETE INFORMATION AND PARTIAL CONSTRAINTS

A digest of technical papers presented at the Topical Meeting on Signal Recovery and Synthesis with Incomplete Information and Partial Constraints, January 12–14, 1983, Incline Village, Nevada.

Cosponsored by:
Optical Society of America

Air Force Office of Scientific Research

In Cooperation with
Acoustics, Speech and Signal Processing Society of IEEE

WEDNESDAY, JANUARY 12, 1983

RUBICON ROOM

8:30 AM OPENING REMARKS
William T. Rhodes, *Presider*

SESSION 1: EXTRAPOLATION AND RESTORATION I
William T. Rhodes, *Presider*

8:45 AM WA1
Retrieval and Synthesis Problems in Optics—An Overview, H. P. Baltes, *LGZ Landis & Gyr Zug AG, Switzerland.* A synopsis of various deterministic and stochastic inverse optical problems in science and engineering is attempted. Two types of grating profile reconstruction problems are discussed as examples. (Invited Paper)

9:15 AM WA2
Four Models for the Bandlimited Signal Extrapolation Problem, Thomas S. Huang and Jorge L. C. Sanz, *University of Illinois at Urbana-Champaign.* We present four basic models for bandlimited signal extrapolation that are useful in understanding the relationships between the continuous and the discrete extrapolation problems.

9:30 AM WA3
Practical Interpolation of 2-D Surfaces using the Gerchberg Algorithm, Mark J. Carlotto and Victor T. Tom, *The Analytic Sciences Corporation.* The interpolation of 2-D wind and hydrographic surfaces is accomplished using the Gerchberg algorithm. Special emphasis is given to algorithm implementation on an array processor.

9:45 AM WA4
On Some Explicit Deconvolution Formulas, C. A. Berenstein, *University of Maryland,* B. A. Taylor, *University of Michigan, and* A. Yger, *Centre de Mathematiques de l'École Polytechnique, France.* Given several measuring devices (convolution operators of compact support) we give explicit formulas to reconstruct exactly the original signal using again convolutors of compact support.

10:00 AM WA5
Spectral Extrapolation of Constrained Signals, R. J. Mammone, *Rutgers University.* A new method of spectral extrapolation based on a novel approach to linear programming is demonstrated. The method provides a distinct computational advantage over previous LP methods.

WEDNESDAY, JANUARY 12, 1983—Continued

10:15 AM WA6
Iterative Image Restoration From Data Available in Multiple Restricted Regions, Yoshiki Yamakoshi and Takuso Sato, *Tokyo Institute of Technology, Japan.* Convergence properties of iterative image restoration procedure when data are available in restricted multiple regions is examined, and optimum design of observation system is discussed.

LOWER LEVEL LOBBY

10:30 AM COFFEE BREAK

RUBICON ROOM

SESSION 2: EXTRAPOLATION AND RESTORATION II
A. A. Sawchuk, *Presider*

10:50 AM WA7
Matched Image Formation and Restoration System, W. Thomas Cathey, *University of Colorado at Denver,* B. Roy Frieden, *University of Arizona,* William T. Rhodes, *Georgia Institute of Technology,* and Craig K. Rushforth, *University of Utah.* Given that image spectral extrapolation or interpolation is to be done, the image gathering and restoration systems can be matched to improve the results.

11:05 AM WA8
Image Restoration by the Method of Projections onto Convex Sets, H. Stark and I. M. Sezam, *Rensselaer Polytechnic Institute.* The method of projections onto convex sets is used to restore an image from incomplete data and the *a priori* known properties of the image.

11:20 AM WA9
Stable, Noniterative, Object Reconstruction from Incomplete Data using Prior Knowledge, A. M. Darling, T. J. Hall, and M. A. Fiddy, *Queen Elizabeth College, England,* J. Abbiss, *Royal Aircraft Establishment, England.* Object reconstruction in weighted Hilbert space is related to Miller regularization theory. Experimental results illustrating the power of the method are presented.

11:35 AM WA10
Frequency-Domain Optimal Inverse Convolution Filtering of Noisy Data, C. B. Chittineni, *Conoco, Inc.* Frequency-domain optimal inverse filtering algorithms are developed and are applied for seismic deconvolution. A blind restoration technique is formulated, and the simulation results are presented.

11:50 AM WA11
Optimal Iterative Image Reconstruction with Incomplete and Approximate Data, J. Gabmann, *Max-Planck-Institut für Plasmaphysik, West Germany.* Reconstruction is decomposed into a two-fold filtering process. The effect of *à priori* information and constraints on the reconstruction is evaluated with respect to missing, inaccurate data.

12:05 PM LUNCH

PROSPECTOR ROOM

SESSION 3: EXTRAPOLATION AND RESTORATION III
C. Rushforth, *Presider*

2:00 PM WA12
Detection of Line Spectra and Point Sources, A. Papoulis, *Polytechnic Institute of New York.* A modification of Levinson's algorithm leads to a method for detecting line spectra in terms of correlation estimates. The method leads to a simple derivation of Wold's decomposition theorem. *(Invited Paper)*

2:30 PM WA13
Reconstruction of Objects from Coded Images by Simulated Annealing, Warren E. Smith, Harrison H. Barrett, and Richard G. Paxman, *University of Arizona.* An optimization process related to the annealing of a melt to form an ordered crystal is applied to the reconstruction of objects from their coded images.

2:45 PM WA14
Signal Deconvolution using Frequency and Time-Domain Magnitude Constraints, H. J. Trussell and P. N. Sura, *North Carolina State University.* The effect of applying magnitude constraints in the frequency and time domain using interative restoration methods are studied. The characteristics of the signals that are most aided by these constraints are discussed.

3:00 PM WA15
Model-Based Restoration Procedure for Small, Low-Resolution Optical Images, John A. Saghri and Andrew G. Tescher, *The Aerospace Corporation.* A deterministic model-based restoration procedure is presented. The algorithm is effective for the restoration of small, coarsely sampled images degraded by diffraction and noise. The method has been tested with a double delta function model through digital simulations.

3:15 PM WA16
Spectrum Extrapolation on a Finite Band, F. Gori, *Università di Roma, Italy,* S. Wabnitz, *California Institute of Technology.* Starting from the image of a coherent finite object formed through a low-pass pupil we obtain the image pertaining to a larger pupil.

WEDNESDAY, JANUARY 12, 1983—Continued

3:30 PM WA17
On an Iterative Algorithm for Stabilized Object Restoration from Limited Spectral Data, J. B. Abbiss and H. S. Dhadwal, *Royal Aircraft Establishment, England,* C. De Mol, *Université Libre de Bruxelles, Belgium.* We describe a regularized iterative procedure for object restoration from a noisy diffraction-limited image and analyze its performance to obtain explicit error bounds.

RUBICON ROOM

3:45 PM COFFEE BREAK

PROSPECTOR ROOM

SESSION 4: SPECTRAL ESTIMATION AND OTHER APPLICATIONS
C. Rushforth, *Presider*

4:05 PM WA18
Estimation of Two Closely Spaced Frequencies Buried in White Noise using Linear Programming, Jaroslav Keybl and George Eichmann, *The City University of New York.* Linear programming is used to estimate two closely spaced frequencies of sinusoidal signals buried in deep white Gaussian noise.

4:20 PM WA19
A Novel Hankel Approximation Method for ARMA Pole-Zero Estimation from Noisy Covariance Data, S. Y. Kung and K. S. Arun, *University of Southern California.* This paper presents a new covariance approximation method for estimating the parameters of AR, ARMA, and sinusoidal processes by a principal component analysis on the Hankel matrix formed from the covariance data.

4:35 PM WA20
Some Signal Processing Issues in Radar Target Identification, E. K. Miller, *Lawrence Livermore National Laboratory.* Three issues in radar target identification are discussed: (1) Prony-processing for EM features; (2) several EM examples; and (3) some results for target identification using poles.

4:50 PM WA21
Restoration of Multichannel Microwave Imagery to Estimate Rainfall Rates in Hurricanes, R. T. Chin, C. L. Yeh, W. S. Olson, and J. A. Weinman, *University of Wisconsin-Madison.* Eight synthetic multichannel microwave images, each having different spatial resolution, were restored to a common optimal resolution to estimate rainfall. The restoration utilizes a constrained iterative technique.

WEDNESDAY, JANUARY 12, 1983—Continued

5:05 PM WA22
Electromagnetic Image Reconstruction Techniques in Inhomogeneous Media Satisfying the Born-Rytov Approximation, Wolfgang-M. Boerner, *University of Illinois at Chicago.* Projection tomography can be extended to vector diffraction tomography assuming that the media are weakly diffracting so that the Born and Rytov approximations are valid.

RUBICON ROOM

6:00 - 8:00 PM DINNER

8:00-10:00 PM INFORMAL CONTRIBUTED SLIDE SHOW

This session will include informal presentations of outstanding slides and observations of unusual phenomena, unexplained optical effects, beautiful photographs and other relevant and stimulating topics. All attendees are welcome to contribute. Please sign up at the OSA registration desk.

THURSDAY, JANUARY 13, 1983

PROSPECTOR ROOM

SESSION 5: RECONSTRUCTION FROM INTENSITY I
J. R. Fienup, *Presider*

8:30 AM ThA1
Phase Problem in Object Reconstruction and Interferometry—An Overview, H. A. Ferwerda, *State University at Groningen, The Netherlands.* Methods for phase retrieval from modulus data in one and two dimensions are reviewed in the context of object reconstruction, interferometry (including speckle). *(Invited Paper)*

9:00 AM ThA2
***Phase Retrieval from Intensity Data Degraded by Shot* Noise,** C. H. Slump and H. A. Ferwerda, *State University at Groningen, The Netherlands.* Object reconstruction is discussed by taking shot noise into account (low-intensity imaging). The case of weak objects is treated in detail.

9:15 AM ThA3
Phase Retrieval for Functions with Sufficiently Disconnected Support, T. R. Crimmins and J. R. Fienup, *Environmental Research Institute of Michigan.* It is shown that the phase retrieval problem almost always has a solution unique among functions with disconnected supports satisfying a certain common separation condition.

9:30 AM ThA4
Mathematical Results of the Phase Retrieval Problem for Bandlimited Functions of Several Variables, Wayne M. Lawton, *Jet Propulsion Laboratory.* Techniques from the theory of several complex variables are used to derive several new results that characterize bandlimited functions of several variables from their moduli.

9:45 AM ThA5
Maximum Entropy Image Reconstruction from Phaseless Fourier Data, John Skilling, *Cambridge University, England.* The maximum entropy technique is used to reconstruct two-dimensional positive images of realistic size from incomplete and noisy Fourier amplitudes without any phase information.

10:00 AM ThA6
Speckle Interferometry: One-Dimensional Image Reconstruction from Zeros of Complex Spectrum, Yuri M. Bruck and Leonid G. Sodin, *Academy of Sciences of the Ukrainian SSR, USSR.* A method is put forward for unique reconstruction of one dimensional images, affected by scattering in a turbulent medium, from minima of their average spectra in the complex plane. The effect of additive noise and type of scattering is discussed.

10:15 AM ThA7
Phase-Retrieval Technique in Pupil Synthesis, Piotr Kiedron, *The Ohmart Corporation.* The uniqueness of the diffraction-limited system is discussed. The idea of complete compensation of aberration with nonnegative apodizer is presented. Several selected open problems are formulated.

RUBICON ROOM

10:30 AM COFFEE BREAK

PROSPECTOR ROOM

SESSION 6: RECONSTRUCTION FROM INTENSITY II
H. Ferwerda, *Presider*

10:50 AM ThA8
Phase Retrieval in Astronomy, J. R. Fienup, *Environmental Research Institute of Michigan.* Methods used for phase retrieval (object reconstruction) in optical astronomy are reviewed, with emphasis on iterative reconstruction from Fourier modulus data *(Invited Paper).*

11:20 AM ThA9
Frequency Sampling of the Short-Time Fourier Transform Magnitude, Thomas F. Quatieri, S. Hamid Nawab, and Jae S. Lim, *Massachusetts Institute of Technology.* Under mild restrictions, a sequence $\underline{x}(\underline{n})$ is uniquely specified by its short-time Fourier transform magnitude at one or two frequencies for each $\underline{n}$.

11:35 AM ThA10
Averaging the Fourier Phase Information in a Signal Ensemble without Calculating Phase, Herbert W. Swan and Joseph W. Goodman, *Stanford University.* This paper shows how the average Fourier phase information of an ensemble of complex signals may be recovered without actually calculating unwrapped phase.

11:50 AM ThA11
Sufficient Condition for Phase Retrieval in Two Dimensions, M. A. Fiddy, *Queen Elizabeth College, England,* B. J. Brames and J. C. Dainty, *The University of Rochester.* Eisenstein's criterion for irreducibility is used to modify an object function, thus ensuring uniqueness of phase retrieval in two dimensions.

12:05 PM LUNCH

PROSPECTOR ROOM

SESSION 7: SYNTHESIS AND RECONSTRUCTION
M. H. Hayes, *Presider*

2:00 PM ThA12
Signal Reconstruction From Partial Fourier Domain Information, Alan V. Oppenheim and Jae S. Lim, *Massachusetts Institute of Technology.* A number of results have been developed recently in the Digital Signal Processing Group at M.I.T. on signal reconstruction from Fourier transform phase or magnitude. This talk summarizes this work, including exact reconstruction from Fourier transform amplitude and exact reconstruction from the magnitude of the short-time (sliding) Fourier transform. *(Invited Paper)*

2:30 AM ThA13
Speckle Interferometry Image Reconstruction Techniques Proceeding from the Phase of the Fourier Transform, Yuri M. Bruck and Leonid G. Sodin, *Academy of Sciences of the Ukrainian SSR, USSR.* Several methods of image reconstruction are suggested for speckle interferometry applications, proceeding from noisy data on the spectrum phase, with lacking information on the position and size of the image. The uniqueness and accuracy of the reconstruction techniques are discussed.

2:45 PM ThA14
Dependent and Independent Constraints for a Multiple Objective Iterative Algorithm, Joseph N. Mait and William T. Rhodes, *Georgia Institute of Technology.* Examples of dependent and independent constraints for the synthesis of complementary pupil functions for bipolar incoherent spatial filtering through constrained iterative algorithms are described.

3:00 PM ThA15
Signal Reconstruction From Fourier Transform Amplitude, Patrick L. Van Hove, Jae S. Lim, and Alan V. Oppenheim, *Massachusetts Institute of Technology.* New results on the reconstruction of a one-dimensional or multidimensional sequence from its Fourier transform amplitude (magnitude and one bit of phase information) are summarized in this paper.

RUBICON ROOM

3:15 PM COFFEE BREAK

SESSION 8: PROBLEMS AND METHODS I
N. C. Gallagher, *Presider*

3:35 PM ThA16
Reconstruction in Electron Microscopy, W. O. Saxton, *Cavendish Laboratory, England. (Invited Paper)*

THURSDAY, JANUARY 13, 1983—Continued

4:05 PM ThA17
Analysis of Time-Sequential Sampling with Spatially Hexagonal Lattice, Robert M. Cramblitt and Jan P. Allebach, *University of Delaware.* The susceptibility of 2-D hexagonal sampling patterns to frequency domain aliasing is analyzed under the constraint that the sampling be time sequential.

4:20 PM ThA18
Image Registration: The Undersampled Case, P. E. Barry, M. Klop, and J. D. Hulsmann, *Grumman Aerospace Corporation.* Image sequence registration is shown to be possible for the undersampled case even though the determination of image-to-image displacement is not.

4:35 PM ThA19
Partial Shape Recognition using Fourier-Mellin Transform Methods, Timothy A. Grogan and O. Robert Mitchell, *Purdue University.* A partial shape recognition method is introduced that uses a global Fourier-Mellin transform for normalization of the curvature function prior to library comparison.

4:50 PM ThA20
Singular Value Analyses of Inversion of Laplace and Optical Imaging Transforms, M. Bertero, *Istituto di Scienze Fisiche dell Universita and Instituto Nazionale di Fisica Nucleare, Italy* and E. R. Pike, *Royal Signals and Radar Establishment, England.* When "object" and "image' domains differ in size the eigenvalue approach to inversion fails and a singular value analysis is required. Superresolution is then achieved.

5:05 PM ThA21
A technique for the calculation of the global extremum of a function of several variables, C. H. Slump and B. J. Hoenders, *State University at Groningen, The Netherlands.* The global extremum of a function is obtained by calculating an integral which equals the exact number of stationary points in the domain of interest.

RUBICON ROOM
5:30 PM REFRESHMENTS

FRIDAY, JANUARY 14, 1983

PROSPECTOR ROOM

SESSION 9: INTERPOLATION AND TOMOGRAPHY I
N. Hurt, *Presider*

9:00 AM FA1
Algorithm for Incomplete Range of Views Reconstruction, Heang K. Tuy, *Hospital of the University of Pennsylvania.* An algorithm for incomplete range of views reconstruction that makes use of *a priori* information is given with some experimental results.

9:15 AM FA2
Geometric Deconvolution of Artifacts in Limited-View Computed Tomography, Rangaraj M. Rangayyan and Richard Gordon, *University of Manitoba.* Limited-view image reconstruction and computed tomography suffer from a systematic geometric distortion. We present here results of Fourier deconvolution techniques to correct the artifacts.

9:30 AM FA3
Localization from Projections ed on Detection and Estimation of Objects, David ssi, *Schlumberger Doll Research,* and Alan S. W v, *Massachusetts Institute of Technology.* An o based probabilistic model is used in reconstruc eld from limited noisy projection measureme performance of maximum likelihood object on estimation is characterized.

9:45 AM FA4
Tomographic Imaging with Limited-View Angle using an Expansion on a Set of Eigenfunctions Adapted to Space-Limited Objects, Line Garnero, *Universite de Paris, France,* and Jean Brunol, *C.G.R., Medecine Nucleaire, France.* In tomographic imaging, reconstruction of an object of known finite spatial extent is performed from an angularly restricted number of projections by a matrix multiplication.

10:00 AM FA5
Linear Estimation with a Size Constraint, M. J. Lahart, *Naval Research Laboratory.* Correlation coefficients between spectral components of a space-limited image are calculated and used to estimate missing components. The technique is used for bandwidth extrapolation and recovery of missing views in computed tomography.

10:15 AM FA6
Bayesian Approach to Limited-Angle CT Reconstruction, Kenneth M. Hanson and George W. Wecksung, *Los Alamos National Laboratory.* Artifact reduction in limited-angle CT reconstruction is demonstrated by use of the Bayesian approach, which estimates an appropriate null-space contribution to the reconstruction.

RUBICON ROOM

10:30 AM COFFEE BREAK

FRIDAY, JANUARY 14, 1983—Continued

PROSPECTOR ROOM

SESSION 10: INTERPOLATION AND TOMOGRAPHY II
J. C. Dainty, *Presider*

10:50 AM FA7
Use of *A Priori* Information in Image Reconstruction from Limited Data, B. P. Medoff, W. R. Brody, and A. Macovski, *Stanford University*. We show how partial knowledge of the underlying density can be used in an iterative image reconstruction algorithm. Examples using real x-ray data are presented.

11:05 AM FA8
Incorporation of Prior Constraints in Tomographic Reconstructions from Coded Images, Richard G. Paxman and Harrison H. Barrett, *University of Arizona*, and Gene R. Gindi, *Yale University*. The iterative Jacobi method is used to reconstruct 2-dimensional functions from 1-dimensional coded image data sets. The folding of prior constraints into the iterations improves the reconstructions.

11:20 AM FA9
Inverse Scattering Reconstructions From Incomplete Fourier Space Data, N. H. Farhat, *University of Pennsylvania*. We show that 3-D tomographic inverse scattering reconstruction of a scattering object is obtainable from data lying on a curved surface, rather than within a volume, of its accessed Fourier space as would ordinarily be required.

11:35 AM FA10
Deblurring and Three-Dimensional Reconstruction from Multiple Linear-Motion Tomograms, Satoshi Kawata, *Osaka University, Japan*, Jack Sklansky, *University of California at Irvine*. A constrained iterative matrix inversion algorithm reconstructs a three-dimensional attenuation function from a few conventional linear x-ray tomograms, leading to deblurred tomograms.

11:50 AM FA11
Two-dimensional Reconstructions from One-dimensional Data by Maximum Entropy, Sibusiso Sibisi, *University of Cambridge, England*. Maximum entropy is used to obtain two-dimensional positive reconstructions from one-dimensional data. A time series is analyzed in both frequency and decay simultaneously.

12:05 PM LUNCH

FRIDAY, JANUARY 14, 1983—Continued

PROSPECTOR ROOM

SESSION 11: PROBLEMS AND METHODS II
J. A. Neff, *Presider*

2:00 PM FA12
Phase Problem of X-Ray Crystallography from the Viewpoint of Signal Recovery, Gerard Bricogne, *Columbia University*. The statistical theory of the X-ray phase problem will be reviewed, and recent work relating it to Jaynes's maximum-entropy formalism will be presented. *(Invited Paper)*

2:30 PM FA13
Image Reconstruction and Partial Deconvolution with Support Constraint: Separation Angle and Least-Squares Interpolation Procedures, A. Lannes, *Laboratoire d'Optique Electronique, France*. The concept of *separation angle* plays an essential part in image reconstruction from projections and in partial deconvolution with support constraint. The corresponding analysis is illustrated geometrically.

2:45 PM FA14
LBI Image Recovery using Sharpness Maximization, D. G. Steer, M. R. Ito, and P. E. Dewdney, *University of British Columbia, Dominion Radio Astrophysical Observatory*. Simulation experiments are described for LBI image recovery by adjusting the antenna phase delays according to the phase closure constraint to give the maximum image sharpness.

3:00 PM FA15
Phase Synchronization of Distorted Imaging Antenna Arrays, Bernard D. Steinberg, *University of Pennsylvania*. Diffraction-limited performance is achieved in a distorted aperture by a wave-front compensator controlled by a radiation field derived from a source of known properties.

3:15 PM SUMMARIZING DISCUSSION
J. R. Fienup, *Presider*

4:00 PM ADJOURNMENT

SESSION I

EXTRAPOLATION AND RESTORATION I

W. T. Rhodes, *Presider*

RETRIEVAL AND SYNTHESIS PROBLEMS IN OPTICS - AN OVERVIEW

H. P. Baltes

LGZ Landis & Gyr Zug AG

CH - 6301 Zug, Switzerland

Motivations for studying inverse optical problems in the Central Laboratory of the Landis & Gyr Corporation are the encoding of information in diffracting or scattering structures tailored to purpose, the reading of such information using optical sources and detectors, and the retrieval of structural information in the presence of random scatterers (such as a rough surface), with emphasis on the automatic optical checking of identity or authenticity. One example of a product based on such procedures is the PHONOCARD currently being manufactured by Sodeco-Saia, a company of the Landis & Gyr group. The PHONOCARD is a pay telephone operated by pre-paid cards (see Fig. 1) with optically stored value units, whose authenticity is checked by an optical reading system.

What distinguishes the "inverse" or "indirect" problem in optical physics from the "direct" or "normal"? The *direct* problem is to predict the emission or propagation of radiation on the basis of a known arrangement of sources or scatterers. Given the source distribution or the incident radiation field and the structure of the scatterer, one seeks to calculate properties of the scattered radiation field, such as the angular distribution of radiated energy or the resulting photon counting statistics. The *inverse* problem is to deduce features of radiation sources or scatterers from detectable properties of the emitted or scattered field. Given, for example, some far-field data (usually up to some error and noise), the aim is to infer properties of the source or the scatterer.

There are many different inverse problems in optical physics, each with its own subtleties. It is perhaps useful to distinguish between the "scientist's inverse problems" - the retrieval from measured data of features characterising particular scatterers or sources - and the "engineer's inverse problems" - to design an object with a structure that yields certain wanted scattering properties. Another useful classification can be based on the statistical character of the radiation field and the information to be retrieved. Deterministic problems involve coherent radiation with a well-defined (complex) amplitude, and scatterers with a well-defined structure; statistical problems deal with partially coherent radiation fields and stochastic scatterers such as rough surfaces or fluctuating media.

In the first part of my presentation I attempt a synopsis of the large variety of inverse problems in optical physics [1,2] with

emphasis on the underlying questions of *uniqueness*, *stability* and *prior knowledge*. In the second part, two specific examples are presented. One is a deterministic problem on the level of *electromagnetic inverse scattering*, namely the design of grating profiles tailored to yield certain desired diffraction efficiencies. The other is a problem of *statistical optics*, namely the question how a phase grating "hidden" behind a diffuser can be retrieved by observing the degree of coherence of the scattered light.

With respect to the first problem, Fig. 2 shows the result of a non-unique profile reconstruction obtained from a reconstruction procedure devised recently [3]. A theoretical result for the second problem is presented in Fig. 3. It shows the far-zone degree of coherence of scattered radiation emanating from a grating hidden behind a diffuser in a situation where the diffuser would prevent the detection of the grating by far-zone intensity measurements.

[1] H. P. Baltes, ed., *Inverse Source Problems in Optics*, Springer-Verlag, New York, 1978

[2] H. P. Baltes, ed., *Inverse Scattering Problems in Optics*, Springer-Verlag, New York, 1980

[3] A. M. J. Huiser, A. Quattropani, H. P. Baltes, Optics Communs. 41 (1982) 149

[4] A. S. Glass, Optica Acta 29 (1982) 575

Fig. 1 PHONOCARD station and cards
(by courtesy of SODECO-SAIA)

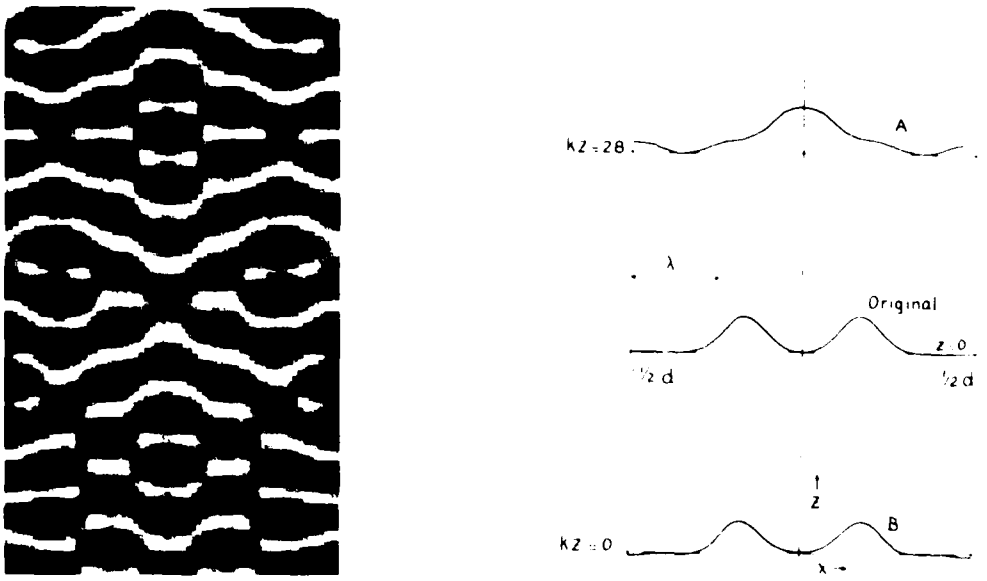


Fig. 2 A non-unique grating profile reconstruction; curves A and B are reconstructions of the original profile shown in the middle (after Ref. 3).

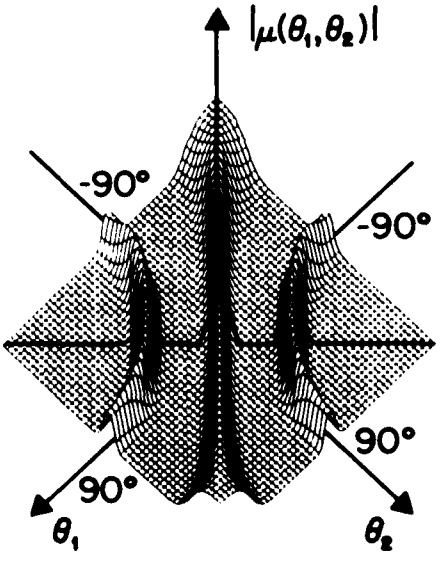


Fig. 3 Degree of coherence of radiation emanating from a lamellar phase grating hidden behind a diffuser (after Ref. 4).

FOUR MODELS FOR THE BAND-LIMITED SIGNAL EXTRAPOLATION PROBLEM

Thomas S. Huang Jorge L. C. Sanz

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
1101 W. Springfield Avenue
Urbana, Illinois 61801

In refs. 2, 4, and 5, two algorithms for solving the continuous band-limited extrapolation problem were developed. However, in practical implementation of these algorithms, discretization is unavoidable. The relationships between the discrete and the continuous algorithms have never been adequately clarified in the literature. In the present paper, we attempt to shed some light on this question.

We first state the problem of continuous band-limited extrapolation. Let f be a finite-energy signal which satisfies $\hat{f}(\omega) = 0$, $\omega \notin [-\Omega,\Omega]$, where $\wedge$ denotes continuous Fourier transform. Since f is analytic, given a piece of f, say $g : [-T,T] \to C : g(t) = f(t)$, $t \in [-T,T]$, we can recover f(t), for $t \notin [-T,T]$. Hence, the continuous extrapolation problem is (under the conditions stated above): Given f(t), $t \in [-T,T]$; find f(t), $t \notin [-T,T]$ (see fig. 1(a)). This model will be referred to as the continuous-continuous model. If we are given a function which is not of finite-energy (for instance, a periodic signal) the extrapolation problem has to be restated. We will concentrate on the periodic case only. Let f be a P-periodic signal defined over the real line: $f(t) = f(t+P)$ for all t, and let f be Ω-band-limited. This means that its distributional Fourier transform $\hat{f}$ is supported in $[-\Omega,\Omega]$ ([6]). In that case it can be proved that f is a linear combination of impulses. The weights in that linear combination are related to the DFT of samples of f. Now, we have another model for the band-limited extrapolation problem, which we call continuous-discrete: given a piece f(t), $t \in [-T,T]$, of a Ω-band-limited P-periodic signal f(t), we want to determine f(t), $t \notin [-T,T]$ (see fig. 1(b)). The solution for this extrapolation model is unique since f is analytic. Moreover, f is a linear combination of exponential functions:

$$f(t) = \sum_{n=-k_o}^{k_o} d_n \, e^{-2\pi i n t/P} \qquad (1)$$

where $k_o = [\frac{\Omega P}{2\pi}]$ (see [7])

The two models given in fig. 1(a) and 1(b) are both concerned with the continuous extrapolation problem; this means that the signals to be extrapolated are continuous in time. Nevertheless, in the first case the Fourier transform is also continuous in time, while in the second case the Fourier transform is a finite array of impulses. There is a major difference in solving the extrapolation problems for the two models: the P-periodic Ω-band-limited function is completely determined by $2k_o+1$ samples in $[-T,T]$. This assertion is a consequence of (1) ([1]). This is not the case for model 1(a).

An attempt to solve the extrapolation problem given in model 1(a) is the following well-known iterative procedure [4],[5]:

$$f_{n+1} = f_n + \mathrm{sinc}_\Omega * (g - J_{[-T,T]} f_n), \quad n \geq 0 \qquad (2)$$

where I is the identity operator and $J_{[-T,T]}$ is the truncation to [-T,T]. One way of getting a numerical implementation of the recursive formula (2) is based on the following discrete recursion:

$$y_{n+1}(j) = y_n(j) + \Delta \sum_{\Delta m \in [-T,T]} \operatorname{sinc}_\Omega(\Delta(j-m))(g(m\Delta) - y_n(m))) \qquad (3)$$
$$y_0(j) = 0$$

where Δ denotes the distance between two consecutive samples and j is any integer number. Now, it is clear that $y_{n+1}(j)$ may not be equal to $f_{n+1}(j\Delta)$. It is well known that $f_n \rightarrow f$ uniformly on the real line. On the other hand, it was shown in [2] that the sequence y_n tends to the minimum norm solution y of the following discrete-continuous extrapolation problem of $g(m\Delta) = z(m)$, $-[T/\Delta] \leq m \leq [T/\Delta]$.

Given z(m), m ε A (A is a finite set of integer numbers) find a finite-energy sequence y(m), m ε Z:

(i) y is band-limited to $(-\Omega,\Omega)$.
(ii) $y(m) = z(m)$, m ε A.

This is the third model sketched in Fig. 1(c). This model appears as a consequence of the numerical implementation of the recursion (2). However, it is also clear that the extrapolation problem which the recursive formula (3) approaches is completely different from the original problem (model 1(a)). In ref. [3] model 1(c) was fully connected with model 1(a).

Another technique of implementing the recursive relationship (2) is by means of the DFT. Equation (2) can be written in the following equivalent form:

$$f_{n+1} = \overset{\vee}{s}_{n+1}$$
$$s_{n+1} = J_{[-\Omega,\Omega]} \cdot (g + (I - J_{[-T,T]} f_n)^{\wedge}$$

where $J_{[-\Omega,\Omega]}$ is the truncation function to $[-\Omega,\Omega]$, $\vee$ and $\wedge$ denote inverse and direct continuous Fourier transform respectively. Replacing $\vee$ and $\wedge$ by their corresponding DFT and implementing the frequency cut-off operator in terms of samples of the frequency space, we get the following recursion

$$\ell_{n+1}(j) = \frac{1}{M} \sum_{m=-N}^{N} \alpha(m) \, e^{2\pi i m j/M} \qquad j \in [-N,N]$$

$$\alpha(m) = \begin{cases} 0 & N \geq |m| > k_o = [\frac{\Omega M \Delta}{2\pi}] \\ \sum_{j=-N}^{N} \beta(j) \, e^{-2\pi i j m/M} & |m| \leq k_o \end{cases} \qquad (4b)$$

$$\beta(j) = \begin{cases} g(j\Delta), & |j| \leq [\frac{T}{\Delta}] \\ \ell_n(j), & N \geq |j| \geq [\frac{T}{\Delta}] \end{cases} \qquad (4c)$$

In formulas (4b) and (4c) Δ denotes the distance between consecutive samples, and $M = 2N + 1$ is the length of the DFT. There is no a priori reason for the convergence of the procedure (4) because $\ell_n(j)$ is not $f_n(j\Omega)$. In ref. [1] it was proven that if $k_o = [T/\Delta] = L$ then ℓ_{n+1} converges to the solution of the following discrete-discrete extrapolation problem

Given the sequence $x(n)$, $-L \leq n \leq L$

Find a sequence $h(n)$, $-N \leq n \leq N$ such that

$$\text{(i)} \quad h(n) \quad \text{is} \quad \text{band-limited} \quad \text{to} \quad [-k_o, k_o] \tag{5.a}$$

$$\text{(ii)} \quad h(n) = x(n), \quad n \in [-L, L] \tag{5.b}$$

The sequence $h(n)$ can also be computed by means of a non-iterative two-step procedure ([9]). Fig. 1(d) sketches this new model for the extrapolation problem. In ref. [1] it was assumed that the number of non-zero frequencies is the same as the number of given samples: $x(n)$, $n \in [-L,L]$. Nevertheless, as i was done for the discrete-continuous model 1(c), it should not be difficult to prove that when $k_o > [T/\Delta]$ the same algorithm (4) converges to the minimum norm sequence $h(n)$ which satisfies (5a) and (5b). The case $k_o < [T/\Delta] = L$ may not yield any solution. It is clear that the discrete-discrete model is completely different from the continuous-continuous model. However, it naturally appears from another numerical implementation of equation (2). Some relationships between models 1(a) and 1(d) are given in [8], but do not get the same class of relationships as the connection between models 1(a) and 1(c) ([3]).

REFERENCES

[1] V. Tom, T. Quatieri, M. Hayes and J. McClellan, "Convergence of Iterative Nonexpansive Signal Reconstruction Algorithms," IEEE Trans. Acoust., Speech, Signal Proc. vol. ASSP-29, no. 5, Oct. 1981.

[2] A. K. Jain and S. Ranganath, "Extrapolation Algorithms for Discrete Signals with Application in Spectral Estimation," IEEE Trans. Acoust., Speech, Signal Proc. vol. ASSP-29, no. 4, August 1981.

[3] J. L. C. Sanz and T. S. Huang, "Discrete and Continuous Band-Limited Signal Extrapolation," submitted to IEEE Trans. Acoust., Speech and Signal Processing, 1982.

[4] A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," IEEE Trans. Circuits Syst. vol. CAS-22, 735-742, Sept. 1975.

[5] R. Gerchberg, "Super-resolution Through Error Energy Reduction," Optica Acta, vol. 21, no. 9, 1974.

[6] Schwartz, "Theorie des Distributions," Dunod, Paris, 1961.

[7] J. L. C. Sanz, "Polar Sampling and Angularly Band-Limited Functions," submitted to IEEE Trans. on ASSP, 1982.

[8] J. L. C. Sanz and T. S. Huang, "Some Aspects of Band-Limited Signal Extrapolation: Models, Discrete Approximations, and Noise," submitted to IEEE Trans. Acoust., Speech, Signal Processing, 1982.

[9] H. Fan, T. S. Huang, J. L. C. Sanz, J. Shafii and B. Tsai "Super-resolution via Band-limited Signal Extrapolation," to be submitted to Applied Optics.

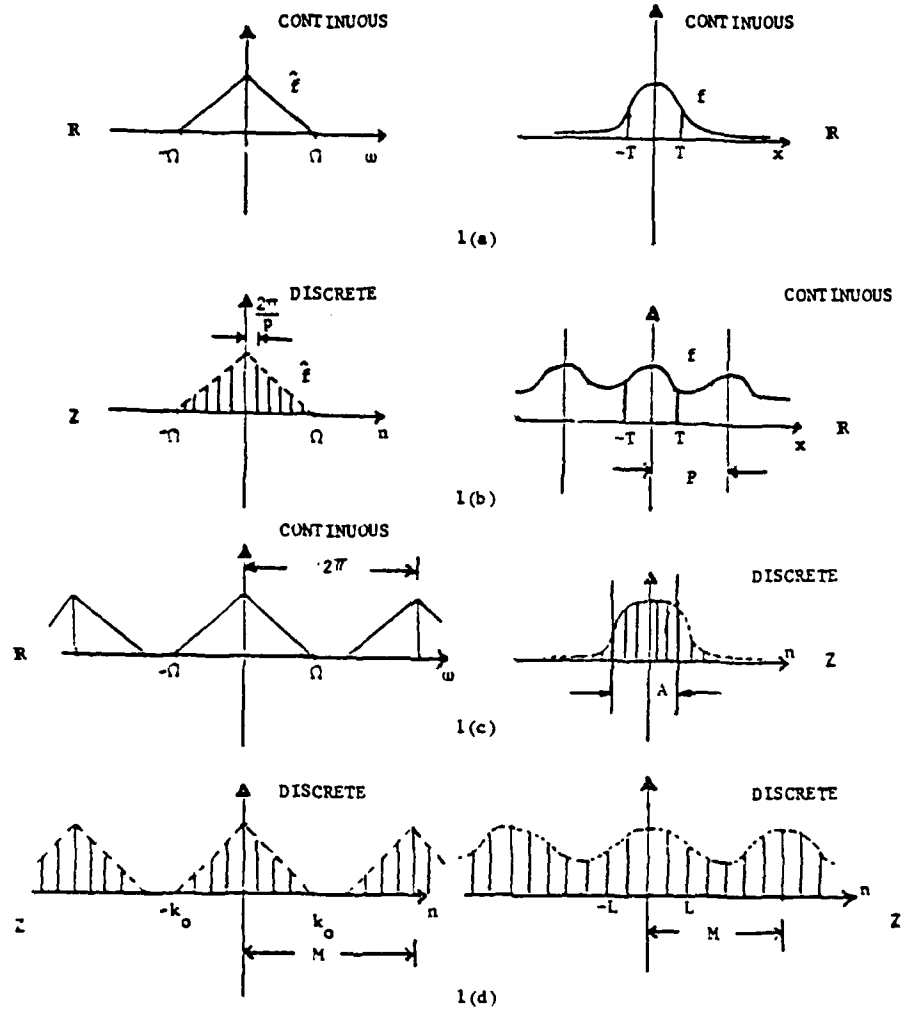


Figura 1 Four basic models for the extrapolation problem.

PRACTICAL INTERPOLATION OF 2-D SURFACES USING THE GERCHBERG ALGORITHM

Mark J. Carlotto and Victor T. Tom
The Analytic Sciences Corporation
One Jacob Way
Reading, MA 01867

Abstract

The interpolation of 2-D wind and hydrographic surfaces is accomplished using the Gerchberg algorithm. Special emphasis is given to algorithm implementation on an array processor.

Introduction

The Gerchberg algorithm has been successfully applied to signal enhancement, reconstruction and extrapolation problems where only partial information is available in the space (time) and frequency domains[1,2]. This algorithm uses an iterative approach and affords effective utilization of partial information in the space and frequency domains. In this paper, a brief description is given of how the Gerchberg algorithm is applied to the iterative interpolation of two-dimensional (2-D) surfaces from irregularly spaced data points. Important implementation issues are discussed which include partitioning large data sets, adapting the spectral bandwidth of the process to the local data, and performing the Gerchberg algorithm on an array processor.

Specific applications presented are: the generation of hydrographic surfaces from bathymetry data obtained from the hydrographic airborne laser sounder (HALS) and the generalization of wind-flows from cloud imagery obtained from the Geostationary Operational Environmental Satellite (GOES). Experimental results obtained using a VAX 11/780 and FPS 120B array processor system are presented.

The organization of this paper is as follows. The next section describes the formulation of the Gerchberg algorithm for interpolation. Issues concerning the implementation of the algorithm on an array processor are then presented in the following section. Two application areas in which this technique has been successfully applied are described in the last two sections.

Interpolation Using the Gerchberg Algorithm

In this section the use of the Gerchberg algorithm to interpolate a 2-D surface to known data subject to a spatial bandwidth (smoothness) constraint is described. Since the desired goal is interpolation rather than extrapolation the solution should be well-behaved. The objective is to interpolate an M x N point surface to a set of data points, d_p, located at (i_p, j_p) whose domain can be written as I_p. By repeatedly substituting the data back into the estimate of the surface and lowpass filtering, the Gerchberg algorithm constructs a solution which is consistent with both the given data and the spatial frequency constraint. At each iteration k, the surface estimate is updated according to

$$s_k = \mathcal{B}s_{k-1} = B\{(I-T)s_{k-1} + Td_p\} \qquad (1)$$

where I is the identity matrix, T is the replacement matrix for the known data, B is the ideal lowpass filtering operation defined by the discrete frequency bandwidths ω_x and ω_y, and

represents the composite operator. Equation 1 has been shown to be stable since Γ is a nonexpansive mapping[3,4]. In addition, if the partial constraints determine a unique solution then Eq. 1 is guaranteed to converge to that solution[5].

Implementation Considerations

This section describes the implementation of an interpolator based on the Gerchberg algorithm on a VAX 11/780 and FPS 120B array processor (AP) system. A comparison of the Gerchberg algorithm with other interpolation techniques (polynomial fitting, least-squares approximation, and local area interpolation, e.g., cubic spline) was performed[4]. The Gerchberg algorithm was found to be computationally more efficient than polynomial techniques in terms of the number of operations performed. The regularity of the computations and the ability to easily change the order of the interpolation, when the known data was spaced irregularly, were significantly better than local area techniques. In addition, the memory requirements of the Gerchberg algorithm were lower than least-square techniques (which require the generation and storage of the orthonormal basis functions). Finally, this technique can be performed in-place reducing storage requirements still further.

The regularity of the alternating spatial replacement and filtering operations allowed an efficient implementation of the Gerchberg algorithm on the AP. A program running in the VAX reads the data files from disk, sections the data set into sub-blocks for processing, and coordinates the operation of the AP. The host supplies a sub-block of data, the number of iterations to be performed, and the filter bandwidths. The AP returns the maximum difference between the interpolated surface and the data:

$$\varepsilon = \text{maximum} \{ s(i,j) - d_p(i,j) \} \quad \text{for } i,j \in I_p \qquad (2)$$

When the difference in Eq. 2 is less than a predetermined threshold ε_T, the interpolated surface is read out of the AP.

In practice, since the original surface is unknown, the bandwidths must be determined indirectly. A technique successfully used to interpolate the HALS and GOES data initializes the filter bandwidths to a low value. After performing a specified number of iterations, if the difference ε is greater than the threshold ε_T, the bandwidths are increased by some predetermined increment. In addition, these differences are an indication of the degree to which the algorithm has converged and can be used to terminate the iteration automatically.

As stated earlier, the M and N array dimensions are determined by the amount of data memory available in the AP. In order not to be limited by the AP memory, a scheme for partitioning large data sets into sub-blocks for processing was developed. Local area techniques have been found to be superior to global interpolation techniques, particularly when the order of the interpolation is large[6]. Since the bandwidths of sub-blocks are adjusted independently, a method of block processing the data which maintains continuity across the data set was implemented. Continuity was maintained by using the edge row and/or column data points from previously processed sub-blocks as true boundary information for the sub-block being processed.

Interpolation of Bathymetry Data

This section describes the interpolation of 2-D surfaces from data obtained from the hydrographic airborne laser sounder (HALS). HALS is a laser bathymetry device for measuring the depth of coastal waters (up to 50 meters in depth depending on water clarity). The data is collected in a spiral sampling pattern (Fig. 1) traced out by a pulsed blue-

green laser beam scanned in an eliptical pattern and translated by the motion of the aircraft. The 256 data points shown in Fig. 1 represents only part of the 4400 depth measurements made in the 1000 X 100 meter survey area processed. The measured depth is proportional to the time-delay between the air-water and bottom return pulses.

The survey area was gridded into a 512 X 64 block of data (spaced 2 meters between samples). The array was partitioned into eight 64 X 64 sub-blocks and was processed as described earlier. In each sub-block, 20 iterations were performed at a given bandwidth. If the peak difference between the data and the interpolated surface was greater than a threshold (ε_T = 3m or 10% of maximum depth), the bandwidth was doubled. The ω_x and ω_y bandwidths were initialized to one frequency bin (0.0039 cycles/meter) at the start of processing each sub-block. The resultant surface is shown in Fig. 2. Since the hydrography was fairly smooth in the outer sub-blocks (0-3 and 6-7), the corresponding bandwidths were small. In the shoal area (sub-blocks 4-5), higher bandwidths were required by the interpolator in order to generate a surface of sufficient order to fit to the measurements. The 512 X 64 sample survey area was processed in approximately 5 minutes under normal system loading.

Generalization of Atmospheric Wind Flows

This section desribes the interpolation of wind vector fields from data obtained from two consecutive frames of GOES cloud imagery (Fig. 3). An initial wind vector field on a square grid was first computed by calculating normalized correlation coefficients ρ for 32x32 sub-blocks and their corresponding search windows. The offsets of the correlation peaks indicate the cloud displacements in x and y, and the maximum correlation values are indicative of the goodness of fit between frames. These initial vectors were then edited to include only those vectors that were due to clouds with 2 kilometer base heights. An additional screening was then performed to retain only those vectors which correspond to high values of ρ, thereby mitigating the effects of changing cloud morphology on this procedure. The remaining vectors (Fig. 3) are then used as the basis for interpolation.

For the present example, the Gerchberg iteration was applied to the x and y components of the wind field separately. For each component, 20 iterations were performed at a given bandwidth, similiar to the bathymetric processing. The interpolated wind field is shown in Fig. 4. This particular example did not utilize the block processing capability, although with much larger data sets, it would have been necessary.

Summary

The Gerchberg algorithm has been shown to be an effective interpolation tool for generating smoothly varying surfaces from incomplete data. A description of the algorithm was presented as well as a discussion of some key issues regarding array processor implementation. Finally the effectiveness of the procedure was demonstrated by interpolating examples of bathymetric data and wind field data.

References

[1] Gerchberg, R.W., "Super-Resolution Through Error Energy Reduction", *Optica Acta*, Vol. 21, No. 9, pp 709-720, 1974.

[2] Papoulis, A., "A New Algorithm in Spectral Analysis and Bandlimited Signal Extrapolation", *IEEE Trans. Circuits Syst.*, Vol. CAS-22, pp. 735-742, Sept. 1975.

[3] Schafer, R.W., Mersereau, R.M., Richards, M.A., "Constrained Iterative Restoration Algorithms", *Proceedings of the IEEE*, Vol. 69, No. 4, 1981.

[4] Carlotto, M.J., *An Iterative Technique for HALS Post-Processing*, TIM-3244-4, The Analytic Sciences Corporation, Aug. 1982.

[5] Tom, V.T., Quatieri, T.F., Hayes, M.H., McClellan, J.H., "Convergence of Iterative Nonexpansive Signal Reconstruction Algorithms", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-29, pp. 1052-1058, 1981.

[6] Conte, S.D., deBoor, C., *Elementary Numerical Analysis: An Algorithmic Approach*, McGraw Hill Book Company, p. 231, 1972.

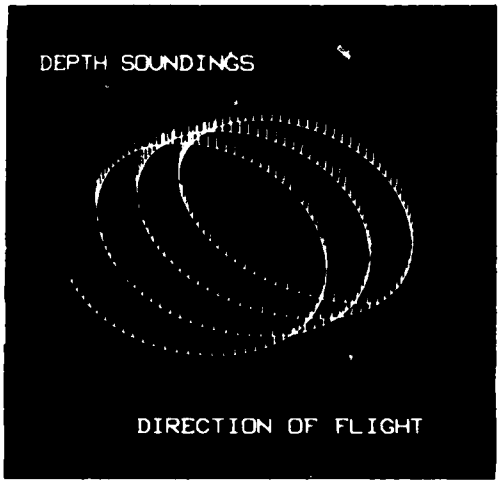


Figure 1 Spiral Pattern of HALS Data

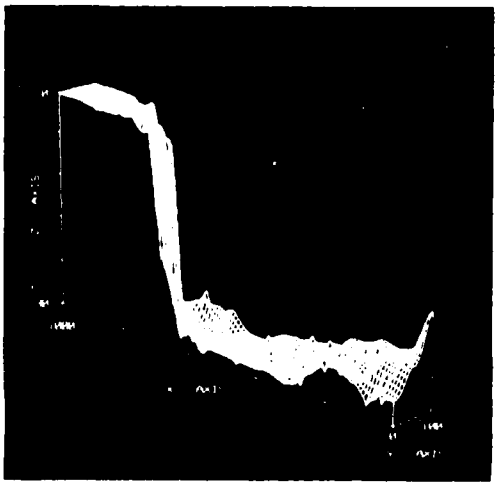


Figure 2 Estimated Surface (σ_1 = 3.0m)

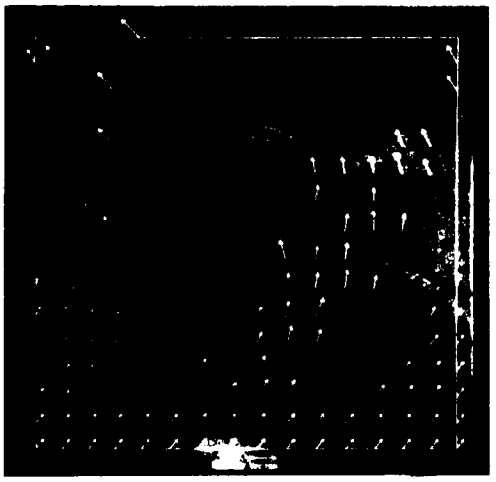


Figure 3 GOES Image and Edited Vectors

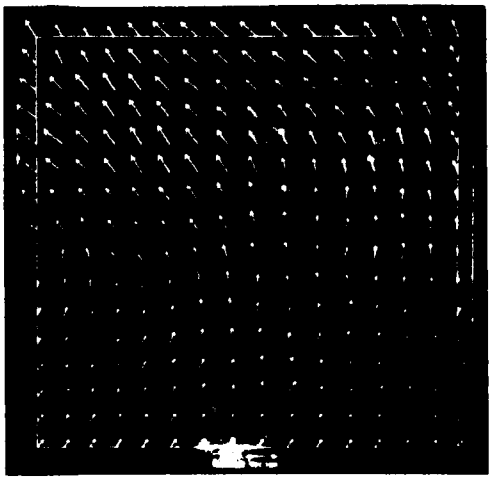


Figure 4 Interpolated Vector Field

On Some Explicit Deconvolution formulas

C. A. Berenstein, Department of Mathematics, University of Maryland, College Park, MD 20742, USA.
B. A. Taylor, Department of Mathematics, University of Michigan, Ann Arbor, Mi 48104, USA.
A. Yger, Centre de Mathématiques de l'École Polytechnique, 91128 Palaiseau Cedex, France.

Given several measuring devices defined by convolution with distributions $\mu_1,\dots,\mu_m$ of compact support in $\mathbb{R}^n$ one would like to construct explicitly *deconvolutors*, i.e. distributions $\nu_1,\dots,\nu_m$, also of compact support, such that

$$(1) \qquad \nu_1 * \mu_1 + \cdots + \nu_m * \mu_m = \delta.$$

This would allow us to reconstruct exactly an arbitrary signal $\phi \in C^\infty(\mathbb{R}^n)$ which was measured as $g_1 = \mu_1 * \phi, \dots, g_m = \mu_m * \phi$ by

$$(2) \qquad \nu_1 * g_1 + \cdots + \nu_m * g_m = \phi.$$

We note that the two main requirements are (i) that the ν_j be of compact support and (ii) that they are *explicit*. By *explicit* we mean that $\nu_1,\dots,\nu_m$ be given by formulas involving the $\mu_1,\dots,\mu_m$, convolution, differentiation, integration and sums. Furthermore, one requires that this procedure be stable in the sense that small errors committed in the implementation of these formulas or even in the construction of the original devices $\mu_1,\dots,\mu_m$ do not affect too much the outcome ϕ given by (2). We remark that the fact that $\nu_1,\dots,\nu_m$ are of compact support should also help with the analysis of noise accompanying the signal ϕ, we plan to come back to this point in another paper.

To put the problem into another, maybe more familiar, perspective we apply the Fourier transform to equation (1) and we get the identity

$$(3) \qquad \hat{\nu}_1(z)\hat{\mu}_1(z) + \cdots + \hat{\nu}_m(z)\hat{\mu}_m(z) = 1 \qquad \text{for all } z \in \mathbb{C}^n,$$

hence we are looking for solutions $\hat{\nu}_1,\dots,\hat{\nu}_m$ of what is sometimes called the *Bezout problem*, these functions must be entire holomorphic functions of n complex variables and satisfy the Paley-Wiener type estimates

$$(4) \qquad |\hat{\nu}_j(z)| \leq A(1+|z|)^N e^{B|\operatorname{Im} z|}.$$

It is immediate from (3) and (4) that to expect a solution to our problem one needs that for some positive constants ε, B, N one has:

$$(5)\qquad \sum_{j=1}^{m} |\hat{\mu}_j(z)| \geq \varepsilon e^{-B|\operatorname{Im} z|}(1+|z|)^{-N}, \qquad \forall z \in \mathbb{C}^n.$$

It is known that this condition is also sufficient for the solvability of the Bezout problem [1]. The difficulty hinges in two things, first, this is purely an existence theorem and, second, even if one could write down the functions $\hat{\nu}_j$ this might not satisfy the stringent conditions required above unless their Fourier inversion were practically immediate. There is one known case of this situation, when all the μ_j have punctual support, then we might as well assume that the $\hat{\mu}_j$ are polynomials in n variables and we look for $\hat{\nu}_j$ also polynomials. In one variable, these $\hat{\nu}_j$ are found via the Euclidean algorithm, in several variables using elimination theory or using an integral expression found by Anderson-Berndtsson [2]. We are interested here in the case where the μ_j no longer have punctual support, and indeed, under some technical conditions on the μ_j, in addition to (5), we can explicitly construct distributions ν_j satisfying (1). The details and its proof appear in [3], [4]. We will restrict ourselves here to discuss one example of particular interest in optics.

Let μ_1, μ_2 be the normalized characteristic functions of planar disks of center the origin and radii r_1 and r_2 respectively. As convolution operators they correspond to difraction by circular lenses. Consider the map

$$(6)\qquad \phi \mapsto (\mu_1 * \phi,\ \mu_2 * \phi),$$

it can be proved that the necessary and sufficient condition for the injectivity of this map is that $r_1/r_2 \notin E = \{\xi/\eta : J_1(\xi) = J_1(\eta) = 0 \text{ and } \xi,\eta \neq 0\}$, where J_1 is the Bessel function of first kind and order 1 (see [5] for references and discussion of this problem.) Noting that

$$(7)\qquad \hat{\mu}_1(z) = \frac{1}{\pi r_1}\frac{J_1(r_1\sqrt{z_1^2+z_2^2})}{\sqrt{z_1^2+z_2^2}}, \qquad \hat{\mu}_2(z) = \frac{1}{\pi r_2}\frac{J_1(r_2\sqrt{z_1^2+z_2^2})}{\sqrt{z_1^2+z_2^2}}$$

we see that the condition that r_1/r_2 is not in the exceptional set E corresponds exactly to the non-existence of common zeroes of the Fourier transforms of μ_1 and μ_2. Here we want to invert the map (6), hence we need the stronger condition (5). In fact, (5) is equivalent in this case to the following arithmetical condition

$$(8)\qquad \exists\, c > 0,\ N > 0 \text{ such that } |(r_1/r_2) - (\xi/\eta)| \geq \frac{c}{|\eta|^N} \text{ for all } \xi/\eta \in E.$$

As a corollary to our theorem one can show that (8) suffices to construct explicitly deconvolutors ν_1, ν_2. The formula we give below corresponds to the optimal case of $N = 2$. Though it is possible that if r_1/r_2 is a rational

number (obviously $\neq 1$) or even a quadratic irrational then (8) holds with $N = 2$ we only know a slightly weaker result, i.e. that there is an algorithmic way of deciding whether or not (8) holds. In less than 2 seconds a computer has shown that all integers in the range from 2 to 200 satisfy (8). By $I_{R,\gamma}$ we denote the distribution supported by the disk $\{x^2 + y^2 \leq R^2\}$ and represented in the interior of that disk by the function

$$(9) \qquad I_{R,\gamma}(x,y) = \frac{2}{\pi^2 R^2} \int_{\sqrt{x^2+y^2}}^{R} \left(\int_t^R \sqrt{R^2 - s^2} \cos \gamma(t-s)ds \right) \frac{dt}{\sqrt{t^2-x^2-y^2}} .$$

When β is a positive zero of the function $z \mapsto J_1(r_1 z)$ we denote

$$(10) \qquad u_1(\beta) = (\beta^5 J_1(r_2\beta)J_2(r_1\beta))^{-1}, \qquad \sigma_1 = \sum u_1(\beta),$$

where the sum takes place over all the positive zeroes of $J_1(r_1 z)$. Similarly if α represents a positive zero of $J_1(r_2 z)$ we define $u_2(\alpha), \sigma_2$. With Δ denoting the Laplace operator, we are ready now to define the two distributions ν_1, ν_2

$$(11) \qquad \nu_1 = -2\pi^2 r_1 \sigma_2 \Delta^3 \mu_2 - 2\pi^2 r_1 \Delta^4 \left(\sum_\alpha u_2(\alpha) I_{r_2,\alpha} \right), \quad \text{and}$$

$$(12) \qquad \nu_2 = -2\pi^2 r_2 \sigma_1 \Delta^3 \mu_1 - 2\pi^2 r_2 \Delta^4 \left(\sum_\beta u_1(\beta) I_{r_1,\beta} \right) + 4\pi^2 \mu_1 - \frac{\pi^2}{2} (r_1^2 + r_2^2)\Delta\mu_1$$

$$+ \frac{\pi^2}{48} (2r_1^4 + 3r_1^2 r_2^2 + 2r_2^2)\Delta^2 \mu_1 .$$

They satisfy the desired identity

$$\mu_1 * \nu_1 + \mu_2 * \nu_2 = \delta .$$

One can see from (11) and (12) that cutting-off the series gives an error that can be estimated in terms of a priori bounds on the derivatives of order at most six of the signal ϕ and, moreover, the finite sum remaining after the cut-off is fairly insensity to the arithmetical nature of r_1/r_2 and hence one uses (8) for theoretical purposes only, i.e. one is allowed errors even in this quotient or, in other words, in the construction of μ_1 and μ_2.

Actually, the above example is really a one dimensional problem due to the radial symmetry of μ_1 and μ_2, but the same kind of analysis can be carried over to genuinely n-dimensional situations, for instance one can study the case where μ_1, μ_2, μ_3 are the characteristic functions of a square and two of its rotations, if they are chosen correctly to satisfy (5) (a rotation of 36° and one of 45° would do the trick) [4].

References

[1] J. J. Kelleher and B. A. Taylor, Finitely generated ideals in rings of analytic function, Math. Ann. 193 (1971), 225-237.
[2] B. Berndtsson, A formula for interpolation and division in $\mathbb{C}^n$, preprint Universidad Nacional de Mexico, 1982.
[3] C. A. Berenstein, B. A. Taylor and A. Yger, Sur quelques formules explicites de decónvolution, preprint Centre de Mathématiques de l'École Polytechnique, 1982.
[4] C. A. Berenstein and A. Yger, Le problème de la decónvolution, preprint Centre de Mathématiques de l'École Polytechnique, 1982.
[5] L. Zalcman, Offbeat integral geometry, Amer. Maths. Monthly 87 (1980), 161-175.

The first two authors would like to acknowledge the support they received from the National Science Foundation in the preparation of this paper.

WA5-1

Spectral Extrapolation of Constrained Signals

R.J. Mammone
Rutgers University
Dept. of Electrical Engineering
New Brunswick, N.J. 08903

There are many applications where it is desirable to restore the high frequency components of a signal. For example, in radar and sonar, greater range accuracy can be obtained by spectral extrapolation. In imaging systems, the restoration of an image will provide increased resolution. The ability to restore the fine details of an image is of obvious value in such diverse fields as astronomy, electron microscopy, satelite surveillance, medical image processing, remote sensing, and in image data compression applications.

The theory of analytic continuation states that a signal of finite duration can be uniquely reconstructed from its bandlimited image[1]. The spectral extrapolation of a signal bandlimited to a cutoff frequency w requires the inversion of the convolution integral operator

$$g(t) = \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{\sin W(t-x)}{\pi(t-x)} f(x)\, dx \qquad (1)$$

where g(t) is the bandlimited signal and f(x) is the original signal of duration T. The discretized form of the problem yields the matrix equation

$$g = Hf + n \qquad (2)$$

where g and f are the sampled vectors of the measured bandlimited signal and the signal to be estimated, respectively, n is the sampled noise vector, and H is a matrix representation of the integral operator obtained by applying an appropriate numerical integration technique. For the spectral extrapolation problem the H matrix is ill conditioned. That is the rank of H is less than the dimension of H. Thus H does not map f uniquely into g and H^{-1} does not exist. The rows of H are not linearly independent. Thus inconsistencies in the equations also occur i.e. the same linear combination of unknowns may be set equal to two different measured values. Since f is not uniquely determined for a given vector g, some additional criteria must be used to select the optimal solution. The addition of constraints on f will aid this selection process.

Recently, a method of constrained inversion has been demonstrated[2,3]. This method makes it possible to impose various constraints on the signal to be recovered. The signal can be restricted to be nonnegative, bounded, smooth i.e. bounded derivatives or in general, any equality or inequality of a linear combination of signal samples can be

implemented. The constraint equations form a region of feasible solutions. The optimal solution should be selected for the signal estimate. This problem can be formulated as a linear programming (LP) problem. The most common method of solving LP problems is called the simplex method[4].

In order to specify the spectral extrapolation problem as an LP problem equation (2) is rewritten

$$g = Hf + n^+ - n^- \tag{3}$$

where the noise vector n has been decomposed into two nonnegative components. The LP formulation can then be written as follows:

$$\text{Minimize: } n_1^+ + n_1^- + n_2^+ + n_2^- + \ldots + n_N^+ + n_N^- \tag{4}$$

$$\text{Subject to: } g = Hf - n^+ - n^-$$

$$0 \leq f \leq f_{max}$$

The function to be minimized (the objective function) is equal to the sum of the absolute value of the elements of the error vector n. This scalar quantity is the l_1 norm of the error. Thus the restored vector f obtained in this way will be optimal in that the l_1 norm of the error is minimized. Minimization of this norm is considered Robust in that it is not sensitive to bursty errors, i.e. a relatively small number of large errors. Minimization of the l_2 (summation of squares) and l_∞ (maximum value) norms of the error are much more sensitive to bursty types of errors. Optimizing these norms will produce estimates which tend to follow the few highly erroneous points. The minimal l_1 formulation thus provides a stabalizing effect. Previously, it was found that the l_2 and l_∞ norms are not as effective as the l_1 norm in solving ill conditioned problems[3]. The l_2 and l_∞ norms are also computationally more expensive. Thus we shall only consider the l_1 norm in this paper.

The simplex method of LP will be used to obtain the minimal l_1 norm of f. The basic premise of the simplex method is that the solution will contain at most r nonzero elements, where r is the rank of the constraint matrix H. The number of nonzero elements of the recovered signal and error vectors can at most be r. The rank of H is approximately $2WT$[3]. This is analagous to the dimension of a communication system[5]. The simplex formulation given by equation (3) can be interpreted as a method of finding the linear combination of column vectors of H which best approximate the measured signal vector g.

LP problems are generally solved in two phases. First, a feasible solution is found. Second, the optimal feasible solution is obtained. Each phase requires approximately the same computational effort. The procedure in the second phase is to iteratively replace one element in the feasible set with one in such a way as to decrease the objective function. The optimal

solution is obtained when no further decrease is possible.

In this paper, it is shown that the first phase of LP can be eliminated for the minimal l_1 formulation. This is accomplished by rewritting formulation (3) as

$$\text{Minimize: } g_{tot} - hf + 2n_1^- + 2n_2^- + \ldots + 2n_N^- \tag{5}$$

$$\text{subject to: } n^+ = g - Hf + n^- \quad \text{and } f, n^+, n^- \geq 0 \tag{6}$$

Where h is a row vector in which the i^{th} element is the summation of the elements in the i^{th} column of H.

g_{tot} is the summation of all the measured elements g.

It has also been assumed for the sake of clarity that all the elements of g are positive. In general if a negative g_i element was obtained then the negative superscripted error element n_i^- would be used in equation (5).

Thus the desired starting point equates the measured signal to the noise vector. Since by assumption g_{tot} is positive it is desired to introduce an element of f into the solution to further reduce the error given by expression (5). The choice as to which f element should replace which n^- element is called the pivot strategy.

The standard simplex algorithm selects the f element which has the largest coefficient in (5). This corresponds to the greatest rate of optimization of expression (5). The n^- element is selected as the first element to become non-positive as the selected f element increases. In this paper we modify the pivot strategy so that the product of the coefficient of f in expression (5) and the value of f is a maximum. This strategy is of particular importance in the case where there is more than one maximal coefficient in expression (5). The memory requirements of the minimal l_1 formulation can also be conserved by storing the H vectors as 1D arrays as opposed to the 2D representations necessary for direct solution by the simplex algorithm.

Conclusion:

A new method of constrained spectral extrapolation is presented. Computer simulations indicate the restorations of high detail in image data.

References:

1. D. Slepian and H.O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty - I," Bell Syst. Tech. 40, 43-63 (1961)
2. R. Mammone and G. Eichmann, "Superresolving image restoration using linear programming," Appl. Opt. 21, 496-501 (1982)
3. R. Mammone and G. Eichmann, "Restoration of discrete Fourier spectra using linear programming," J. Opt. Soc. Am. 72, 987-992 (1982)

4. Dantzig, G.B., Lineare Programmierung und Erweiterungen, (Springer, Berlin, 1966)
5. J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering,(John Wiley and Sons, New York, 1965)

Iterative Image Restoration From Data Available In Multiple Restricted Regions

Yoshiki Yamakoshi and Takuso Sato

Tokyo Institute of Technology, The Graduate School at Nagatsuta, 4259 Nagatsuta, Midori-ku, Yokohama-shi, 227 Japan

I. Introduction

Restoration of the original image from data available in restricted regions has been the subject of many studies in the field of optical, X-ray and acoustical imagings. Several concrete means for extrapolating of the observed data beyond the observed region have been proposed.[1-5] Among them, iterative means for the image restoration is considered as one of the most effective methods, because of its stability even under faily noisy observed data.

In this paper, the iterative image restoration procedure for the case where the data are available in multiple restricted regions is treated generally and the convergence properties of this procedure and an optimum construction of the observation system are examined. Multiple restricted regions may be considered, for instance, when the obstacles such as the object holders exist or when the data with high signal to noise ratio can be obtained only in restricted regions due to the existence of highly absorbing parts in the object.

II. Formulation of the iterative image restoration procedure

The observation system under consideration is shown schematically in Fig.1. Only one dimensional cases are considered. In this system, it is assumed that the data are available in n restricted regions in the image plane. Hence, the observed image g(x) is given as follows

$$g(x) = \sum_{k=1}^{n} \{P_{D,k}(x)\, f(x)\} \qquad (1)$$

where

$$P_{D,k}(x) = \begin{cases} 1 & X_k - \frac{\Delta X_k}{2} \le x \le X_k + \frac{\Delta X_k}{2} \\ 0 & \text{otherwise} \end{cases}$$

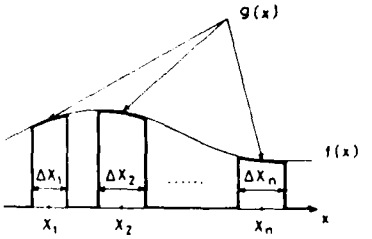


Fig.1 Available data in the image plane

and f(x) is the original image to be restored. As for f(x), we assume that i) it is square integrable in each region and ii) it is band limited within $\pm\Omega$ in the spatial frequency plane.

If we consider a set of prolate spheroidal wave functions[6] (PSWF) $\{\psi_{\ell,j}\}$ which are complete in the ℓ-th restricted region, then the expansion coefficients of the observed image are as follows

$$a_{g,\ell,j} = \int_{-\infty}^{\infty} g(x)\, \psi_{\ell,j}(x)\, dx = \sum_{k=1}^{n} \int_{-\infty}^{\infty} P_{D,k}(x)\, f(x)\, \psi_{\ell,j}(x)\, dx \qquad (2)$$

Expanding f(x) by using $\{\psi_{k,j}\}$ for k-th restricted region, we obtain

$$\mathbf{a}_{g,\ell} = \mathbf{A}\, \mathbf{a}_{f,\ell} \qquad (3)$$

where

$$\tilde{\mathbf{a}}_{g,\ell} = [\, \tilde{a}_{g,\ell,0}\ ,\ \tilde{a}_{g,\ell,1}\ ,\ \cdots\ ]^t$$

$$\mathbf{a}_{f,\ell} = [\, \tilde{a}_{f,\ell,0}\ ,\ \tilde{a}_{f,\ell,1}\ ,\ \cdots\ ]^t$$

$$\mathbf{A} = \sum_{k=1}^{n} \mathbf{\Psi}_{k\to\ell}\, \boldsymbol{\lambda}_k\, \mathbf{\Psi}_{\ell\to k}\ , \qquad (\mathbf{\Psi}_{k\to\ell})_{i,j} = \int_{-\infty}^{\infty} \psi_{k,j}(x)\, \psi_{\ell,i}(x)\, dx \qquad (4)$$

$$\boldsymbol{\lambda}_k = \mathrm{Diag}(\lambda_{k,j})$$

and $\{\lambda_{k,j}\}$ are the eigen values of $\{\psi_{k,j}\}$. Thus, the desired coefficient vector $\mathbf{a}_{f,\ell}$ of the original image is obtained from the observed data by solving the following linear equation.

$$\mathbf{a}_{f,\ell} = \mathbf{A}^{-1}\, \tilde{\mathbf{a}}_{g,\ell} \qquad (5)$$

It can be shown that Eq.(5) is equivalent to the following revisions.

1) Revision in the image plane

$$h_i'(x) = \sum_{k=1}^{n} P_{D,k}(x)\, f(x) + \left(1 - \sum_{k=1}^{n} P_{D,k}(x)\right) h_i(x) \qquad (6)$$

2) Revision in the spatial frequency plane

$$H_{i+1}(\omega) = P_B(\omega)\, H_i'(\omega) \qquad (7)$$

where

$$H_i'(\omega) = \mathfrak{F}\{h_i'(x)\}, \quad h_{i+1}(x) = \mathfrak{F}^{-1}\{H_{i+1}(\omega)\}$$

$$P_B(\omega) = \begin{cases} 1 & |\omega| \le \Omega \\ 0 & \text{otherwise} \end{cases}$$

Fig.2 shows these procedures.

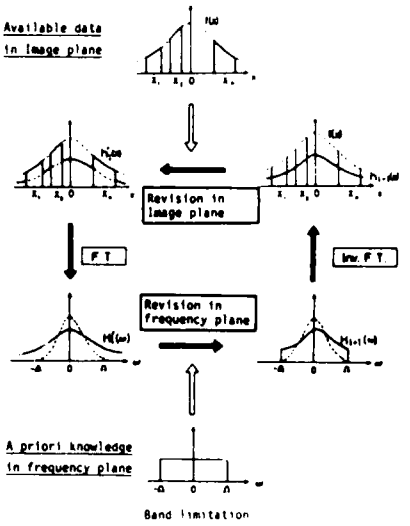


Fig. 2
Schematic diagram of the iterative image restoration procedure

III. Convergence properties of the iterative revision procedure

For the estimation of the mean square error of the iteratively revised image, we use eigen functions $\{\phi_j(x)\}$ which have the expansion coefficient vectors $\{\mathbf{x}_j\}$ of the matrix $\mathbf{A}$ and the eigen values $\{\sigma_j\}$.

Let us consider the mean square error of the i-th revised image. We assume that the observed image g(x) is written as

$$g(x) = \sum_{k=1}^{n} P_{D,k}(x)\, f(x) + \sum_{k=1}^{n} P_{D,k}(x)\, n(x) \qquad (8)$$

where n(x) is the additive noise. This model represents the imaging system such as X-ray radiorogy, since the fluctuation of the photon counting may be interpreted as the noise. If we assume that f(x) and n(x) have the band limited white spectrum within $\pm\Omega$,

$$E\{f(x)f(x')\} = \tilde{a}_f^{\,2}\, \sin\Omega(x-x')/\pi(x-x') \qquad (9)$$

$$E\{n(x)n(x')\} = \tilde{a}_n^{\,2}\, \sin\Omega(x-x')/\pi(x-x') \qquad (10)$$

then the expectations of the mean square error $E\{\varepsilon_i\}$ is given by

$$E\{\varepsilon_i\} = E\{\int_{-\infty}^{\infty}(f(x)-h_i(x))^2dx\} = \sum_j (1-\sigma_j)^{2i}\, a_f^{\,2} + \sum_j \{1-(1-\sigma_j)^i\}^2\, a_n^{\,2} \quad (11)$$

Here, we will examine Eq.(11) for the case where the sum of the widths of the restricted regions is fixed. We only discuss the case where the sum of the widths of these regions is small so that $\sum_{k=1}^{n} \Delta X_k \ll \pi/\Omega$ is satisfied. Examination for this case seems of great importance in practical cases, since we are considering the problem of the restoration of the original image from incomplete data and this condition corresponds to the extreme case of this kind. In this case, the convergence of the iterative procedure must be slow because the available information about the object is restricted seriously.

When the iterative procedure is truncated at a small number, the mean square error is approximated as

$$E\{\varepsilon_i\} = \sum_j a_f^{\,2} - 2i\, a_f^{\,2}\,\alpha + \{i(2i-1)\, a_f^{\,2} + i^2\, a_n^{\,2}\}\,\beta \quad (12)$$

where α and β are the sum of the eigen values and the sum of the squares of the eigen values respectively.

When the sum of the widths of the regions is fixed, α is constant, since $\alpha = \sum_{k=1}^{n} (\Omega/\pi)\,\Delta X_k$. Hence, the observation system with smaller β gives the smaller mean square error for the revised image.

IV. Construction of the optimum observation system

Next, we consider the construction of the observation system which gives small β, that is, the system which gives small mean square error. The β is written as follows

$$\beta = \sum_{k=1}^{n}\sum_j \lambda_{k,j}^{\,2} + \sum_{k=1}^{n}\int_{-\infty}^{\infty} P_{D,k}(x)\,dx \sum_{\substack{k'=1 \\ (k'\neq k)}}^{n}\int_{-\infty}^{\infty} P_{D,k'}(\eta)\{\sin\Omega(x-\eta)/\pi(x-\eta)\}^2 d\eta \quad (13)$$

The first term of Eq.(13) is the sum of the squares of the eigen values of the PSWF, which depends only on the width of each region. On the other hand, second term presents the cross interaction among these regions which depends on the widths of the regions and the distances among them.

When the movement of the regions is allowed, β depends only on the second term of Eq.(13). Hence, $E\{\varepsilon_i\}$ decreases with the decrease of this cross interacting term and we can see that this term decreases when the distances between the neighboring regions are increased.

The evaluation of the system based on β may also interpreted from the following diffrent point of view.

The sum of the squares of the eigen values β corresponds to the variance of the distribution of the eigen values, since the sum of the eigen values α is constant. Hence, if β is large, the distribution of the eigen values should be something like a step function. Thus, eigen values should take values close to one or close to zero. On the other hand, if β is small, the eigen values should be distributed between one and zero. These situations are shown in Fig.3. Thus, the observation system with a small β may be considered to give information about larger number of the eigen functions compared to the system with large β. The information for

each eigen function, however, may be degraded in this case. Hence, if we can use the iterative means for the restoration of these degraded information, it may be said that the system with small β is prefered from the point of view of the mean square error of the image after revisions of a finite number.

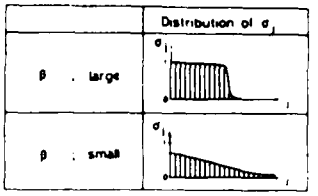


Fig. 3
Distribution of the eigen values

V. Computer simulations

In Fig.4, the properties of β for the case where only two regions exist are shown. From the figure, β decreases with the increase of the distance between two regions, but it remains almost constant when the distance becomes greater than π/Ω. Fig.5 shows the effect of the number of divisions and the variation of the widths of the regions. It shows that β decreases with the decrease of the variation of the widths. It increases with the increase of the number of divisions. Fig.6 shows the comparison between β and the normalized mean square error of the revised images. The original image has the band limited white spectrum. The results show that the optimum construction which gives the minimum mean square error for the first time with the increase of the distance corresponds almost exactly to that for the β.

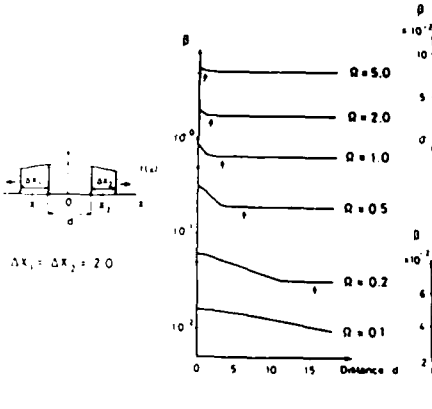


Fig.4
Effects of the distance between the regions

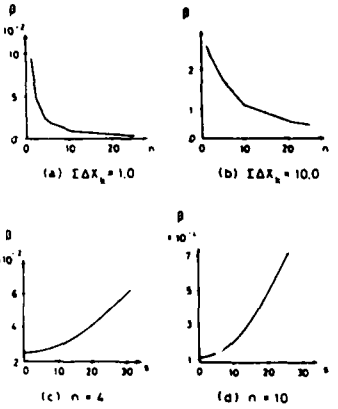


Fig.5
Effects of the number of divisions (a)and(b), and the variation of the widths (c)and(d). (s is the standard deviation of the widths)

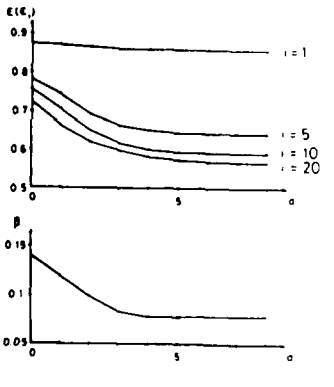


Fig.6
Comparison of the mean square error and β. The width of each region is 1.0 and the band width is 0.61.

VI. Conclusions

We have examined the properties of the iterative image restoration for the case where the data are available in multiple restricted regions. When the sum of the widths of the regions is small, we show a means for the construction of the optimum system, wich is based on the evaluation of the sum of the squares of the eigen values of the system.

References

1. R.W.Gerchberg, Optica Acta, vol.21, no.9, 709 (1974)
2. A.Papolis, IEEE trans. on Circuits and Systems, vol.22, no.9, 735 (1975)
3. T.Sato, et al. Appl. Opt. vol.20, no.3 395 (1980)
4. T.Inouye, IEEE trans. on Nucl. Sci. vol.26, no.2, 2666 (1979)
5. D.C.Youla, IEEE trans. on Circuits and Systems, vol.25, no.9, 694 (1978)
6. D.Slepian and H.O.Pollak, Bell Syst. Tech. J. vol.40, 43 (1961)

NOTES

SESSION II

EXTRAPOLATION AND RESTORATION II

A. A. Sawchuk, *Presider*

Matched Image Formation and Restoration System*

W. Thomas Cathey
Dept. of Electrical and Computer Engineering
University of Colorado at Denver
Denver, Colorado 80202

B. Roy Frieden
Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

William T. Rhodes
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Craig K. Rushforth
Dept. of Electrical Engineering
Univ. of Utah
Salt Lake City, Utah 84112

In the past, attempts to increase the resolution of images by modification of the imaging system or by image enhancement techniques have been two separate fields of endeavor. Research on imaging systems concentrated on resolution increase by use of such approaches as special phase masks, the rearrangement of the spatial bandwidth of a system, and trades of temporal bandwidth for spatial bandwidth [1-7]. The image restoration techniques have concentrated on spectral extrapolation using maximum entropy, Bayes' theorem, Gerchberg's algorithm,or some other iterative approach [8-12].

The goal of this work is to encourage an integrated approach to the problem. We present results of analysis of the effect of various distributions of the spatial bandpass of a system for three problems: (1) accurate location of a point source, (2) restoration of an image in noise, and (3) restoration of an object having a flat spectrum. We then discuss two ways of synthesizing an imaging system to gather the data with the desired spatial bandpass.

First, let us consider the simpler case of determining the location of a one-dimensional point source in the presence of noise. Two cases are explored. One is an aperture that forms a perfect lowpass spatial filter. The other is a perfect bandpass filter with the same total bandwidth. With no noise, it is intuititively clear that an aperture composed of two separated regions, which gives a narrower main lobe, is best if

*This work was supported in part by an Army Research Office Palantir Study.

there is no ambiguity problem. With noise, we used as our measure of resolution the variance of the maximum likelihood estimate of the location. If the S/N is high, we find the variance σ^2, to be

$$\sigma^2 = \frac{3}{\rho^2(a^2+12b^2)}$$

where ρ is a measure of the S/N, a is the width of the spatial bandpass, and b is its center frequency. We see that σ^2 decreases monotonically with b, the limit being set by the probability of the point being located with a secondary lobe.

A second problem is image restoration when the aperture can be split into three sub-apertures, one of which is centered. The restoration procedure used was a least-squares algorithm using singular-value decomposition. We assumed a known finite spatial extent of the object, but did not use a positivity constraint on the object [13-14]. It was found that by an appropriate choice of the distribution of the three apertures (one centered) the mean square error in the restoration of the chosen image was reduced from 0.355 for a conventional aperture to 0.0182 for the optimum three-element aperture.

The final analytical problem allowed the use of several properly phased apertures. It was shown that the optimum distribution of the aperture depends on the S/N. A Wiener-Helstrom filter was used and flat object and noise spectra were assumed. The general trend is that for a high S/N, the best array is one similar to a non-redundant array; and for a low S/N, the best approach is to cluster all the arrays to form a normal aperture, which results in a low-pass spatial filter.

Once a desired spatial frequency transfer function is selected, it is necessary to synthesize the aperture distribution to obtain that distribution. Image gathering systems are grouped into those having multiple receiving apertures and those having a single, fixed receiving aperture but several transmitting apertures. The former group has been studied extensively, and. if it is assumed that a phase reference is available, the apertures used in the three analyses could be synthesized. We devote our attention to the problem of a single receiving aperture and multiple transmitting apertures. Two cases are considered. In the first, we assume that we can record the amplitude and phase of the received wave, and in the second case, we record the intensity.

The transfer function for a single transmitter and several coherent receiving apertures is not the same as the transfer function for a single receiving aperture and several coherent transmitters, but they are similar. If the amplitude and phase are recorded, the effect of each transmitter can be considered separately. Because the angle of illumination is different for each transmitter, the receiving aperture collects different regions of the angular spectrum of the object. Use of a phase reference allows each of these regions to be recorded and recombined. The result is the building up of the spatial spectrum beyond what would be possible otherwise.

If the intensity is to be recorded, a high spatial frequency fringe structure can be projected onto an object by using two sources. The source angles can be changed to beat down different bands of spatial high frequencies. An analogous approach with multiple receiving apertures is the technique of interferometric imaging.

References

1. G. Toraldo di Francia, "Super Gain Antennas and Optical Resolving Power", Supplemento al Voi. IV, Serie IX del Nuovo Cimento, N.3, 426-438 (1952)

2. B. Roy Frieden, "On Arbitrarily Perfect Imagery With a Finite Aperature", Optica Acta 16, 795-807 (1969)

3. B. Roy Frieden, "The Extrapolating Pupil, Image Synthesis, and Some Thought Applications", Appl. Opt. 9, 2489-2496 (1970)

4. G. R. Boyer, "Pupil Filters for Moderate Superresolution", Appl. Opt. 15, 3089-3093 (1976)

5. R. Boivin and A. Boivin, "Optimized Amplitude Filtering for Superresolution Over a Restricted Field", Optica Acta 27, 587-610 (1980)

6. M. A. Grimm and A. W. Lohmann, "Super-Resolution Image for One-Dimensional Objects", J. Opt. Soc. Am. 6, 1151-1156 (1966)

7. A. W. Lohmann and D. P. Paris, "Super-Resolution for Nonbirefringent Objects", Appl. Opt. 3, 1037-1043 (1964)

8. R. W. Gerchberg, "Super-Resolution Through Error-Energy Reduction", Optica Acta 21, 709-720 (1974)

9. B. Roy Frieden, "Restoring with Maximum Likelihood and Maximum Entropy", J. Opt. Soc. Am. 62, 511-518 (1972)

10. Kikuchi and B. H. Soffer, "Maximum Entropy Image Restoration. I. The Entropy Expression", J. Opt. Soc. Am. 67, 1656-1665 (1977)

11. W. H. Richardson, "Bayesian-Based Interative Method of Image Restoration", J. Opt. Soc. Am. 62, 55-59 (1972)

12. C. K. Rushforth and R. L. Frost, "Comparison of Some Algorithms for Reconstructing Space-Limited Images", J. Opt. Soc. Am. 70, 1539-1544 (1980)

13. C. K. Rushforth, A. E. Crawford, and Y. Zhous, "Least-Squares Reconstruction of Objects with Missing High-Frequency Components", J. Opt. Soc. Am., 72, 204-211

14. Y. S. Shim and Z. H. Cho, "SVD Pseudoinversion Image Reconstruction", IEEE Trans, ASSP-29, 904-909 (1981)

IMAGE RESTORATION BY THE METHOD OF PROJECTIONS ONTO CONVEX SETS

by
H. STARK AND I.M. SEZAN
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Troy, New York 12181

SUMMARY

1. INTRODUCTION

In this paper we use the method of projections onto convex sets (POCS) to restore an image for which only partial knowledge of the Fourier transform is available. A spectacular improvement over the Gerchberg-Papoulis restoration is observed when the number of iterations is small.

The iterative solution for finding a point in the intersection of closed convex sets by projection was furnished by Gubin et al in [1]. Youla [2] extended Gubin's results to image processing and developed additional theory regarding the properties of compositions of projection operators onto closed convex sets with non-empty intersections. Lent and Tuy [3] were the first to apply the method to the extrapolation of bandlimited functions.

2. THEORY OF IMAGE RESTORATION BY THE METHOD OF POCS

Every known property of an unknown image f is viewed as a constraint that restricts the image to lie in a well-defined closed convex set C_i. Thus for m properties, there are m sets C_i i=1,...,m and the image must lie in the intersection $C_o \triangleq \bigcap_{i=1}^{m} C_i$. The problem is then to find a point of C_o, given the sets $\{C_i\}$ and the operators $\{P_i\}$ that project onto C_i. Now assume an arbitrary $f \in H$ where H is the Hilbert space consisting of the $L_{2\times 2}$ space of square integrable functions with inner product ($g \in H, h \in H$)

$$(g,h) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)h(x,y)dxdy \tag{1}$$

and norm

$$||g||^2 = (g,g)^{1/2}. \tag{2}$$

The projection $h \triangleq P_i f$ of f onto C_i is that element that satisfies

$$\min_{y \in C_i} ||f-y||^2 = ||f-h||^2. \tag{3}$$

Every point of C_o is a fixed point of the operator $T \triangleq P_m P_{m-1} \ldots P_1$. Conversely every fixed point of T is a point of C_o if C_o is not empty. To find a fixed point of C_o, the following iterative algorithm can be used

$$f_{k+1} = Tf_k \qquad f_o \text{ is arbitrary}, \tag{4}$$

where f_{k+1} converges weakly (i.e., inner product convergence) to the original image $f \in C_o$. The same conclusion holds when in Eq. (4) we use, $T \triangleq T_m T_{m-1} \cdots T_1$ where

$$T_i = 1 + \lambda_i (P_i - 1) \qquad 0 < \lambda_i < 2. \tag{5}$$

The λ_i's are called relaxation parameters and can sometimes be used to improve the rate of convergence.

The algorithms actually used in the restorations shown in the next section were cyclic-sequential:

$$f_{n+1} = T_{\alpha(n)} f_n \qquad f_o \text{ arbitrary}, \tag{6}$$

where $\alpha(n) \triangleq 1+n \bmod m$. Equation (6) has exactly the same convergence properties as Eq. (4). Note that m iterations (one cycle) of Eq. (6) is one iteration in Eq. (4).

3. EXPERIMENT OF RESULTS

Figure 1 shows the object f consisting of three nested rectangles on a dark background; the brightness values are 1.0, 0.8, 0.4 in going from smallest to largest rectangle. Figure 2 shows the discrete Fourier transform (DFT) of Fig. 1. Figure 3 is the portion of the Fourier data from which the restoration is attempted. In total, we are given the following a priori constraints, all of which are convex-set restrictions on f 1) P_1:f is space-limited, 2) P_2:$||f||^2 \leq 268.5$ ($||f||^2$=267.0), 3) P_3:$0 \leq f \leq 1$, 4) P_4: the Fourier data G given in Fig. 3. The explicit form of the projection operators are given in [2] and [4].

Figure 4 shows the results after 30 iterations. The original scene is shown in the upper-left; the restoration using projections P_1, P_2, and P_4 is shown in the lower left. Shown in the upper right is the restoration using $T_1 = 1 + 1.9995(P_1-1)$, $T_2 = 1 + 1.9995(P_2-1)$ and $T_4 = 1 + 1.75(P_4-1)$. In the lower right is shown the Gerchberg-Papoulis result.

Figure 5 is the same as Fig. 4 except that in the lower left is shown the restoration using P_1, P_2, P_3 and P_4 while in the upper right is shown the restoration using T_1=1 + 1.9995(P_1-1), T_2=1 + 1.9995(P_2-1), T_3=1 + 0.25(P_3-1), T_4=1 + 1.925(P_4-1). As before the G-P is shown in the lower right.

It is easily seen that the method of POCS furnishes far better restorations than the Gerchberg-Papoulis method. When appropriate relaxation parameters are used, the normalized error after 30 iterations is around 17 percent for POCS and around 39 percent for G-P.

4. REFERENCES

1. L.G.Gubin, B.T. Polyak and E.V. Raik, "The Method of Projections for Funding the Common Point of Convex Sets", U.S.S.R. Computational Mathematics and Mathematical Physics, 7, No. 6, pp. 1-24 (1967).

2. D.C.Youla and H. Webb, "Image Restoration by the Method of Convex Projections, Port I-Theory", submitted for publication.

3. A-Lent and H. Tuy, "An Iterative Method for the Extrapolation of Bandlimited Functions", Jour. of Math. Analysis and Applications, 83, pp. 554-565 (1981).

4. I.M. Sezan and H. Stark, "Image Restoration by the Method of Convex Projections, Part 2-Applications", submitted for publication.

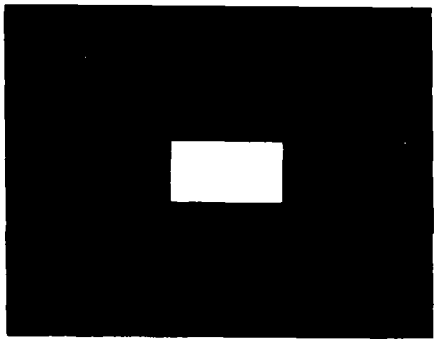


Fig. 1 - The Scene

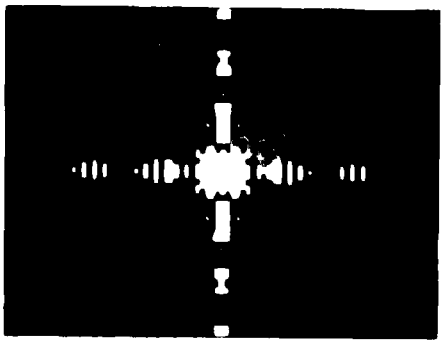


Fig. 2 - FT of Fig. 1

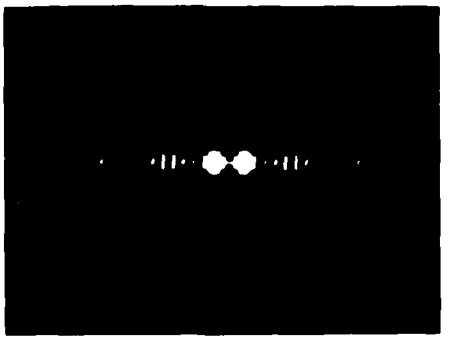


Fig. 3 - The given FT data

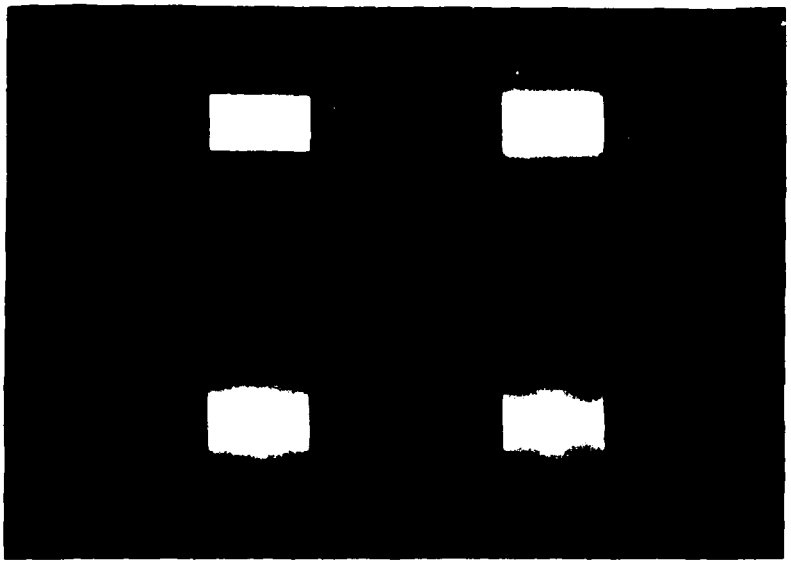
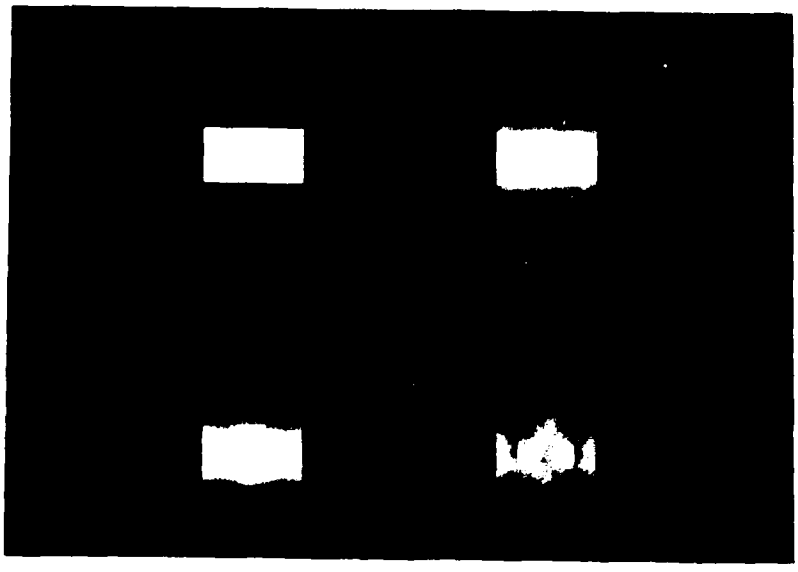


Fig. 4 - Original scene and restorations after 50 iterations



Stable, non-iterative, object reconstruction from incomplete data using prior knowledge

A.M. Darling, T.J. Hall, M.A. Fiddy
Physics Dept., Queen Elizabeth College (University of London), Campden Hill Road, London, W8 7AH, U.K.

The non-uniqueness and instability of object reconstruction from incomplete data can only be resolved by a priori constraints restricting the set of admissible solutions. A successful approach is to choose the object consistent with the data and of minimum norm in a weighted Hilbert space [1,2]. The weight is chosen to reflect our prior knowledge of the solution. The algorithm involves the solution of a set of linear equations with Toeplitz structure which can be efficiently solved in a finite number of steps by the Levinson recursion [3]. We show the equivalence between this method and Miller regularisation [4,5] for ill-posed problems. Experimental results demonstrating the effectiveness of the method are shown in the presentation [see also ref. 2].

Let f be an element of the Hilbert space F and suppose we have data consisting of the values C_n of a finite number of bounded linear functionals $F_n(f)$ defined on F. For example the sampled Fourier transform of f:

$$\hat{f}(x_n) = \int_{\Omega} f(t) \exp(itx_n)\,dt \tag{1}$$

may be considered as a bounded linear functional on $L^2(\Omega)$ whose value is the sample $\hat{f}(x_n)$. By the Riesz representation theorem [6] there exist unique $\phi_n \in F$ such that:

$$C_n = (f, \phi_n) \tag{2}$$

By the projection theorem [6] we may always decompose $f \in F$ into $f=u+v$ where $u \in \psi$, $v \in \psi\perp$ and $\psi \oplus \psi\perp = F$. Let $\psi = \text{span}\{\phi_n\}$ then we may write:

$$f = \sum_n a_n \phi_n + v \tag{3}$$

Substituting Eq. (3) into Eq. (2) we obtain the linear system:

$$\sum_m (\phi_m, \phi_n) a_m = C_n \tag{4}$$

These equations tell us nothing about v. Noting, however, that $(u,v)=0$ and so $\|f\|^2 = \|u\|^2 + \|v\|^2$ then the image $\tilde{f}$ of f formed from Eq. (3) by setting $v=0$ minimises $\|\tilde{f}\|$ and hence is the minimum norm solution consistent with the data. Since the error $f-\tilde{f}=v$ is orthogonal to ψ and hence to f, $\tilde{f}$ is also the solution which minimises $\|f-\tilde{f}\|$ subject to $\tilde{f} \in \psi$. Our reconstruction procedure is thus an orthogonal projection onto ψ. ($\psi\perp$ contains all the functions which give rise to zero data.) Let $F=L^2(\Omega; p^{-1})$ which has inner product:

$$(f,g) = \int_{\Omega} p^{-1} f g^* \,dt \tag{5}$$

where the prior knowledge weight p is real and positive. Then taking Eq. (1) as the data we have corresponding to Eqs. (3) & (4)

$$\tilde{f} = p(t) \sum_n a_n \exp(-ix_n t) \tag{6}$$

$$\sum_m p(x_m - x_n) a_m = \hat{f}(x_n) \tag{7}$$

$$P(x) = \int_{\Omega} p(t) \exp(ixt)\, dt \tag{8}$$

f will obviously have the form of the prior knowledge p yet the choice of the coefficients a_n ensures that f is consistent with the data and that $\|f-\tilde{f}\|$ is minimised. The inverse weighting in $\|f-\tilde{f}\|$ by p is interesting since if p is close to $|f|$ it forces the estimate to be more accurate where $|f|$ is small enhancing low contrast features. The procedure is stable against noise if p is chosen to have a small non-zero value even in regions where f is known not to be supported, reflecting our prior knowledge that there will always be out of band noise. For regularly sampled data $\{P(X_m-X_n)\}_{nm}$ is Hermitian Toeplitz so Eq. (7) may be efficiently solved using the Levinson recursion [3]. When $p(t)=1$ $t \in [-\Delta\pi, \Delta\pi]$ $\Delta < 1$, $p(t)=0$ elsewhere and the samples are at integer positions Eqs. (6)-(8) are the same as those solved iteratively by the Gerchberg/Papoulis bandlimited extrapolation procedure [7].

Let $f \in F$, our object, and $g \in G$, our data, be related by $Af+n=g$ where $A:F \rightarrow G$ is a linear continuous operator with non-existent or unbounded inverse and n is additive noise. Miller's method [4,5] consists of searching for an f satisfying:

$$\|Af-g\|_G \leq \epsilon \qquad \|Bf\|_H \leq E \tag{9}$$

where $B:F \rightarrow H$ is a linear constraint operator densely defined on F and with bounded inverse. Introducing the functional $\Phi(f) = \|Af-g\|^2_G + (\epsilon/E)^2 \|Bf\|^2_H$ it is easy to show that any function satisfying $\Phi(f) \leq \epsilon^2$ satisfies Eq. (9) and any function satisfying Eq. (9) satisfies $\Phi(f) \leq 2\epsilon^2$. We therefore chose our reconstruction as the function f which minimises $\Phi(f)$ this is given by:

$$(A^*A + (\epsilon/E)^2 B^*B)\tilde{f} = A^*g \tag{10}$$

Taking $F=L^2(\Omega;p^{-1})$ and A as the sampled Fourier transform operator:

$$Af = \sum_n (f,\phi_n)_F e_n = \sum_n \hat{f}(x_n) e_n = g \tag{11}$$

where e_n is the natural basis in $G=L^2(w)$ which has inner product $(u,v)_G = \sum_n w_n u_n v_n^*$ and w_n are real and positive, we then have:

$$A^*g = \sum_n (g,e_n)_G \phi_n \tag{12}$$

so if $B^*B=I$ (i.e. a constraint on the norm) we conclude again that $\tilde{f}$ has the form of Eq. (6) and where the a_n are now given by:

$$\sum_m [P(x_m-x_n) + w_n^{-1}(\epsilon/E)^2 \delta_{nm}] a_m = \hat{f}(x_n) \tag{13}$$

When $(\epsilon/E)^2 \rightarrow 0$ (no noise) the two formulations are identical. For $W_n=1$ it is easily shown that the term in $(\epsilon/E)^2$ in Eq. (13) is equivalent to setting the prior knowledge to $p+(\epsilon/E)^2, t \in [-\pi,\pi]$ (for samples at integer positions), solving for the coefficient a_n using $p+(\epsilon/E)^2$ in place of p in Eq. (8) and then forming the reconstruction Eq. (6) with p left as it is. That is, the method is equivalent to letting p take a small non-zero value even where f is not supported and then filtering the noise away from the reconstruction. Miller's theory is particularly useful as it can provide us with precise

Frequency Domain Optimal Inverse Convolution Filtering of Noisy Data

C. B. Chittineni
Conoco Inc.
P. O. Box 1267
Ponca City, OK 74603

SUMMARY

In many physical experiments, the observations r(m,n) can be represented as the convolution of a model p(m,n) with a source function s(m,n) subject to additive noise e(m,n). The ultimate goal of deconvolution schemes[1,2] is to recover both the model and the source function using observations and all other available information. Let $d(m,n) = p(m,n) * s(m,n)$. The signal-to-noise ratio α is defined as $\alpha = \Sigma\Sigma d^2(m,n)/\Sigma\Sigma e^2(m,n)$.

Usually, some knowledge about the noise is available. In this paper, it is assumed that α is known. In practice, α can either be measured or estimated from the data. Assuming the source function is known, algorithms are developed in the following for the estimation of the model by taking into account the signal-to-noise ratio.

A. Signal Extraction Through Constrained Wiener Estimation

An estimate $\hat{p}(m,n)$ of the signal p(m,n) is sought in the form $\hat{p}(m,n) = h(m,n) * r(m,n)$, where h(m,n) is the constrained Wiener filter. Let $\hat{d}(m,n) = \hat{p}(m,n) * s(m,n)$ and $\hat{e}(m,n) = r(m,n) - \hat{d}(m,n)$. For a particular $\hat{p}(m,n)$, an estimate $\hat{\alpha}$ of the signal-to-noise ratio α can be written as $\hat{\alpha} = \Sigma\Sigma \hat{d}^2(m,n)/\Sigma\Sigma \hat{e}^2(m,n)$. Then the problem is to estimate h(m,n) such that $\epsilon^2 = E[\Sigma\Sigma(p(m,n) - \hat{p}(m,n))^2]$ is minimized, subject to the constraint $\alpha\Sigma\Sigma e^2(m,n) - \Sigma\Sigma d^2(m,n) = 0$, where E is an expectation operator and p(m,n) is the desired but unknown signal. The solution in the frequency domain is given by:[3]

$$H(u,v) = \frac{S^*(u,v)\, P_p(u,v) + \lambda\alpha\,|R(u,v)|^2\, S^*(u,v)}{[E(|R(u,v)|^2) + \lambda(\alpha - 1)\,(MN)^2\,|R(u,v)|^2\,|S(u,v)|^2]}$$

where $P_p(u,v)$ is the power spectral density of p(m,n) and λ is the Lagrangian multiplier. It is seen that when $\lambda = 0$, H(u,v) becomes a Wiener filter.[1] For a particular λ, let $\rho(\lambda)$ be the constraint. It can be shown[3] that the $\rho(\lambda)$ is a monotonically decreasing function of λ and can be adjusted very efficiently in the frequency domain by a Newton-Raphson-like iterative procedure.

B. Signal Extraction Through Constrained Restoration

In the case of seismic data, p(m,n) represents the reflection series and hence will have some quality of "spikiness." The wavelet s(m,n) will be

smooth compared to p(m,n). These characteristics can be taken into account through the measures for spikiness or smoothness. Then the estimation problem is to find an estimate $\hat{p}(m,n)$ of p(m,n) by minimizing $\Sigma\Sigma\big(c(m,n) * p(m,n)\big)^2$, the criterion of smoothness, subject to a constraint on the signal-to-noise ratio, where c(m,n) is a smoothing array. The frequency domain estimate for p(m,n) can be obtained as:[3]

$$\hat{P}(u,v) = \frac{\lambda\ \alpha\ R(u,v)\ S^*(u,v)}{MN[\,|c(u,v)|^2 + \lambda(\alpha - 1)\ |S(u,v)|^2]}$$

where λ is a Lagrangian multiplier. The constraint $\rho(\lambda)$ is a monotonically decreasing function of λ, and hence λ can be adjusted very efficiently in the frequency domain.

C. Signal Extraction Through Constrained Mapping

The p(m,n) is estimated as $\hat{p}(m,n) = r(m,n) * h(m,n)$. The filter h(m,n) is determined such that it transforms s(m,n) into a desired signal x(m,n) (e.g., a Dirac delta function) subject to a constraint on the signal-to-noise ratio of the estimated signal p(m,n). Let $y(m,n) = s(m,n) * h(m,n)$; $f(m,n) = e(m,n) * h(m,n)$; and $d(m,n) = p(m,n) * x(m,n)$. Then $p(m,n) = d(m,n) + f(m,n)$. The signal-to-noise ratio of p(m,n) is given by $\alpha_0 = \Sigma\Sigma\, d^2(m,n)/\Sigma\Sigma\, f^2(m,n)$. In the frequency domain, an estimate for the filter can be obtained as:[3]

$$H(u,v) = \frac{S^*X + \lambda\ \alpha_0\ MN\ R^*\ P\ X}{MN[\,|S|^2 + \lambda\ \alpha_0\ |R|^2]}$$

The constraint $\rho(\lambda)$ is a monotonically decreasing[3] function of λ and hence can be adjusted very efficiently in the frequency domain.

Some experimental results of the techniques described in the paper are presented below. The data are given in Figure 1. The model p(m) and the source function s(m) are shown in Figures 1a and 1b, respectively. The trace r(m) with $\alpha = 2$ is given in Figure 1c. The noise is added only for half of the trace. Figure 1d shows the predictive deconvolution of the signal in Figure 1c.

Figure 2 shows the iterative estimates of s(m) and p(m) obtained in the following manner. Initial estimate of p(m) is taken as the output of predictive deconvolution algorithm. Initial estimate of s(m) is obtained using constrained restoration technique, with c(m) chosen as (2, -1, -2, -1, 2) for m = 1, 2, ..., 5 and c(i) = 0 for $i \neq m$. The estimate of s(m) is updated using constrained Wiener restoration algorithm, and the estimate of p(m) is updated using the constrained mapping technique. $E\big(|R(u)|^2\big)$ is replaced with $|R(u)|^2$. The procedure is repeated two times. It can be seen from Figure 2 that reasonably good estimates are obtained for both s(m) and p(m).

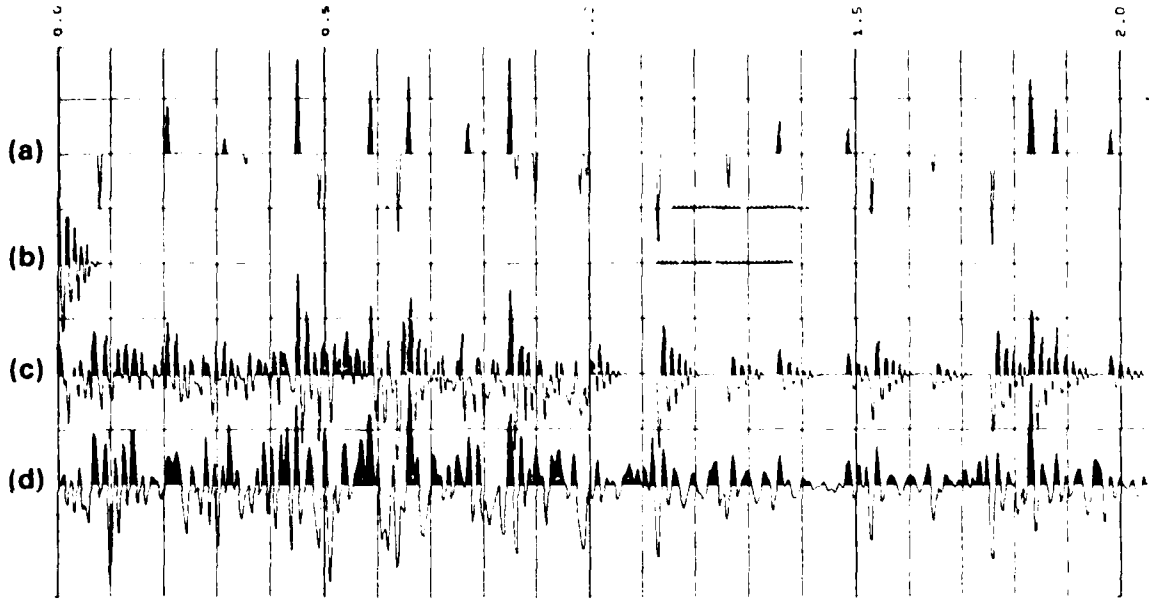


Figure 1. Plotted in succession are (a) the model p(m) (b) the wavelet s(m) (c) the trace r(m) with signal to noise ratio 2. (d) the predictive deconvolution of the trace in (c)

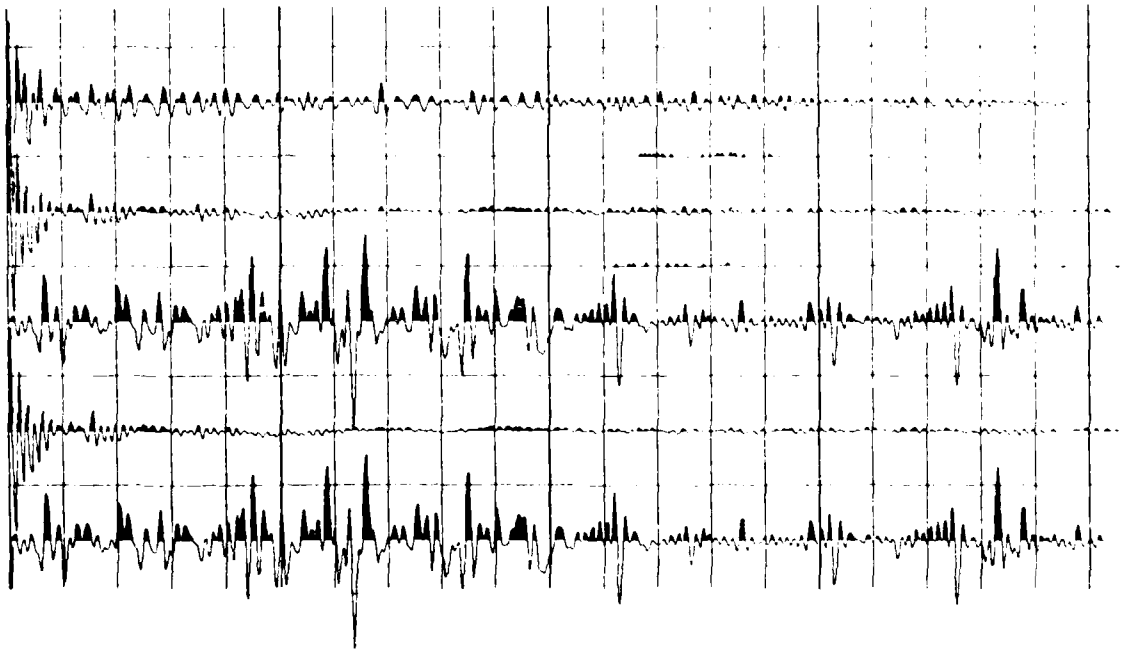


Figure 2. Blind restorations of s(m) and p(m) for signal to noise ratio 2

REFERENCES

[1]Andrews, H. C., and Hunt, B. R., Digital Image Restoration, Prentice-Hall: New Jersey, 1977.

[2]Webster, G. M., ed., Deconvolution, Vols. I and II, Geophysics Reprint Series, Society of Exploration Geophysicists, Tulsa, Oklahoma, 1978.

[3]Chittineni, C. B., Frequency domain optimal inverse convolution filtering of noisy data, Research Report 5523-7-1-82, Conoco Inc., Ponca City, Oklahoma, July 1982.

Optimal Iterative Image Reconstruction with Incomplete and Approximate Data

J. Gassmann, Max-Planck-Institut für Plasmaphysik, 8046 Garching, FRG

For the determination and improvement of phases in X-ray analysis a procedure called "phase correction" or "density modification" has been developed[1,2]. It is applicable to all types of images when the spectral density is known. It has been applied to various structure determinations of small molecules[3], resolution extension of macro molecules[4], phase improvement in a virus structure[5] and reduction of the effect of missing projection data in electron microscopy.

This iterative process of structure determination has been analysed in details with reference to information theory[6] and electron microscopy[7]. The phasing procedure consists generally of an iterative weighted filtering in direct and reciprocal space according to a Wiener-Hopf optimization. The phase probabilities have the general form $W = I_1(X)/I_0(X)$, where I_0 and I_1 are modified Bessel functions. Criteria for the modification process $\varrho \rightarrow \varrho^*$ of the image density ϱ are density cutoff, positivity, finite extension, maximal brightness, symmetry and statistical information. Independent of the modification process a residual value $R = \int (\varrho - \varrho^*)^2 dV$ must be minimized, equivalent to $\int \varrho \varrho^* dV = max$ where the improved image density ϱ^* is obtained from the initial density by application of some modification operator $\varrho^* = D\{\varrho\}$[8].

The effect of a specific modification operator on the phasing process varies with the type of and information about the image. The phase probabilities for the correctness of a modification are represented by weighting of the filtering process.

Statistical information independent of the image is used in the so-called "maximum entropy approach"[9,10]. This approach applied as an iterative process is a special case of "phase correction". This additional statistical information has an insignificant effect on the iteration process compared with the effect of other modification operators and is unsuitable for objects occurring in X-ray analysis and astronomy[11,12].

Different statistical modifications ("minimal entropy") likewise do not improve appreciably the convergence of the phasing process[13].
Criteria for the convergence of the phasing process are obtained with respect to the information content of the image. The objects in X-ray analysis (molecules) are sparsely distributed in many resolution elements. Therefore the solution of the phase problem is generally possible by iterative phase determination in this field. A numerical method of evaluating the necessary initial image information based on the relative importance of correct phase values against accurate spectral amplitude values is developed. The necessary amount and accuracy of the starting information for convergence of the iteration process is deduced from this calculation.

1) W. Hoppe, R. Huber and Gaßmann:
Acta Cryst. 16(1963) A 4.

2) W. Hoppe and J. Gaßmann:
Acta Cryst. B 24(1968) 97.

3) J. Gaßmann and K. Zechmeister:
Acta Cryst. A 26(1972) 27o.

4) R.W. Schevitz, A.D. Podjarny, M. Zwick, J.J. Hughes, P.B. Sigler:
Acta Cryst. A 37(1981), 669

5) G. Bricogne:
Acta Cryst. A 32(1976), 832

6) J. Gaßmann:
Acta Cryst. A 32(1977), 474

7) J. Gaßmann:
Optik 48 (1977), 347

8) J. Gaßmann:
Acta Cryst. A 32 (1976), 274

9) Frieden B. (1972) J. Opt. Soc. Am. 62, 511

1o) Kikuchi R., Soffer B. (1977) J. Opt. Soc. Am. 67, 1565

11) Wernecke S., D'Addario L. (1977) IEEE Trans. Comp. C 26, 351

12) Gull S., Daniell G. (1978) Nature 272, 686

13) J. Gaßmann:
DFG-Kolloquium Digitale Signalverarbeitung, Göttingen Okt. 1978

SESSION III

EXTRAPOLATION AND RESTORATION III

C. Rushforth, *Presider*

Detection of Line Spectra and Point Sources

A. Papoulis
Polytechnic Institute of New York
Route 110
Farmingdale, N. Y. 11735

Summary

The problem of detecting an object consisting of point sources in a noisy environment in terms of a single sample of a diffraction limited image is considered. The proposed method in its simplest form can be phrased as follows:

We assume that the autocorrelation of a process $s[n]$ is of the form

$$R[m] = \lambda_o\, \delta[m] + \sum_{i=1}^{N} \gamma_i\, e^{j\omega_i m} \tag{1}$$

We wish to estimate its parameters λ_o, γ_i, and ω_i in terms of the estimates $\hat{R}[m]$ of $R[m]$ obtained from a single sample of $s[n]$.

The underlying investigation is based on a theorem by Caratheodory[1] used by Pisarenko in the context of the above problem in its deterministic form where it is assumed that $R[m]$ is known for $|m| \le N$. We present here a modified form of the solution as a variation of Levinson's algorithm and we extend the investigation to the random case.

As we recall, in the maximum entropy (ME) method of estimation, it is assumed that

$$S(\omega) = \frac{e_N}{|1 - a_1^N e^{-j\omega} - \cdots - a_N^N e^{-jN\omega}|^2} \tag{2}$$

The parameters a_k^N are the coefficients of the MS predictor

$$\hat{\underset{\sim}{s}}[n] = a_1^N \underset{\sim}{s}[n-1] + \cdots + a_N^N \underset{\sim}{s}[n-N] \tag{3}$$

and

$$e_N = E\{|\underset{\sim}{s}[n] - \hat{\underset{\sim}{s}}[n]|^2\} = \Delta_{N+1}/\Delta_N \tag{4}$$

where Δ_N is the determinant of the correlation matrix.

These parameters can be found recursively:

$$e_{N-1}\Gamma_N = R[N] - \sum_{k=1}^{N-1} a_k^{N-1} R[N-k]$$

$$a_k^N = a_k^{N-1} - \Gamma_N a_{N-k}^{N-1} \qquad 1 \le k \le N-1 \tag{5}$$

$$a_N^N = \Gamma_N \qquad e_n = (1-\Gamma_N^2)e_{N-1}$$

The recursion starts with $e_o = R[0]$.

We establish Caratheodory's result using the following property of the roots z_i of the error filter

$$E_N(z) = 1 - a_1^N z^{-1} - \cdots - a_N^N z^{-N}$$

In all cases, $|z_i| \le 1$ because $|\Gamma_i| \le 1$. If

$$\Delta_{N+1} = 0 \qquad \Delta_N \neq 0$$

then all roots are on the unit circle

$$z_i = e^{j\omega i} \qquad R[m] = \sum_{i=1}^{N} \gamma_i e^{j\omega_i m} \qquad (6)$$

The above is a limiting form of the ME entropy method but it is not the ME solution to the problem. In this case, $S(\omega)$ is not as in (2) but it consists of lines

$$S(\omega) = 2\pi \sum_{i=1}^{N} \gamma_i \delta(\omega - \omega_i) \qquad |\omega| < \pi \qquad (7)$$

Furthermore, $\Gamma_N = 1$, $e_N = 0$, and $s[n]$ is predictable.

If $\Delta_{N+1} \neq 0$, then we form the modified autocorrelation

$$R^o[m] = R[m] - \lambda_o \delta[m] \qquad (8)$$

and we select λ_o such that the corresponding determinant Δ^o_{N+1} is zero. We then proceed as before.

In the stochastic case, we estimate $R[m]$ from the given data and we attempt to determine the statistical properties of the estimates $\hat{\lambda}_o$, $\hat{\gamma}_i$, and $\hat{\omega}_i$ of the unknown parameters. The results are preliminary.

The approach used leads to a simple derivation of Wold's decomposition theorem in the context of the Levinson algorithm.

References

[1] U. Grenauder and G. Szego, Toeplitz Forms and Their Applications, Berkley Univ. Press, 1958.

[2] V.F. Pisarenko, "The Retrieval of Harmonics", Geoph. J.R. Astro. Soc., 1973.

[3] A. Papoulis, "Maximum Entropy and Spectral Estimation: A Review", IEEE, Trans. ASSP, pp. 1176-1187, December, 1981.

Reconstruction of Objects from Coded Images by Simulated Annealing

Warren E. Smith

Harrison H. Barrett

Richard G. Paxman

Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

and

Radiology Department
Arizona Health Sciences Center
University of Arizona
Tucson, Arizona 85724

Summary:

Coded apertures are useful for imaging sources of x rays and gamma rays in nuclear medicine, x-ray astronomy, laser fusion studies, and nuclear reactor safety research. While conventional apertures such as a pinhole or multibore collimator provide only two-dimensional (2D) projections of a three-dimensional (3D) source, coded apertures give some information about the 3D structure of the source. The easiest way to see this is by considering a point source some finite distance from the aperture plane. The coded image in this case is merely an enlarged shadow of the aperture transmission function. This shadow

contains all the necessary information to fully characterize the source. The scale of the shadow gives the longitudinal coordinate of the point (i.e., its distance from the aperture plane), and the lateral position of the shadow on the detector is simply related to the lateral position of the point in a plane parallel to the aperture. The strength of the source is related to the number of recorded photons in the coded image. Thus the shadow uniquely determines the 3D location of the source and its strength if we know a priori that it is a point source. Of course, we are usually interested in more complicated sources, and the central problem of coded-aperture imaging may be stated as follows: Given one or more 2D coded images and possibly some a priori information about the source, estimate the 3D distribution of the source.

In the absence of a priori information, this problem is clearly unsolvable. A 2D coded image contains, say, N^2 independent pixels, while a 3D object represented digitally at the same resolution as the coded image has N^3 independent volume elements (voxels). The problem is thus equivalent to solving N^2 equations in N^3 unknowns if only one coded image is available.

Fortunately, some very useful a priori information is readily available. The object distribution is inherently non-negative and is usually known to be entirely contained in some specified volume, such as the patients' body in the case of nuclear medicine. However, it is by no means clear whether this information is sufficient to guarantee a unique, stable solution. Indeed, it is possible to construct counter-examples where several different non-negative objects of bounded support lead to the same coded image. Nevertheless, some degree of success has been achieved in solving this problem by means

of iterative, constrained image-processing algorithms. This work is reported elsewhere at this meeting by Paxman, Gindi, and Barrett.

Recent work by Kirkpatrick et al., at IBM suggests a radically different approach to this problem, as well as to many other problems of constrained optimization. These workers have been addressing the problem of optimal layout of components and connecting wires on an integrated circuit chip. Building on an idea due to Metropolis in the 1950's, they have developed an extraordinarily successful method called "simulated annealing". The name derives from a close analogy with statistical mechanics. When a liquid freezes to form a solid, the lowest-energy state of the solid is often a perfect single crystal. However, in practice, the actual solid is more or less disordered and perhaps even completely random as with glasses. To get a good crystal, the temperature must be lowered slowly and gradually, especially near the freezing point.

The analogy with optimization problems is that some cost function plays the role of an energy. In the IBM problem, the cost function is related to the complexity of connecting wires and number of pins on the IC chip. In image processing, we are trying to find an estimate of an object that is consistent with measured data (such as a coded image), and a natural cost function is the RMS difference between measured and estimated data. A difficulty with traditional iterative approaches to this problem is that the object estimate can get stuck in a local minimum where the RMS error is indeed minimal with respect to small perturbations of the estimate, but which is quite far from the true global minimum.

Simulated annealing avoids this problem by not demanding that all modifications of the object estimate reduce the cost function. Rather, some modifications that increase the cost function ("energy") by an amount ΔE are accepted with a probability $\exp(-\Delta E/kT)$, where kT is a controllable parameter. Since T is naturally interpreted as a temperature, several other concepts from statistical mechanics can be exploited. One can monitor the "specific heat" or "entropy" of the system, thereby detecting phase transitions where the "temperature" should change more gradually. In this way, an optimum "annealing schedule" can be developed.

We have applied this idea to the coded-aperture problem with striking results. The geometry simulated used two 1D coded apertures, each consisting of five pinholes, and oriented orthogonally to each other. The object was two-dimensional (64×64 pixels), and the data were two separate 1D coded images (256 pixels each). Thus we attempted to reconstruct an object of 4096 points using 512 measured data points. The algorithm involved randomly adding or subtracting points from the object estimate in a Monte Carlo fashion. If addition or subtraction of a point lowered the energy (RMS error of the coded image), that modification of the object estimate was always accepted. If it raised the energy by ΔE, it was accepted with probability $\exp(-\Delta E/kT)$. After some 10^5 points were either added or subtracted in this manner, very good object estimates were obtained.

Acknowledgment: We are grateful to J. Fienup for calling to our attention the work of Kirkpatrick et al., and for suggesting its application to image processing. This work was supported by the National Cancer Institute.

Signal Deconvolution Using Frequency and Time Domain Magnitude Constraints

H.J. Trussell and P.N. Sura
Department of Electrical Engineering
North Carolina State University
Raleigh, NC 27650

Introduction

In many signal restoration problems we have very limited knowledge about the statistics required for implementation of the most common restoration techniques, for example, Wiener filtering. We do, however, possess some practical a priori knowledge about the nature of the signal we week to estimate. Because of this a priori knowledge, it may be suboptimal to use methods which make few assumptions about the nature of the ideal signal, for example, maximum entropy restoration [1]. Iterative restoration methods can be easily modified to incorporate such a priori knowledge while requiring little statistical information [2].

A common application of bounding constraints using iterative methods is in the restoration of non-negative signals [2,3]. Other uses of this type of method have been in applying frequency domain constraints [4]. For example, [5] assumes exact knowledge of the power spectrum over some limited region of the frequency domain and obtains superresolution by applying a finite spatial support constraint. The work reported here will use some of these techniques and extend the use of magnitude constraints to both upper and lower bounds as well as applying finite support constraints in both the space/time and frequency domains.

Implementation of Constraints

The iterative restoration method used in this study was the modified Landweber iteration [6]. The method is designed to restore signals degraded by a linear system. Using algebraic notation

$$\underline{g} = \underline{H}\underline{f}+\underline{n} \tag{1}$$

where $\underline{g}$ is degraded signal

$\underline{f}$ is ideal signal about which we have some a priori knowledge of its characteristics

$\underline{H}$ is a degrading function usually representing convolution

$\underline{n}$ is noise with known statistical properties.

The iterative method is defined by

$$\underline{f}_{k+1} = C[\underline{f}_k + \underline{D}(\underline{g} + \underline{H}\underline{f}_k)] \tag{2}$$

where C is an operator which enforces the constraints. The iteration is said to have converged when

$$||\underline{g} - \underline{\underline{H}}\underline{f}_k|| \approx ||\underline{n}|| \qquad (3)$$

This convergence criterion is discussed in [7,8].

The constraints which were applied fall into three classes:

I. Upper and lower limits on $\hat{\underline{f}}$ in the space/time domain;

II. Upper and lower limits on the power spectrum of $\hat{\underline{f}}$;

III. Finite regions of support in frequency and space/time domain.

In order to be effective the constraining boundaries must be close to the actual limits of the true signal. Nothing is accomplished by a constraint which is satisfied by an unconstrained algorithm. It is the purpose of this research to determine the usefulness of a priori knowledge of boundary limits. The usefulness of such knowledge depends on the characteristics of the original signal, $\underline{f}$.

Much work has been done applying the nonnegativity constraint to signal resotration [2-4]. The examples are usually signals which consist of a few nonzero points on a zero (or near zero) background. It was noted during this earlier work that a small bias could result in a greatly distorted restoration. This simply confirms the statement about the utility of a priori knowledge in the previous paragraph.

Results

1. Application of Upper and Lower Bounds in Space/time Domain.

The signal must have a significant number of points near the boundary for the constraint to be effective. As mentioned before the lower bound of zero is very beneficial in the restoration process. An upper bound applied to this type of signal is useless since only a few points would be effected. In practice the problem is to get a restored estimate with values as high as the true signal.

The signal which benefited the most from the space/time bounds was one composed of several distinct levels. In this case the upper and lower bounds were known and the restoration produced using the constraints was obviously superior to unconstrained restoration. The effect of a small error in the bounds was noticeable indicating that the exact knowledge is important. Another effect of this type of error is that the iterative method fails to converge.

A more typical signal was a scan line taken from a two dimensional image. The line was blurred and noise added. The restoration of this signal showed little benefit from the boundary constraints even though the exact values were used.

2. Application of Upper and Lower Bounds in Frequency Domain.

It is generally true that most digital signals have more energy in the lower frequencies than in the higher frequencies. This is true in large part because the analog signals from which most digital signals are derived are sampled well above the Nyquist rate. Restoration methods may increase the power in the high frequency portion of the spectrum beyond what is appropriate for a given signal. This is true even of the Wiener filter. Because of these properties it is reasonable to expect a frequency dependent bound to have positive effects on the restoration.

The bounds used for this study were generated by using a percentage of a smoothed spectral estimate of the original signal. The smooth signals and the scan line data benefited most from this constraint. The signal composed of several distinct levels had more variation in its power spectrum and the restoration was actually degraded by constraint. Furthermore, the restoration of the smoother signals was not very sensitive to error in the bounds.

3. Application of Finite Regions of Support.

The constraints of regions of support is mathematically an adaptive boundary constraint in which the upper and lower bounds are zero outside the region of support. The determination of the region of support is usually easier in the space/time domain than in the frequency domain. In either case, the region of support is directly effected by the noise level. When the noise power is sufficiently greater than the signal power in a particular region, the signal may be neglected.

The signals which benefited most from the spatial support constraints were the ones whose unconstrained restoration had ringing beyond the region of support. Both finite level and smooth signals benefited from this constraint. The finite level signal benefited by a greater reduction of mean square error as would have been predicted.

The effect of enforcing a finite support region in the frequency domain had predictably different results. The signals which were smooth have little high frequency power and thus benefited most. The signals have sharp edge transitions benefit from this type of constraint only when the signal-to-noise ratio is very low.

4. Application of Combinations of Constraints.

It was found to be very useful to combine several of the constraints discussed earlier. The results while not additive in any sense are at least qualitative extensions of the individual constraints. For example, the signal with only a few distinct levels is benefited more by a combination of upper and lower bounds and finite support than by either individually. The additional benefit is not large but is noticeable.

Comments on Comparisons

It is always most convenient to compare results by examining their mean square errors (mse). This method used alone would yield misleading conclusions in the application of constraints as studied here. In most cases the application of the constraint resulted in a reduction of mse but also positively affected the subjective quality of the solution. Several times an inappropriate constraint reduced the mse but adversely affected the subjective quality.

References

[1] B.R. Frieden, "Restoring with Maximum Likelihood and Maximum Entropy," J. Opt. Soc. Am., 62, 4, pp. 511-518, April 1918.

[2] R.M. Mersereau, R.W. Schaffer and M.A. Richards, "Constrained Iterative Restoration Algorithms," Proc. IEEE, 69, 4, pp. 432-450, April 1981.

[3] H.J. Trussell, "Maximum Power Signal Restoration," IEEE Trans. Acoust. Speech, and Signal Processing, ASSP-29,5, pp. 1059-1061, October 1981.

[4] J.R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," Proc. Soc. Photo-Optical Inst. Engineers, 207-02, Applications of Digital Image Processing III, Aug. 28, 1972.

[5] R.W. Gerchberg, "Super-resolution through error energy reduction," Optica Acta, 21,v9, pp. 709,720, Dec. 1977.

[6] O.N. Strand, "Theory and methods related to the singular-function expansion and Landweber's iteration for integral equations of the first kind," SIAM J. Numer. Anal.,v11, pp 798-825, Sept. 1974.

[7] H.J. Trussell, "Convergence of Iterative Restoration Methods," Proc. IEEE Int. Conf. on Acoust., Speech and Signal Processing, 1982, Paris France, Vol. 3, pp. 1527-1530.

[8] H.J. Trussell, "Convergence Criteria for Iterative Restoration Methods," IEEE Trans. Acoust., Speech and Signal Processing, to be published.

WA15-1

A Model-Based Restoration Procedure For Small, Low-Resolution Optical Images

John A. Saghri and Andrew G. Tescher

Advanced Concepts Office, The Aerospace Corporation
P. O. Box 92957, Los Angeles, California, 90009

Introduction

Compensation of image degradations caused by inaccuracies in the optics or sensing mechanism, transmission error, or reflection from spurious objects is commonly referred to as image restoration. Restoration techniques require at least limited information about the degradation phenomenon (e.g., point spread function, characteristics of the noise and the sensing mechanism). These techniques are often based on classical one-dimensional signal processing. An efficient restoration method should use the maximum amount of a priori information about the object.

Consider the discrete-discrete linear image model representation in vector form,

$$s = Hf + n \tag{1}$$

where s = recorded image intensity
f = object intensity
H = point spread function or blurring matrix
n = noise process

The classical restoration techniques are maximum posteriori (MAP), maximum entropy (ME), maximum likelihood (ML), and the least mean square error (LMSE). In the MAP restoration[1], f^* is found by maximizing the conditional probability density of the object, i.e.;

$$f^*_{MAP} = \max_f P(f/s) \tag{2}$$

Knowledge of the probability distribution of the object and the noise are required. In the ME restoration[2,3]

$$f^*_{ME} = \max_f (-f^T \ln f) \tag{3}$$

$$\text{subject to } \|s - Hf\|^2 = \|n\|^2$$

In ML[4] restoration, the solution is the f^* for which the noise probability is maximized. For additive noise we have

$$f^*_{ML} = \max_f P(s/f) = \max_f P(s - Hf) \tag{4}$$

In LMSE[5] f^* is selected when

$$f^*_{LMSE} = \min_f \|s - Hf\|^2 \tag{5}$$

For a limited number of image samples (less than 10) the classical techniques may be inappropriate. For such cases, an analytical model-based (deterministic) solution via numerical analysis can yield good image restoration.

This paper presents a model-based restoration algorithm. The discussion is for the continuous-discrete case, that is, the original image to be reconstructed is continuous and the observed image is discrete. The objective is to estimate the parameter values of the assumed object model based on the sampled observation of a degraded image. The technique is similar to the LMSE restoration; however, a deterministic rather than a stochastic object model is assumed.

Simulation results with a doublet source have been promising. The noise model is an additive, zero-mean, uncorrelated Gaussian process with constant variance. Other noise models[2,6] will be considered in later studies.

Problem Statement

Consider the image of an object being sensed by a detector array. Let the vector $\underline{X} = [x,y]^T$ represent the spatial coordinates of an object point. The object is assumed to be modeled parametrically in terms of N parameters $\underline{P} = [P_1, P_2, \ldots P_N]^T$.

For a linear incoherent imaging process, the image intensity at point $\underline{X}_i$ is given by

$$g(\underline{X}_i) = \int\int_{-\infty}^{\infty} f(\underline{X}, \underline{P}) \cdot h(\underline{X}, \underline{X}_i) d\underline{X} \tag{6}$$

where $f(\underline{X}, \underline{P})$ is the parametric image model and $h(\underline{X}, \underline{X}_i)$ is the optical system point spread function.

The image plane is sampled by an array of detectors. The response of the j-th detector, S_j, is proportional to the summation of the radiant energy over the detector area.

$$S_j = K \int_{A_j} \left[\int\int_{-\infty}^{\infty} f(\underline{X},\underline{P}) \cdot h(\underline{X},\underline{X}_i) d\underline{X} \right] d\underline{X}_i + E_j \tag{7}$$

where A_j is the area of the j-th detector, K is the known constant of proportionality and E_j is the noise effect associated with the j-th detector.

The problem is to estimate the set of N parameters $\underline{P} = [P_1, P_2, \ldots P_N]$ given M detector responses $(M \geq N)$.

Solution

Two effective techniques to estimate $\underline{P}$ are maximum likelihood and least mean square error restoration processes. The joint probability density of the noise, $P(E_1, E_2, \ldots E_M)$ is assumed known in the first algorithm. The solution is the parameter set $\underline{P}$ for which the noise probability density is maximized.

The second technique is more practical

because implementation does not require complete knowledge of the noise statistics. The LMSE solution is the parameter set $\underline{P}$ for which the expectation of the squared error between the original object and the estimated object is minimum.

For the model-based estimation, a method similar to LMSE is chosen. The restoration problem is the previously presented continuous-discrete case. The solution is the optimum parameter set $\underline{P}^*$, such that the expectation of the error between the estimated samples of the noise-free image and the actual detected samples is minimum.

Based on Equation (1), a system of M nonlinear equations is established relating each of the M actual detector responses to its paramtetric value.

$$\begin{cases} S_1 = K \int_{A_1} \left[\int\int_{-\infty}^{\infty} f(\underline{X},\underline{P}) \cdot h(\underline{X},\underline{X}_1) d\underline{X} \right] dx_i = E_1 \\ \vdots \\ S_M = K \int_{A_M} \left[\int\int_{-\infty}^{\infty} f(\underline{X},\underline{P}) \cdot h(\underline{X},\underline{X}_1) d\underline{X} \right] dx_1 = E_M \end{cases} \tag{8}$$

S_1 thru S_M represent the M actual detector responses.

The solution is the parameter set $\underline{P}^*$ for which

$$\left\{ \sum_{i=1}^{M} E_i^2 \right\}^{1/2} \text{ is minimized} \tag{9}$$

The solution to this maximization problem is based on a modification of Levenberg-Marquart algorithm which eliminates the need for explicit derivatives[7].

Application to a Doublet

This section presents the derivation of the integral of Equation (6) for the specific case of a doublet object. Only one-dimensional case is considered. Extension to two-dimensions is straight foward.

Consider two point sources (Figure 1) of amplitudes A and B, and separation distance D. The relative position of the first point source with respect to the leading edge of the sampling detector is referred to as the phase and is denoted by C. The parametric object model is

$$F(\underline{X},\underline{P}) = A\,\delta(x - C) + B\,\delta(x - C - D)$$

where $\underline{x} = x$ (10)

$\underline{P} = [A, B, C, D]^T$

We assume a Gaussian point spread function for the system

$$h(x) = (1/\sqrt{2\pi}\sigma)\exp(-x^2/2\sigma^2) \tag{11}$$

Assuming shift invariance, the convolution integral of Equation (6) becomes

$$g(x_1) = \frac{1}{\sqrt{2\pi\sigma^2}} A \exp \frac{-(x_1 - C)^2}{2\sigma^2} + B \exp \frac{-(x_1 - C - D)^2}{2\sigma^2} \tag{12}$$

For a detector of width Δ and the inter-detector gap of γ (Figure 1) the noise-free value of the j-th sample is

$$\begin{aligned} S_j &= \int_{j(\Delta+\gamma)}^{j(\Delta+\gamma)+\Delta} g(x_1)dx_1 \\ &= A\,\mathrm{erf}\left(\frac{j(\Delta+\gamma)+\Delta-C}{\sigma}\right) - A\,\mathrm{erf}\left(\frac{j(\Delta+\gamma)-C}{\sigma}\right) \\ &\quad + B\,\mathrm{erf}\left(\frac{j(\Delta+\gamma)+\Delta-C-D}{\sigma}\right) - B\,\mathrm{erf}\left(\frac{j(\Delta+\gamma)-C-D}{\sigma}\right) \end{aligned} \tag{13}$$

where

$j = \ldots, -1, 0, 1, \ldots$

Thus, with a minimum of four samples, the unknown parameters can be estimated via the LMSE solution to the nonlinear system of Equation (8).

Simulation Results

The discussed restoration algorithm was simulated on a digital computer and applied to the case of doublet restoration in one-dimension.

Noise-Free Case

The doublet-detector configuration of Figure 1 was chosen. Table 1 presents the selected parameter values for this system. Note that the blur circle is large relative to the size of the detector. Typically, the detector width is equal to the diameter of the 90% power circle, i.e., $\Delta = 4.3\sigma$. Here $\sigma = \Delta$ to ensure that the signal power spreads over more adjacent detectors thereby producing more nonzero samples. Based on Equation (8) a minimum of four distinct detector responses are needed to solve for the four unknown parameters. When the doublet is less than one detector wide, the large blur circle ensures that this minimum number of required samples are generated.

Table 1. Selected Values for the System Parameters

Parameters		Relative Value
Detector width	Δ	
Detector gap	γ	$0.1\,\Delta$
PSF Sigma	σ	$1.0\,\Delta$
First Point Source Intensity	A	1.0
Second Point Source Intensity	B	2.0
Phase	C	$0.125\,\Delta$
Doublet width	D	$1.5\,\Delta$

The detector-doublet configuration of Figure 1 is actually drawn to scale for the specific system whose parameter values appear in Table 1. Note that even before sampling the analog focal plane image, $g(x)$, does not indicate the assumed doublet source.

Figure 2 illustrates the calculated response of the detector array for the specified parameter values. The algorithm attempts to do the calculations in reverse order, that

is, given the detector responses, it traces them back to the unknown parameters A, B, C, and D.

The simulated detector responses were the inputs to the iterative LMSE procedure. The initial guess on the unknown parameter set was arbitrarily chosen as zero (i.e., $P^0 = 0$). The exact unknown parameter values for the given system were generated after 19 iterations. Table 2 indicates the estimated value of the parameters at each iteration. Other sets of parameter values were then selected. For each set, the exact solution was obtained after a certain number of iterations.

Table 2. Estimated Values of the Parameters at Each Iteration

Iteration	A	B	C	D
Initial Values	0.000	0.000	0.000	0.000
	.821	.821	0.000	0.000
	.819	.819	.309	.618
	1.404	1.404	-.338	3.548
	1.388	1.388	-.302	3.454
5	1.336	1.336	-.181	3.135
	1.209	1.209	.085	2.398
	1.391	1.393	.301	1.249
	1.091	1.871	.148	1.130
	1.160	1.773	.174	1.176
10	1.268	1.605	.229	1.229
	1.207	1.707	.306	1.390
	1.199	1.772	.315	1.423
	1.179	1.817	.298	1.429
	1.017	1.982	.176	1.450
15	1.006	1.991	.143	1.487
	1.001	1.998	.128	1.498
	1.000	2.000	.125	1.500
	1.000	2.000	.125	1.500
Final Values	1.000	2.000	.125	1.500

For decreasing separation of the two point sources, the speed of convergence of the algorithm also decreased. The number of required iterations increased exponentially as separation distance D approached zero. Figure 3 presents the plot of the number of iterations versus separation distance. For D = 0, convergence was much faster because the problem reduced to determining only three parameters, (A + B), C and D.

Note that the rate of convergence can be kept constant as D decreases by allowing both the optical system PSF and the detector width Δ to be proportionally reduced. However, in practice, if only σ is allowed to decrease, we may experience undersampling in which case the signal can not be uniquely reconstructed. For the system of Table 1 where $D = 1.5\Delta$, undersampling occurs for $\sigma < 0.2\Delta$. For $D=0.5\Delta$ this occurs at $\sigma < 0.6\Delta$. Because the objective is to resolve the doublet separation distance beyond the detector resultion, the PSF σ is chosen to the detector size.

Simulation In The Presence of Noise

A Monte Carlo simulation was carried out to determine the effectiveness of the estimation process in the presence of noise. The noise is modeled as a white Gaussian process of known standard deviation. The SNR was defined as

$$SNR = \sqrt{\sum_{i=1}^{M} s_i^2} \Big/ \sqrt{\sum_{i=1}^{M} n_i^2}$$

where

M = total number of samples
n = the set of noise values

For each SNR the standard deviation of each estimated parameter was calculated.

Figure 4 displays plots of the standard deviations of the estimated parameters B, and D versus SNR for the system parameters of Table 1. The plots reveal the expected exponential behavior.

In addition to SNR, the precision of the estimation process highly depends on the separation distance D. Figure 5 presents plots of normalized sigma versus SNR for various expected values of doublet separation D. Normalization is with respect to the detector width Δ. As seen, the technique requires a substantially high SNR to resolve a doublet that falls within one detector. On the other hand, when the doublet is more than one detector wide, the procedure can quite accurately resolve it (to within fractions of the detector resolution Δ) at even a low SNR. If we place a limit on the maximum allowed estimation uncertainty σ, then the relationship between the amount of noise and the resulting resolution d can be readily derived from the figure.

Conclusion

It was shown that the model-based approach is an effective restoration technique applicable to coarsely-sampled optical images degraded by diffraction and noise. Simulation results indicated that the knowledge of the PSF, together with the a priori information about the object model, permits object reconstruction from a limited number of image samples. The dominant source of error is the background noise which is assumed to be a zero-mean, uncorrelated Gaussian process. The simulation results indicated that the technique required a high SNR to resolve the doublet when its geometric image is within a single detector cell. On the other hand, when the doublet is more than one detector wide, the procedure can accurately resolve it (to within fractions of the detector resolution) even at a low SNR.

In this paper the restoration procedure was applied to the specific case of a doublet object because it can best describe the resolution enhancement capability of the algorithm. In general, the procedure can be applied to any other object model.

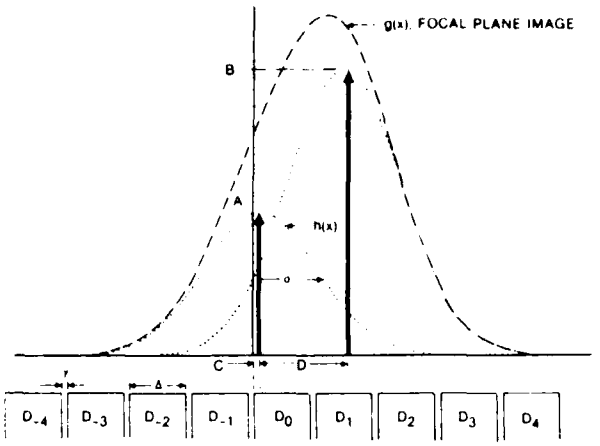


Figure 1 Imaging geometry of the doublet by a one-dimensional detector array

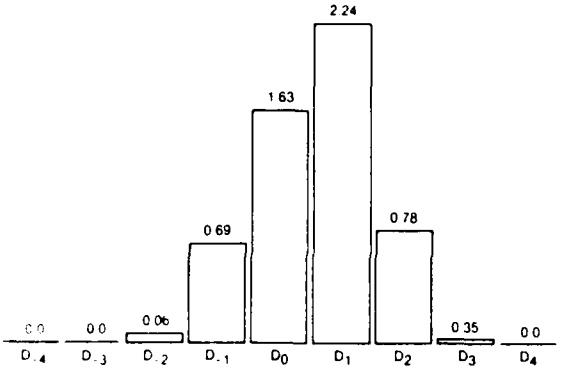


Figure 2 Noise free response of the detector array for the system of Table 1

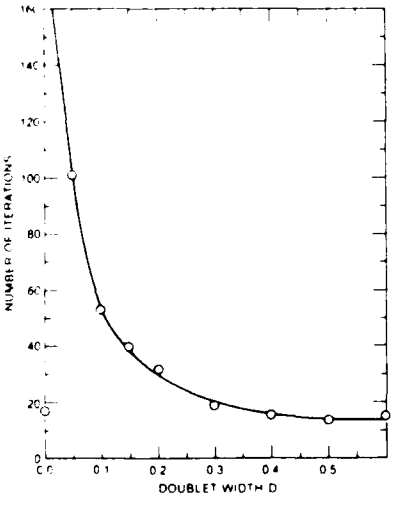


Figure 3 Number of iterations (inversely proporational to the convergence rate) vs doublet width expressed in detector width units Δ

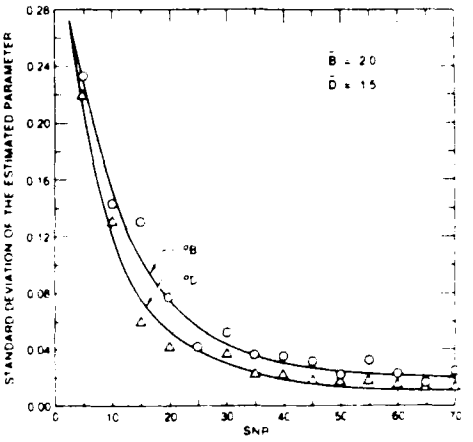


Figure 4 Standard deviation of the estimated parameters B and D vs SNR for mean doublet width of D = 1.5 Δ

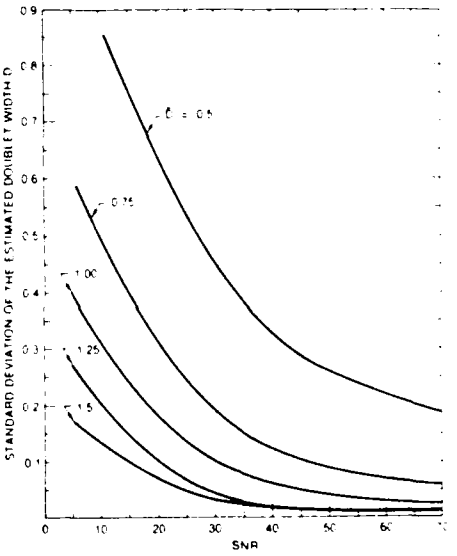


Figure 5 Standard deviation of the estimated doublet separation vs SNR for different mean doublet width D

REFERENCES

1. H. J. Trussel and B. R. Hunt, "Improved Methods of a Maximum A Posteriori Restoration", IEEE Trans. Comput., vol. C-27, No. 1, pp. 57-61, Jan. 1979.

2. B. R. Frieden, "Restoring with Maximum Entropy. III. Poisson Sources and Background", J. Opt. Soc. Am., vol. 68, No. 1, pp. 93-103, Jan. 1978.

3. S. J. Wernecke and L. R. D'Addario, "Maximum Entropy Image Reconstruction", IEEE Trans. on Comput. vol. C-26, No. 4, pp. 351-364, April 1977.

4. B. R. Frieden, "Restoring with Maximum Likelihood and Maximum Entropy", J. Opt. Soc. Am., vol. 62, No. 4, pp. 511-518, April 1972.

5. F. von der Heide, "Least Squares Image Restoration", Optics Communications, vol. 31, pp. 279-284, Dec. 1979.

6. C. M. Lo and A. A. Sawchuk, "Nonlinear Restoration of Filtered Images with Poisson Noise", SPIE, Vol. 207, pp. 84-95, 1979.

7. K. M. Brown and J. E. Dennis, "Derivative Free Analogues of the Levenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares Approximations", Numerische Mathematik, vol. 18, pp. 289-297, 1972.

Spectrum extrapolation on a finite band

F.Gori - Istituto di Fisica-Facoltà di Ingegneria-P.le A.Moro,5 Roma 00185, Italy
S.Wabnitz - Calif.Inst.of Technology-Pasadena, Calif.91125,USA

1. Introduction

The iterative method of Gerchberg (GM) for extrapolating the whole spectrum of a finite support object[1] has been analyzed and generalized by several authors [2-14]. In principle, during each iteration of the GM an infinite band of frequencies should be handled. At first sight, it seems that the (obvious) existence of a cut-off frequency in any practical implementation of the GM should simply allow the spectrum extrapolation to be achieved only below such a frequency. This is not the case, as we shall presently show, in that the extrapolated spectrum obtained by this method up to the cut-off frequency does not coincide with the true spectrum. In this paper we present a modified version of the GM that allows to obtain a spectrum extrapolation on a finite band of frequencies. This is of use both to limit the storage and computation time requirements and to reduce the sensitivity to high frequency noise.

To simplify the exposition, we shall restrict ourselves to the one-dimensional case. The extension to the two-dimensional case is then straightforward.

The following operators are used throughout the paper: a) the band-limiting operator B_j $(j=1,2)$. It truncates the Fourier transform of a function $f(x)$ to the frequency interval $(-\nu_j, \nu_j)$; b) the space-limiting operator T. It truncates a function $f(x)$ to a basic domain $(-a,a)$; c) the identity operator δ which leaves unchanged any function.

2. The Gerchberg method with finite bandwidth constraint.

Let a coherent object $o(x)$ vanishing outside a finite interval $(-a,a)$ be imaged through a space-invariant and aberration-free optical system whose rectangularly shaped pupil extends from $-\nu_1$ to ν_1. The resulting image $i^{(1)}(x)$ can be written

$$i^{(1)}(x) = B_1 To(x), \tag{1}$$

where the previously defined operators are used and where the identity $To(x) \equiv o(x)$ is exploited. The GM for recovering $o(x)$ is based on the following iterative algorithm [1,4]

$$o_N(x) = i^{(1)}(x) + (\delta - B_1) To_{N-1}(x), \qquad (N=1,2,\ldots) \tag{2}$$

with the initial estimate $o_0(x) = i^{(1)}(x)$.

The previous algorithm is implemented by alternate truncations in the space and frequency domains. When in the frequency domain, obvious limitations make δ be substituted by a finite band-limiting operator B_2. The actual algorithm becomes

$$q_N(x) = i^{(1)}(x) + (B_2 - B_1)Tq_{N-1}(x), \tag{3}$$

with the initial estimate $q_0(x) = i^{(1)}(x)$.
One might be led to conclude that the sequence of $q_N(x)$ converges to the image which would be formed through B_2, namely $i^{(2)}(x) = B_2\, To(x)$. However, this is not the case as we shall presently show.

Let us consider the band-pass operator $(B_2 - B_1)T$. This is easily shown to be a positive definite operator. Denoting its eigenfunctions by $\Phi_k(x)$, they satisfy the equation

$$(B_2 - B_1)T\Phi_k(x) = \mu_k \Phi_k(x) \qquad (|x| \leq a),\ k = 0, 1, \ldots) \tag{4}$$

where μ_k are the (strictly positive) eigenvalues. Scarce information is available about the eigenfunctions and the eigenvalues. For our purpose, however, it is sufficient to know that the eigenfunction set is complete in $L^2(-a,a)$ and that the eigenvalues are less than unity. We now expand the truncated image $Ti^{(1)}(x)$ into a series of the $\Phi_k(x)$

$$Ti^{(1)}(x) = \sum_{k=0}^{\infty} \beta_k T\Phi_k(x) \tag{5}$$

By using eqs.(3),(4) and (5) it is easily shown that the truncated iterates tend to

$$Tq_\infty(x) \to \sum_{k=0}^{\infty} \frac{\beta_k}{1-\mu_k} T\Phi_k(x). \tag{6}$$

We want to show that such an asymptotic expression differs from $i^{(2)}(x)$. To this end, let

$$To(x) = \sum_{k=0}^{\infty} \alpha_k T\Phi_k(x) \tag{7}$$

be the expansion of the object into a series of $\Phi_k(x)$. The following expressions can be given for $i^{(1)}(x)$ and $i^{(2)}(x)$

$$i^{(1)}(x) = B_1\, To(x) = \sum_{k=0}^{\infty} \alpha_k B_1 T\Phi_k(x), \tag{8}$$

$$i^{(2)}(x) = B_2\, To(x) = \sum_{k=0}^{\infty} \alpha_k B_2 T\Phi_k(x). \tag{9}$$

We now subtract eq.(8) from eq.(9) and use eq.(4). The result is as follows

$$i^{(2)}(x) - i^{(1)}(x) = \sum_{k=0}^{\infty} \alpha_k \mu_k \Phi_k(x). \tag{10}$$

On using eqs.(10) and (5) we obtain

$$Ti^{(2)}(x) = \sum_{k=0}^{\infty} (\beta_k + \alpha_k \mu_k)\, T\phi_k(x). \qquad (11)$$

Let us compare eqs.(6) and (11). Were the two series identical, the following relations should hold

$$\alpha_k = \frac{\beta_k}{1 - \mu_k}, \qquad (k = 0,1,\ldots) \qquad (12)$$

But this would entail (see eq.(7)) $Tq_{\infty}(x) \equiv To(x)$, and this cannot be true unless $B_2 = \delta$.

We now give a modified GM that allows $i^{(2)}(x)$ to be recovered. To this aim, consider the following simple identities

$$B_2 To(x) = B_2 Ti^{(1)}(x) + B_2 To(x) - B_2 TB_1 To(x), \qquad (13)$$

$$B_1 To(x) = B_1 Ti^{(1)}(x) + B_1 To(x) - B_1 TB_1 To(x). \qquad (14)$$

From eq.(13) we obtain the iterative algorithm

$$p_N(x) = B_2 Ti^{(1)}(x) + p_{N-1} - B_2 Tr_{N-1}(x), \qquad (15)$$

where $p_N(x) = B_2 To_N(x)$ and $r_N(x) = B_1 To_N(x)$. Such an algorithm requires at each step both $B_2 To_{N-1}(x)$ and $B_1 To_{N-1}(x)$. The latter can be also iteratively obtained (see eq.(14))

$$r_N(x) = B_1 Ti^{(1)}(x) + r_{N-1}(x) - B_1 Tr_{N-1}(x). \qquad (16)$$

Convergence of the sequences (15) and (16) follows from the analogous property of the sequence (2) taking into account the continuity of the operators B_1T and B_2T. In conclusion, the required image $i^{(2)}(x)$ is obtained through the combined use of the sequences (15) and (16). It is apparent that the computations in the frequency domain involve only data below the cut-off frequency ν_2.

3. Numerical results

The preceding theory has been numerically tested. A sample of results is given below. A coherent object consituted by two gaussian pulses is imaged through a low-pass pupil. The space-bandwidth product equals one. The truncated, noise-free image $i^{(1)}(x)$ as well as the image $i^{(2)}(x)$ obtained by increasing the pupil width by a factor 1.67 are shown in fig.1. Also shown are the reconstructions $q_{200}(x)$ and $p_{200}(x)$ achieved with the algorithms of eqs.(3) and (15). As can be seen the algorithm of eq.(3) fails to approach the correct image $i^{(2)}(x)$ whereas this goal is attained by means of the modified algorithm. The performance of the latter method in the presence

of noise is illustrated in fig.2. The noisy image, say $\overline{i^{(1)}(x)}$, the (noise-free) image $i^{(2)}(x)$ relating to the enlarged pupil and $p_{200}(x)$ are drawn. The recovery is still satisfactory.

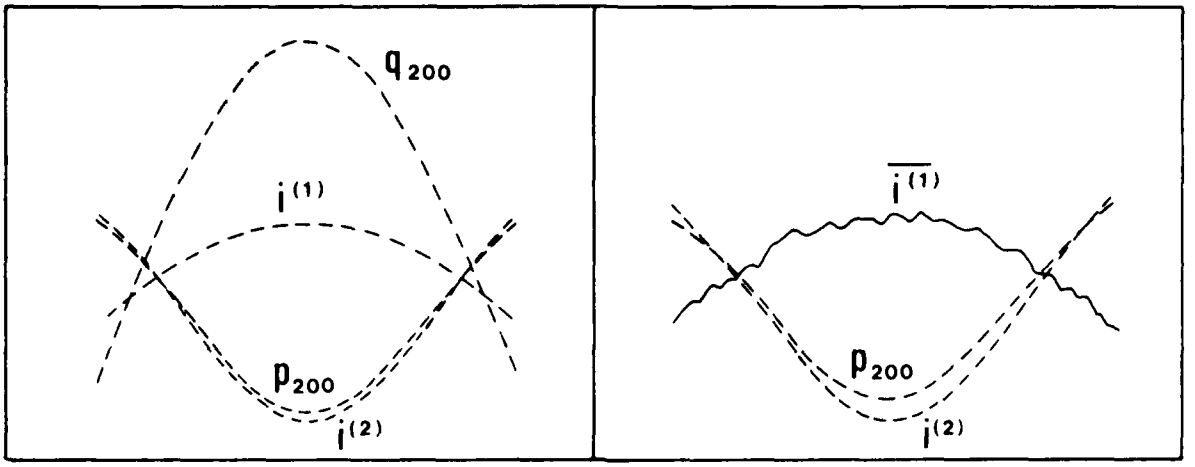


Fig.1

Fig.2

1. R.W.Gerchberg, Opt.Acta, 21, 709 (1974)
2. P.De Santis and F.Gori, Opt.Acta, 22, 691 (1975
3. A.Papoulis, IEEE Trans.Circuits Syst., CAS-22, 735 (1975)
4. F.Gori, IEEE Catalog No.75 CH0941-SC (1975)
5. P. De Santis, F.Gori, G .Guattari and C.Palma, Opt.Acta, 23, 505 (1976)
6. G.Cesini, G.Guattari, G.Lucarini and C.Palma, Opt.Acta, 25,501 (1978)
7. R.J.Marks II, Appl.Optics, 19, 10 (1980)
8. J.B.Abbis, C.De Mol and H.S.Dhadwal, Controller HMSO London (1982)
9. J.A.Cadzow, IEEE Trans.Acoust.Speech, Signal Processing ASSP-27, 4 (1979)
10. H.Maitre, Opt.Acta, 28, 973 (1981)
11. D.C.Youla, IEEE Trans.Circuits Syst., CAS-22, 735 (1978)
12. C.K.Rushfort and R.L.Frost, J.Opt.Soc.Am., 79, 1539 (1980)
13. R.W.Shafer, R.M.Merserau and M.A.Richards, Proc.IEEE, 69, 432 (1981)
14. A.K.Jain and S.Ranganath, IEEE Trans.Acoust.Speech, Signal Processing, ASSP-29, 830 (1981)

On an iterative algorithm for stabilised object restoration from limited spectral data

J. B. Abbiss
Royal Aircraft Establishment, Farnborough, Hants, UK

C. De Mol
Dept of Mathematics, Université Libre de Bruxelles, Belgium

H. S. Dhadwal
Royal Aircraft Establishment, Farnborough, Hants, UK

Summary

We analyse the problem of object restoration in the presence of noise, when the coherent image is formed by a space-invariant system consisting of a one-dimensional clear pupil extending over $(-\Omega, \Omega)$. If the object distribution $f(x)$ lies between $-X$ and $+X$, the noiseless image $\bar{g}(y)$ formed by such a system would be given by the equation

$$\bar{g}(y) = \int_{-X}^{X} \frac{\sin \Omega(x-y)}{\pi(x-y)} f(x)dx \quad , \qquad \text{all } y \, . \qquad (1)$$

In Fourier space, the solution to this equation is equivalent to infinite extrapolation of the truncated spectrum.

The problem of estimating $f(x)$ when the data $\bar{g}(y)$ have been contaminated in some way, for instance by measurement noise, is ill-posed; *ie* the reconstruction will not in general depend continuously on the data[1,2], even if it is unique. In Ref 3 it was shown how the methods of regularisation theory could be combined with an iterative Fourier transform procedure[4,5] to yield a stabilised solution to this problem.

Let us suppose that the effects of the noise can be described by a purely additive function $r(y)$. Then equation (1) becomes, in operational notation,

$$g = B_\Omega D_X f + r \qquad (2)$$

where g is the noisy image and the operators B_Ω and D_X are defined by

$$(B_\Omega h)(x) = \int_{-\infty}^{\infty} \frac{\sin \Omega(x-y)}{\pi(x-y)} h(y)dy \quad , \qquad \text{all } x \ ,$$

$$(D_X h)(x) = \begin{cases} h(x) \ , & |x| \leqslant X \ , \\ 0 \ , & |x| > X \ . \end{cases}$$

Since r is unknown, equation (2) cannot be solved directly. On the other hand, the attempt to determine some function ϕ satisfying the alternative equation

$$g = B_\Omega D_X \phi$$

will not in general yield a meaningful solution, and may fail completely. However, by applying the results of regularisation theory, which makes use of known object constraints, we are able to reformulate the problem, and modify equation (2) in such a way that the solution is now constrained to be 'close to' the true object in some (well-defined) sense and to respond stably to variations in the noise level. In Ref 3 a regularised approximation to the true object obtained in this way was shown to be the solution $\tilde{f}$ of the equation

$$(D_X B_\Omega D_X + \alpha D_X)\tilde{f} = D_X B_\Omega g \qquad (3)$$

where α is the regularisation parameter. From this equation a procedure for calculating $\tilde{f}$ by iteration can be derived, the Nth estimate being

$$\tilde{f}_N = D_X B_\Omega g + \left\{(1-\alpha)D_X - D_X B_\Omega D_X\right\}\tilde{f}_{N-1} \quad . \qquad (4)$$

As in the case of the original unregularised algorithm, the behaviour of the iterative solution can be analysed by means of the prolate spheroidal wave functions[5,6]. We carry out this analysis, and also derive explicit error bounds for the regularised estimate of the object as a function of the number of iterations N and of the regularisation parameter α .

References

1 A.N. Tikhonov, V.Y. Arsenin — *Solutions of ill-posed problems.* V.H. Winston & Sons, Washington DC (1977)

2 M. Bertero, C. De Mol, G.A. Viano — In *Inverse scattering problems in optics.* Topics in current Physics, Vol.20, edited by H.P. Baltes, Springer-Verlag, Berlin (1980)

3 J.B. Abbiss, C. De Mol, H.S. Dhadwal — Regularised iterative and non-iterative procedures for object restoration from experimental data. To be published in Optica Acta

4 R.W. Gerchberg — Optica Acta, 21, 9, 709 (1974)

5 A. Papoulis — IEEE Trans. Circuits and Systems, CAS-22, 9, 735 (1975)

6 P. De Santis, F. Gori — Optica Acta, 22, 8, 691 (1975)

SESSION IV

SPECTRAL ESTIMATION AND OTHER APPLICATIONS

C. Rushforth, *Presider*

Estimation of Two Closely Spaced Frequencies Buried in White Noise Using Linear Programming

Jaroslav Keybl and George Eichmann
Department of Electrical Engineering
The City College of the City University of New York
New York, N. Y. 10031

ABSTRACT

Linear programming is used to estimate the spectrum of two sinusoids signals closely-spaced in frequency buried in deep white gaussian noise by employing a-priori knowledge of the spectrum. The method will be illustrated by a number of examples.

INTRODUCTION

In the calculation of the spectrum of a discrete-time signal consisting of two sinusoids with closely-spaced frequencies embedded in white gaussian noise, problems arise when only a small portion of the signal is available. This is known as windowing of the data and becomes evident as "leakage" in the spectral domain, i.e. energy in the main lobe of a spectral response "leaks" into the sidelobes obscuring other present spectral responses. There are many methods used to estimate the spectrum. The periodogram [1] performance is poor for short data lengths. The Blackman-Tukey method [2] is also hampered by spectral distortion. The Burg method [3], a high frequency-resolution technique, even in the absence of noise, yields spectral line splitting. The recent algorithm of Cadzow [4] outperforms the Burg method in a low noise environment.

In this paper, we estimate the spectrum of two sinusoidal signals closely-spaced in frequency by employing a-priori knowledge of the Fourier spectrum of the signal in the form of linear inequalities. The advantage of imposing constraints in spectral restoration process has been pointed out [5]. In the linear programming formulation, there are many solutions. Here, we select a solution which minimizes the l_1 norm of the Discrete Fourier transform(DFT) of the measured signal and its estimate.

DETERMINATION OF THE SPECTRUM

We assume that the signal estimate s(k) can be represented by a weighted sum of past signals

$$s(k) = -\sum_{i=1}^{p} a(i)s(k - i) + e(k) \qquad k = 1,2,\ldots,m \qquad (1)$$

where a(i) are unknown weighting coefficients and e(k) is an error term. This expression is a linear predictor[6]. By taking

the DFT of Eq.(1), we have

$$S(n) = -\sum_{i=1}^{p} a(i)\exp\{-j2\pi in/N\}S(n) + E(n) \quad 1 \le n \le N \qquad (2)$$

where S(n) and E(n) are the DFT of s(k) and e(k), respectively. Because S(n) and E(n) are complex, and using Euler's formula, we rewrite Eq.(2) as

$$S_R(n) = -\sum_{i=1}^{p} a(i)[S_R(n)\cos v + S_I(n)\sin v] + E_R(n) \qquad (3a)$$

$$S_I(n) = -\sum_{i=1}^{p} a(i)[S_I(n)\cos v - S_R(n)\sin v] + E_I(n) \qquad (3b)$$

where $v = 2\pi in/N$. We have 2N equations with p unknowns. When the signal is real, we can reduce it to N equations with p unknowns. Assuming that two sinusoidal signals are present, it can be shown, the a(i) coefficients are the pole coefficients of the z-transform of two sinusoidal signals. Therefore, p = 4 and the range of the a(i)'s are

$$-4 \le a(1) \le 4, \quad -2 \le a(2) \le 6, \quad a(3) = a(1), \quad a(4) = 1 \qquad (4)$$

Eq.(4) adds 6 more equations, to the N equation generated from Eq.(3), for a total of N + 6 equations used in the linear programming formulation. From Eq.(3)and (4), we can now solve for the a(i) coefficients by minimizing the l_1 norm of the error - E(n).

NUMERICAL RESULTS

To test this method we have used the time series

$$s(k) = A_1\cos(2\pi f_1 k) + A_2\cos(2\pi f_2 k + \phi) + w(k) \qquad (5)$$

with $1 \le k \le N$ and w(n) as white Gaussian noise with zero mean and variance σ^2. The two sinusoidal frequencies are normalized so that f = 0.5 corresponds to the Nyquist rate. The individual sinusoidal signal-to-noise ratio's (SNR) is given by $20\log (A_k/\sqrt{2}\sigma^2)$ for k = 1,2. For all of our examples, we chose the signal amplitudes $A_1 = \sqrt{2}$, $A_2 = \sqrt{20}$ and the signal frequencies f_1 = 0.2168 and f_2 = 0.2265. In two examples we introduced a forty-five degree phase difference between the two sinusoids.

In Fig. 1. the variance of of the noise σ^2 = 0.94 and M = 192. This corresponds to 0.54 db SNR on the weaker signal and an time-bandwidth product (TBP) of 1.50. The spectrum calculated using the periodogram shows random fluctuations and it resolves both frequencies. Our method yields peaks at frequencies f_1 = 0.2168 and f_2 = 0.2265 which shows the ability to resolve the frequencies with a small error in such a low SNR environment without fluctuation. It should be pointed out that the reason both peaks are of equal amplitude is because we are calculating the poles which make the function blow up. In Fig. 2, we introduce a forty-five degree phase shift between the

sinusoids. Here the SNR is - 0.26 db. We see that there is no line splitting as is often the case with other algorithms. The peaks occur at f_1 = 0.2207 and f_2 = 0.2265, again resolving the frequencies in a low SNR environment. Fig. 3 shows the estimated results when the SNR is - 14.2 db. The frequencies f_1 = 0.2188 and f_2 = 0.2305 yield good estimates for a very low SNR. In Fig.4, we added a forty-five degree phase shift with a SNR - 15 db. We see that the periodogram is unable to resolve the two frequencies but our method peaked at f_1 = 0.2188 and f_2 = 0.2324. Again we note that there is no line splitting. In the next set of experiments,the TBP is reduced. In Fig.5, the SNR is 0 db and the TBP is 1.00. The periodogram shows peaks while our method gives the frequency estimates f_1 = 0.2168 and f_2 = 0.2305 showing that we get good quality estimates even when the number of samples is reduced. Finally, in Fig.6, we plot the spectrum for a SNR of - 14.5 db. The estimated frequencies are f_1 = 0.2168 and f_2 = 0.2363. The larger frequency error can be attributed to the very low SNR and the small TBP environment.

SUMMARY

We have shown a new method for determining the spectrum of two sinusoidal signals closely-spaced in frequency, where the weaker signal has a very low SNR. We show that the method is not affected by spectral line splitting when the two sinusoids have a phase difference of 45 degres. Computer generated results of this method have been presented.

ACKNOWLEDGEMENT

This work is supported in part by a grant from the Air Force Office Scientific Research AFOSR 81-0169 and a contract from the Rome Air Development Center F19628-80-C-0095

REFERENCES

[1] S.M. Kay and S.L. Marple,Jr., "Spectrum Analysis - A Modern Perspective", Proc. IEEE, vol.69,pp. 1380-1419, Nov.1981.

[2] R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra From the Point of View of Communications Engineering (New York:Dover 1959)

[3] A. Papoulis, "Maximum Entropy and Spectral Estimation: A Review," IEEE Trans. Acoust. Speech Signal Proc., vol.ASSP-29,pp. 1176-1186,1981

[4] J.A. Cadzow, "High Performance. Spectral Estimation - A New ARMA Method", IEEE Trans. Acoust.Speech Signal Proc., vol.ASSP-28, 524-529, Oct. 1980.

[5] R.J. Mammone and G. Eichmann, "Restoration of the discrete Fourier spectra using linear programming", J. Opt. Soc. America, vol.72, No.8, August 1982

[6] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc.IEEE, vol. 63, pp. 561-580, April 1975.

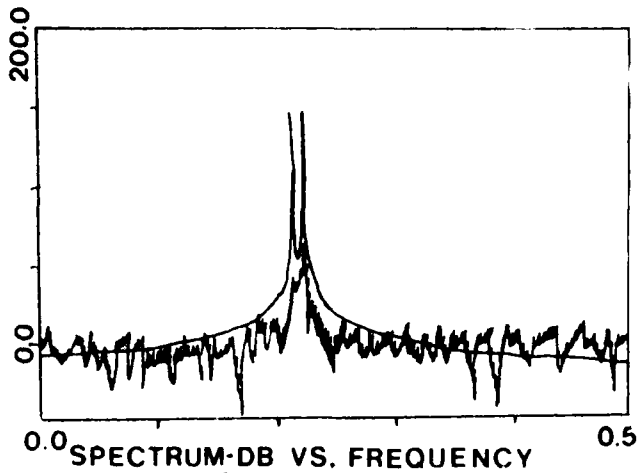


Fig. 1. Spectrum using new method and periodogram $\sigma^2=0.94$ m=192

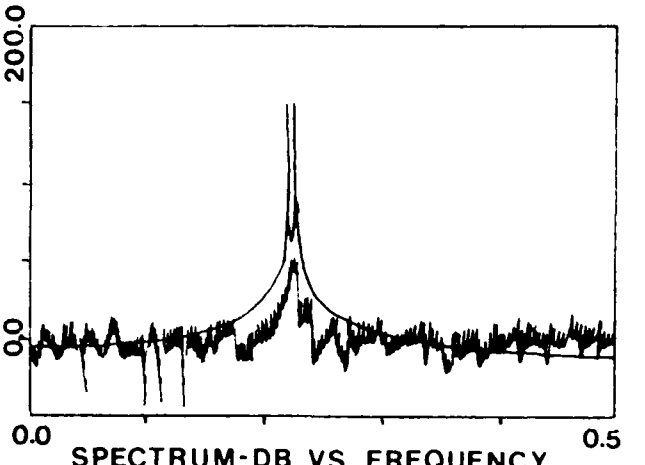


Fig. 2. Spectrum using new method and periodogram $\sigma^2=1.03$ m=192 $\phi=\pi/4$

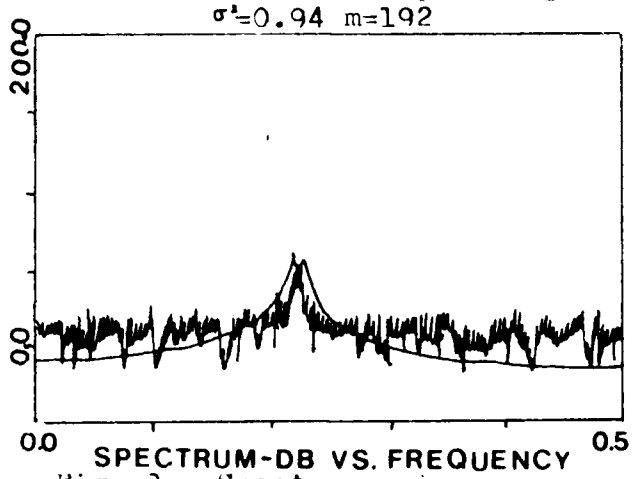


Fig. 3. Spectrum using new method and periodogram $\sigma^2=5.12$ m=192

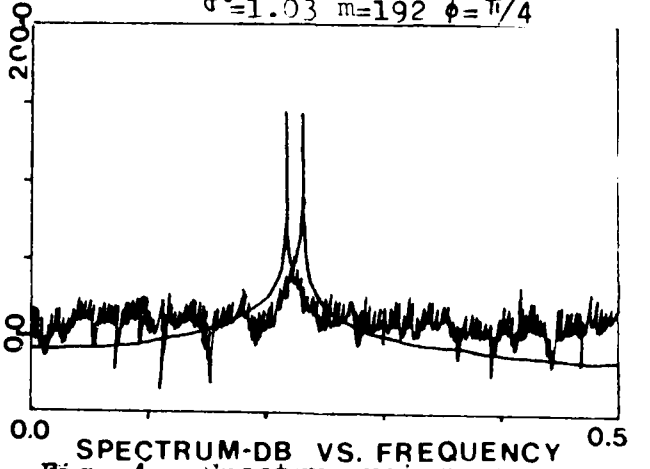


Fig. 4. Spectrum using new method and periodogram $\sigma^2=5.60$ m=192 $\phi=\pi/4$

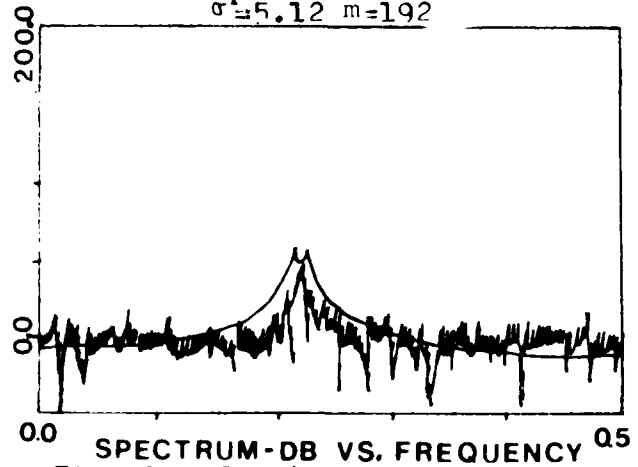


Fig. 5. Spectrum using new method and periodogram $\sigma^2=1.00$ m=128

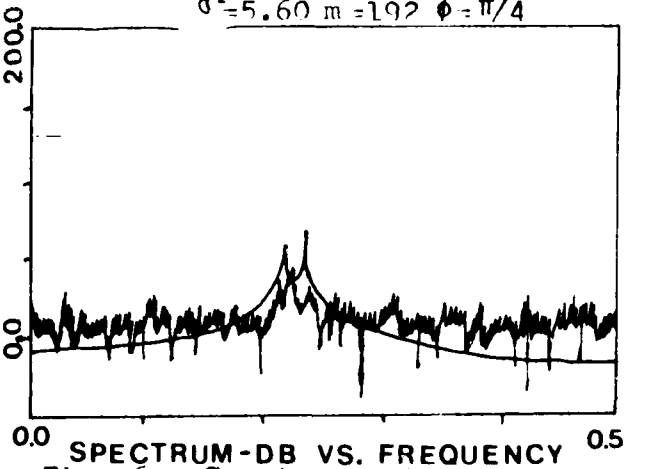


Fig. 6. Spectrum using new method and periodogram $\sigma^2=5.27$ m=128 $\phi=\pi/4$

A NOVEL HANKEL APPROXIMATION METHOD FOR ARMA POLE-ZERO ESTIMATION FROM NOISY COVARIANCE DATA*

S.Y. Kung and K.S. Arun

Dept. of Electrical Engineering-Systems,
Univ. of Southern California,
Los Angeles, CA 90089-0272

1. Introduction

Model based methods have been gaining popularity in high resolution spectral estimation, and have recently demonstrated a great deal of success. Such methods allow us to parameterize the spectrum in terms of a relatively small number of unknown parameters, and thus reduce the spectral estimation problem to that of first, selecting the appropriate model, and second, estimating its parameters. The most popular models used today, are

1) Autoregressive model (AR),
2) Sinusoids plus noise model (S+N) and
3) Autoregressive moving average model (ARMA)

The problem addressed in this paper is that of estimating the parameters for the above models from a finite length of the covariance sequence of the stochastic process. Given only the raw data (i.e. the time series itself), one could first estimate the covariance sequence before estimating the model parameters. We shall treat covariance estimation errors as perturbations in the covariance sequence. There exist other methods that estimate the model parameters directly from the time series, without the covariance estimation step, but in this paper, we are only concerned with parameter estimation from the covariance sequence.

The Maximum Entropy Method[JPB] for AR models and Pisarenko's method[PIS] for S+N models, both work very well when the covariance lags are known exactly. But, MEM cannot account for additive noise, and suggested modifications[KAY],[ERT] do not work well when the noise is colored. At the same time, Pisarenko's method is not meant for colored noise, and it turns out that both these methods are very sensitive to any perturbations in the covariance data[HC].

The only way out of these rather restrictive limitations appears to lie in covariance approximation methods [BS],[SK2]. Toeplitz approximation seems to work very well for the S+N model, and many versions have been

*This research was supported in part by the Office of Naval Research under contract no. N00014-81-K-0191and by the Army Research Office under grant no. DAAG 29-79-C-0054.

proposed[SW2],[CYB],[KT]. In this paper, we present a Hankel approximation method for ARMA models, that is based on a principal component analysis on the Hankel matrix formed from the covariance sequence. Both the AR and S+N models are special cases of the ARMA model, hence this new algorithm provides a general approximation method for all the three models.

Since we are interested only in finding a rational approximation to the given covariance sequence (not the covariance matrix), it is not immediately obvious that we should try to approximate a Toeplitz matrix. Instead, why not use the Hankel matrix? Hankel approximation is a popular approach to linear system model order reduction[AAK],[SK1],[PL], and the experience gained there may be put to use in the covariance approximation problem. In fact, it can be shown (c.f. Sect.2) that the rank of the infinite Hankel matrix formed from the covariance sequence, equals the order of the generating model. Singular value decomposition is a very good way of examining how close the Hankel is to a low rank matrix. By dropping the smaller singular values, and taking only the principal components, one gets a low rank approximation to the Hankel. The approximant itself need not be Hankel, yet one can find a least squares estimate [SK1] of the corresponding state space triple (F,g,h) whose order is equal to the rank of the approximant. The eigenvalues of F are, then, the poles of the model.

This method can thus be used for identifying the pole locations of both AR and ARMA models, and is well suited to the frequency estimation of harmonic processes. Being a covariance approximation technique, one can expect it to be robust with respect to covariance perturbations, either due to the presence of colored noise in the signal, or due to covariance estimation errors. This method is thus, a robust generalization of the Levinson(MEM) and Pisarenko methods, and is applicable to colored noise environments. For ARMA models, the zeros can be estimated by the use of an algorithm similar to the minimal stochastic realization algorithm due to Faurre[FRE].

2. Theory and development

The problem at hand is to find a rational covariance sequence that is closest in some sense, to the given covariance sequence. A covariance sequence is rational if it can be exactly modelled as the covariance of the output of an ARMA system to white noise input.

Assume that a rational covariance sequence $\{r(k)\}$ corresponds exactly to an ARMA model whose transfer function is $H(z) = b(z)/a(z)$. Then, the two sided z-transform $S(z)$ of $\{r(k)\}$ will be equal to

$$S(z) = H(z).H(z^{-1}) = \frac{b(z).b(z^{-1})}{a(z).a(z^{-1})}$$

But, if we take only the causal part of $\{r(k)\}$, and define the one sided z-transform

$$R(z) = x(z)/y(z) = \frac{1}{2}.r(0) + \sum_{k>0} r(k).z^{-k},$$

then, $S(z) = R(z) + R(z^{-1})$

$$\Rightarrow \frac{b(z).b(z^{-1})}{a(z).a(z^{-1})} = \frac{x(z).y(z^{-1}) + x(z^{-1}).y(z)}{y(z).y(z^{-1})}$$

Therefore, $a(z) = y(z)$ and

$$b(z).b(z^{-1}) = x(z).y(z^{-1}) + x(z^{-1}).y(z) \qquad (1)$$

Thus, the causal part of $\{r(k)\}$ is the impulse response of a modified ARMA model $x(z)/a(z)$ that has the same poles as the original ARMA model $b(z)/a(z)$. Since, both models have the same poles, we can find the system poles by treating the causal part of $\{r(k)\}$ as a deterministic impulse response sequence, and using the principal Hankel component (PHC) algorithm[SK1]. Kronecker's theorem tells us that the rank of the infinite Hankel matrix formed from the causal part of the covariance sequence is equal to the order of the system $x(z)/a(z)$. Thus, if the given covariance sequence is close to a rational covariance, singular value decomposition on H should help us break up H into principal components and perturbation components. Selecting only the principal components will lead to a rational approximant, which need not be Hankel. Now, using the PHC factorization algorithm, one can obtain the state space parameters (F,g,h) for $x(z)/a(z)$. The steps are:

1) Do an SVD on the Hankel matrix $H=U.\Theta.V^T$
2) Select the dominant components only, and factorize $H=\mathcal{O}.\mathcal{C}$ where $\mathcal{O}=U.\Theta^{0.5}$ and $\mathcal{C}=\Theta^{0.5}.V^T$ are the observability and controllability matrices.
3) Now evaluate the state space parameters
$F = \bar{\mathcal{C}}.\mathcal{C}^{+}$
g = 1st column of $\mathcal{C}$
h = 1st row of $\mathcal{O}$.

Since (F,g,h) is the state space triple for $x(z)/a(z)$, the eigenvalues of F are the poles of $x(z)/a(z)$ and consequently, the desired poles of the ARMA model $b(z)/a(z)$. If the given covariance sequence is exact and truly rational, the algorithm should give us the system poles exactly. If the covariance sequence is perturbed, then the approximation step in the PHC algorithm gives us a rational approximant, that is close to the given covariance sequence. The PHC algorithm is relatively insensitive to perturbations in the Hankel, and so, we obtain a good estimate of (F,g,h) and the poles of the ARMA model.

The zeros of (F,g,h) will give us the roots of $x(z)$, and not the desired zeros of $b(z)/a(z)$. Yet, equation(1) tells us that $x(z)$ and $a(z)$ contain information about $b(z)$ as well, and so, it should be possible to evaluate the zeros of the ARMA, from (F,g,h). The triple (F,g,h) now, corresponds to a rational covariance sequence $r(k) = h.F^{k-1}.g$, and it has been shown[AKE],[FRE] that every rational covariance sequence can be written as the covariance of a time series $\{y(t)\}$ with the Markovian representation:

$$x(t+1) = F.x(t) + T.v(t)$$
$$y(t) = h.x(t) + v(t)$$

where $v(t)$ is the driving white noise input, and there exists column vector g such that $r(k) = h.F^{k-1}.g$.

Since the PHC algorithm gives us (F,g,h) satisfying these conditions, we now only need to find T, to obtain the state space parameters (F,T,h,1) of the ARMA model b(z)/a(z). Faurre's minimal stochastic realization algorithm[FRE] gives us a way of determining T and the input noise power ρ. This algorithm requires us to solve the algebraic Riccati equation

$$P = F.P.F^T + (g-F.P.h^T).(r(0)-h.P.h^T)^{-1}.(g-F.P.h^T)^T$$

to find the state variance P.

$$\text{Then, } \rho = r(0) - h.P.h^T$$
$$\text{and } T = (g - F.P.h^T)/\rho$$

Then, the ARMA zeros are the eigenvalues of (F - T.h), since the inverse system a(z)/b(z) has the state space parameters (F-T.h,T,-h,1).

3. Results

For illustrating the power of this method, simulations were conducted on two sinusoidal signals in additive colored AR(1) noise at 0dB. The amplitudes of both sinusoids was set at 1 and the noise had covariance $r(k)=\alpha^{abs(k)}$. The sinusoid frequencies were selected to be 0.25π and 0.30π, while the Markov parameter for the noise was picked to be $\alpha=0.8$. The size of the Hankel matrix used was 10x10. The results are summarized below.

Covariance sequence used	Estimated poles
True covariance	$1.00000\exp\{\pm j0.25000\pi\}$
	$1.00000\exp\{\pm j0.30000\pi\}$
	$0.80000\exp\{ j0.00000\pi\}$
True covariance + uniformly distributed zero mean random perturbations of variance 0.0001	$0.99829\exp\{\pm j0.25207\pi\}$
	$0.99813\exp\{\pm j 0.30616\pi\}$
	$0.80116\exp\{ j0.00000\pi\}$
Unbiased covariance estimate from 1024 sample data.	$0.99163\exp\{\pm j0.24299\pi\}$
	$0.98841\exp\{\pm j0.30467\pi\}$
	$0.76021\exp\{ j0.00000\pi\}$

When the covariance lags are known exactly, the poles and zeros of the ARMA model are identified exactly by the new Hankel approximation method. In the presence of white noise, the stochastic process can be modelled by a new ARMA model that has the same poles as the original ARMA model. Thus, pole identification of AR and ARMA models is totally unaffected by white noise, and the method performs excellently for frequency estimation of sinusoids in white noise.

But the more salient advantage lies in the fact that SVD is a robust and numerically stable approximation algorithm, and it makes this method insensitive to covariance perturbations, including the effects of colored noise and covariance estimation errors. Thus, this method provides a robust way of generalizing both the Levinson and Pisarenko methods to colored noise and other forms of covariance perturbations, that are not necessarily positive semi definite. It also provides a way of separating narrow band signals from

broadband signals, on the basis of the distance of the pole from the unit circle.

The second step in this method, that uses Faurre's minimal stochastic realization algorithm for estimating the zeros, contains no aproximation. It appears that covariance approximation is carried out mainly in determining the pole locations. Hence, we are currently looking into alternate schemes for computing the zeros in a robust fashion.

4. References

[AAK] V.V.Adamjon, D.Z.Arov and M.G.Krein, "Analytic properties of Schmidt pairs for a Hankel operator, and the generalized Schur-Takagi problem", Math. USSR Sbornik, vol.15, no.1, 1971

[AKE] H.Akaike, "Stochastic theory of minimal realization", IEEE T.AC-19, Apr. 1974

[BS] A.A.Beex and L.L.Scharf, "Covariance sequence approximation for parametric spectrum modelling", IEEE T.ASSP-29 no.5, Oct. 81

[CYB] G.Cybenko, "Moment problems and low rank Toeplitz approximations", Intl. Symp. on rational approximation of systems, Katholieke Univ. Leuven, Aug.31-Sep.1, 1981

[FRE] P.Faurre, "Stochastic realization algorithms" in System Identification: Advances and case studies, Ed. R.K.Mehra and D.G.Lainiotis, Academic Press, 1976

[FRT] O.L.Frost III, "Resolution improvement in AR spectral estimation", EASCON, 1977

[JPB] J.P.Burg, "Maximum entropy spectral analysis", Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975

[KAY] S.M.Kay, "Noise compensation for AR spectral estimates", IEEE T.ASSP-28, no.3, Jun.1980

[KL] S.Y.Kung and D.W.Lin, "Recent progress in linear system model reduction via Hankel matrix aproximation", ECCTD, The Hague, Netherlands, Aug. 1981

[KT] R.Kumaresan and D.W.Tufts, "Data adaptive principal component signal processing", IEEE Conf. on Decision and Control Alburquerque, NM, Dec.1980

[PIS] V.F.Pisarenko, "The retrieval of harmonics from a covariance function", Geophys. J. Royal Astron. Soc. 33, 1973

[SK1] S.Y.Kung,"A new identification and model reduction algorithm via singular value decomposition", 12th Asilomar Conf. on Circuits, Syst. and Comp., Pacific Grove, CA, Nov 1978

[SK2] S.Y.Kung, "A Toeplitz approximation method and some applications", Intl. Symp. on Math. theory of netw. and syst., Santa Monica, CA, Aug. 1981

[UC] T.J.Ulrych and R.W.Clayton, "Time series modelling an max entropy", Physics of the interplanetary interiors, 12, 1976

Some Signal Processing Issues in Radar Target Identification

E. K. Miller
Lawrence Livermore National Laboratory
P. O. Box 5504, L-156
Livermore, CA 94550

Introduction

When an object (target) is illuminated by an electromagnetic field, the energy it scatters to a given observation point depends upon its geometry and material characteristics, the conditions of illumination (angle of incidence, polarization, frequency(ies), etc.) and the medium through which the field propagates. It is intuitively obvious, and demonstrated by analysis and measurement, that the scattered field (which will also contain, in general, interference and/or noise energy) has impressed or encoded on it information about the target which might somehow be extracted for purposes of: 1) detection (i.e., is there a target?); 2) classification (determine whether the target belongs to a class of interest); 3) recognition (if in the class of interest, which one of that class is it?) and 4) imaging or inversion (reconstruct the target's geometry and/or material properties).*

The quality and amount of target-scattered energy that are available place limits on how much information an observer can extract solely from measurement. Rarely, however, is the observer dependent wholly on the measurement alone to accomplish the goal of target identification, but can, in addition, make use of a priori information (e.g., atmospheric noise data, knowledge that airborne targets are planes or helicopters) and/or modify the measurement itself (change frequencies, polarization, pulse shape, etc.). These factors suggest that the observer should process his data in such a way as to fully utilize all a priori information, and to the extent possible, adapt the controllable measurement conditions to the characteristics of the target. In both cases, knowledge of at least some of the characteristics of expected targets is an essential requirement for improving the signal processing.

Target Characterization

Target characteristics may be recognized to belong to (at least) two domains. One, the target domain, is that in which the target is a basic entity having either directly measurable properties (such as mass, volume, shape) or properties which are obtainable from those measurements (moment of inertia and higher-order moments, harmonic expansion of the surface profile), either of which can be used to uniquely describe the target to some degree of accuracy. The other, which we term the data domain, is that in which the target exists only indirectly, in terms of its effect upon a measured observable (a scattered EM field, for example), and from which various properties may be derived. When these derived, or data domain,

*Unless a specific item from this list is being discussed, they will be lumped together in the single term "target identification."

properties can be uniquely related to the basic, or target domain, properties, then it can be said that an inverse problem has been solved. For many applications, however, classification or recognition alone may suffice or even be all that is possible to achieve, in which case the data-domain properties may be all that is needed. In the former application, the problem is more absolute in nature, i.e., (determine the target) whereas in the latter, it is more comparative or relative (i.e., differentiate between targets). In either case, the way in which target-domain and data-domain properties are related is crucial to a successful outcome.

Another key to characterizing a target is the number of independent properties that are needed either to describe it to some specified degree of completeness or to separate it from other targets of interest. This number, which we may call the target rank, can depend on the characterization being used and availability of *a priori* information. For example, if all targets are known to be penetrable, homogeneous spheres, then the only basic properties of concern are radius and composition. We might also expect that a desirable property of the characterization is that it be parsimonious, i.e., with minimum rank that it include all independent properties over some set of measurements. This statement implies that the characterization in either the target domain or data domain is probably not unique, and that alternative property sets are available.

Besides the issues of which target or data properties may be best, and what their associated ranks may be, there is the closely related question of the information content of the measurement. This might be loosely defined to be the total number of bits required to represent the maximum number of independent properties (basic or derived) obtainable from the data. Implicit in this definition is a tradeoff between the number of properties and their complexity, as we speculate that the total information represented by the data is fixed. A priori knowledge, however, can greatly influence how the data are best used, changing the problem from one of system identification (determining a model for the data) to parameter estimation (given a model, estimating the model parameters). In other words, by reducing the number of unknowns via a priori knowledge (a model), the data can yield either more complete or more accurate information about the model parameters (the properties).

Some EM Feature Sets

In radar terminology, the EM properties of a target are sometimes called its features, and the ad hoc nature of the features that have at one time or other been used is demonstrated by some examples which include: the maximum value of target cross section at a given frequency; the aspect-averaged cross section over a specified frequency band; and the ratio of cross sections for two orthogonal polarizations of the incident field. But target features such as these can vary due to the aspect, frequency and polarization dependence of its scattering properties. There is a need for a feature set that is a fundamental property of the target and is independent of how the target is excited.

An example of such a feature set is the complex source-free resonances, or poles, popularized in recent years in connection with the Singularity Expansion Method. These frequency-domain poles are dependent only on the target geometry (and impedance, if it is imperfectly conducting) with the oscillatory or imaginary parts established by the length(s) of propagation path(s) on the object, and the damping or real parts associated with energy loss (radiative and dissipative) over that path. While significant analytical questions remain (higher-order layer poles, branch cuts, singularities at infinity, etc.), from a pragmatic viewpoint the poles seem to be unique properties of a target and, therefore, provide an eminently attractive mechanism for target identification. In some sense, too, they seem to represent a kind of minimum-parameter target description which has interesting implications for the inverse problem from an information theoretic viewpoint. Finally, they may provide useful insight concerning other feature sets that have been or might be used for target identification.

Considering the use of frequency-domain poles for target recognition presumes the availability of wide-band or impulsive scattering data for a target. An alternate approach is to make the measurement as a function of angle in a bistatic or monostatic scattering mode at a fixed frequency. In this case, a different kind of pole (or more properly a singularity) arises due to the equivalent or actual point sources which give the scattered field. Such sources can correspond to scattering centers and points of radiation from the target, and thereby might also serve as a feature set. To the extent that they are related to the target geometry, these space singularities could provide a more direct link between the target and data domains that do the complex frequency-domain poles. For example, the maximum physical extent of the target might be obtained directly from extrema of the space-singularity distribution, but could only be inferred from the wavelength of the smallest-value frequency-domain pole.

Space singularities can also be associated with measurements made as a function of position (e.g., along a line) or as a function of frequency. In the former situation, the singularities correspond to the direction angles of the plane-waves which make up the total field. The result is that a spatial array can image a distribution of sources located on the far-field sphere. An analogous situation arises in measuring the field of a set of point sources as a function of frequency, but at a fixed point in space, where the singularities then correspond to the source positions. By making measurements from three orthogonal (or linearly independent) directions, the possibility arises of imaging a three-dimensional source distribution.

Signal Processing for EM Features

The various data types mentioned above have a common exponential form, given generically by

$$F(x) = \Sigma\, R_\alpha\, e^{X_\alpha x} \quad \frac{[\mathscr{L}]_1}{[\mathscr{L}]^{-1}} \quad F(X) = \Sigma\, R_\alpha/(X-X_\alpha) \qquad (1)$$

where we refer to these as the waveform and spectral representations, respectively. For the cases discussed above, it follows that:

Representation	$x =$	$X =$	$X_\alpha \sim$
Time/Frequency	t	$i\omega$	Complex resonance.
Frequency/Space	ω/c	$i\hat{k}.\bar{R}$	Source location along a line.
Space/Angle	x	$ik \cos \theta$	Direction cosine of plane wave.
Angle/Space	$\cos \theta$	ikd	Source position along a line.

An interesting property of these representations is their close connection with widely used signal-processing techniques. For example, the linear-predictor filter implicitly incorporates an exponential model, as can be demonstrated by applying Prony's Method to a set of uniformly spaced (in x) samples of f(x). Prony's Method and related techniques provide a way to estimate values of X_α and R_α from the f(x) samples. Thus, there is the possibility of not only obtaining parameters which may be useful EM feature sets, but also that of processing EM data for target recognition.

In this paper, we will consider three issues relating to the above discussion:

1) The performance characteristics of Prony-like signal processing for estimating EM features, from the viewpoint of an input-output information transformation;
2) Pole (and singularity) sets of EM features obtained from computed EM data for several situations;
3) Some initial results of using frequency-domain poles for target identification.

It will be demonstrated for data similar to that which occurs in sampling EM fields, that Prony's Method seems to preserve information, that is give poles values whose accuracy on the average equals the accuracy of the input data. However, the imaginary (oscillation) components are almost always found with greater accuracy than the real (decay) components, whose accuracy seems more closely correlated to that of the predictor coefficients. The richness of the pole-based representation for EM problems will be shown by giving sample results from transient data, linear-array patterns, and plane wave scattering from a half space. Finally, various target-identification schemes using frequency-domain poles will be compared as a function of signal-to-noise level, with the tentative finding that improved performance seems to occur at the expense of increased computational effort.

Restoration of Multichannel Microwave Imagery To Estimate Rainfall Rates in Hurricanes

R. T. Chin, and C. L. Yeh and W. S. Olson, and J. A. Weinman
Dept. of Elect. and Comp. Engr. Dept. of Meteorology

University of Wisconsin-Madison
Madison, Wisconsin 53706

I. Background

Multichannel microwave radiometers on the Seasat and Nimbus 7 satellites offer a quantitative method to measure rainfall amounts over the ocean. The emissivity of the ocean surface is low and varies predictably with wind speed; it thus provides a good background for observing precipitation. The theory and initial validation of this concept was given by Wilheit et al. [1]

Recently Olson [2] employed a radiative transfer model to simulate the polarized brighteness temperatures that a Scanning Multichannel Microwave Radiometer (SMMR) would measure from hurricanes above sea surfaces at several frequencies. These brightness temperatures depend upon the rainfall rates, rain column heights, and the emissivities of the wind roughened sea surfaces. The information content of each channel is a variable function of these parameters and their relationships are nonlinear. A piecewise-linear regression algorithm was then applied to the synthetic data in the manner discussed by Smith and Woolf [3] to infer rainfall rates. The regression method employs data from eight of the SMMR channels.

Unfortunately, the size of the antenna of the SMMR on Nimbus-7 imposes a diffraction limit on the angular resolution such that the relative angular response, [H], of the radiometer is

$$|H(\theta)| = \left[\frac{2J_1(ka \sin\theta)}{ka \sin\theta} \right]^2 \qquad (1)$$

where a is the antenna radius, θ is the angular deviation from the antenna centerline, $k = 2\pi/\lambda$ is the wave number and $J_1(\bullet)$ is the first-order Bessel function. The various channels of SMMR therefore each have a different footprint size (i.e., the 6.6, 10.7, 18.0, and 37.0 GHz channels with two polarizations have footprints of 148 x 95, 91 x 59, 55 x 41, and 27 x 18 km respectively). It is difficult to apply the regression algorithm unambiguously to real SMMR data because each channel measures radiation from an area which may contain differing amounts of rain. This study overcomes the diffraction limitation imposed on spatial resolution by means of image restoration.

II. Restoration of Spatial Resolution

The distribution of rain bearing clouds, $\vec{f}$, the observed microwave image, $\vec{g}$, the noise distribution, $\vec{\varepsilon}$, and the point spread function of the degradation, [H], are related by the following linear equation:

This work was supported in part by the NOAA under Grant MO-A01-78-00-4320; and by the University of Wisconsin-Madison WARF Foundation under Grant 135-2028.

$$\vec{g} = [H]\vec{f} + \vec{\varepsilon} \qquad (2)$$

The inverse problem, in which $\vec{f}$ is derived from the measured $\vec{g}$ in the presence of noise, requires the inversion of Fredholm integral equations of the first kind. This linear inversion requires the existence and uniqueness of an inverse transformation $[H]^{-1}$. However, even if $[H]^{-1}$ exists and is unique, it may be ill-conditioned, so that a trival perturbation in $\vec{g}$ can produce nontrivial perturbation in $\vec{f}$. Thus, an ill-conditioned problem can produce undesirable effects such as noise amplification, resulting in grainy images.

This problem can be avoided by using a constrained iterative restoration algorithm to reconstruct the multichannel images. It is well established in the literature that a finite object has an analytic spectrum. Analyticity implies that knowledge of only part of the spectrum is sufficient to uniquely determine the remainder of the spectrum. Hence, the complete spectrum may be derived from the diffraction limited image of a finite object. This remarkable property has been applied with some success by Gerchberg [4], Papoulis [5], Papoulis and Chamzas [6], Howard [7], and Rushforth et al., [8] to invert one and two dimensional signals. This property also provides the theoretical basis for our method to match the footprints of multichannel microwave images.

The spatial image restoration can be considered as an operator, $\mathcal{O}$, chosen to estimate the portion of the spectrum of the actual rain field missed by diffraction limited imaging. The operataor $\mathcal{O}$ can be defined as the two-dimensional FFT of $\vec{g}$ within some known extent. The known extent of the hurricane is determined by the a priori information provided by the 37 GHz channel (the highest frequency channel of the SMMR with the best spatial resolution), and possibly visible and infrared images. The additional spectral components generated by $\mathcal{O}$ when added to the incomplete spectrum of the observed image, restores the resolution of the image. This process is then iterated to achieve optimal resolution.

III. Results

We initially utilized the 37 GHz channels to provide an a priori estimate of the spatial extent of the rain cells that is incorporated into the algorithm. Results of our initial investigation are shown in Fig. 1. A synthetic hurricane image was created in a 16 x 16 image field. Noise free antenna temperatures of both the 37 GHz and the 6.6 GHz channels were generated. The 37 GHz image regions that have no rain were used as a constraint to extend the resolution of the 6.6 GHz image data. Fig. 1a shows the cross sections of the rain cells in a synthetic image of a hurricane before and after the restoration. The degraded 6.6 GHz image resolution has been enhanced to a large degree after a few iterations. The synthetic images of the entire model hurricane that would be measured at 37 GHz, the original degraded 6.6 GHz, and the enhanced 6.6 GHz model hurricane rain cells are shown in Figs. 1b, 1c and 1d, respectively.

Data distorted by noise has rendered it difficult to continue the spectrum beyond the original diffraction limitation. Additional spatial and spectral constraints were therefore introduced to restore the noisy images.

The spectral constraints are based on a knowledge of the highest cutoff frequency (i.e., 37 GHz), and a knowledge of the degradation point spread function. More precisely, the magnitude and the phase information of a few low-

frequency components of [H] are derived and they are used to replace the magnitude and phase of the spectrum of the degraded image.

Spatial constraints are based on the estimated extent of rain areas, derived from 37 GHz, visible and infrared images, the upper and lower bounds of the measured brightness temperatures, and some physical attributes of hurricanes. Wind patterns around a hurricane are approximately axially symmetric with respect to the hurricane eye and we have incorporated this constraint as a preprocessing step to smooth out some of the noisy data. It has been demonstrated in our simulation that by performing a running average of the wind speeds along a circular sector around the hurricane, substantial noise is reduced.

The procedure to restore low frequency bandlimited images (i.e., the 6.6, 10.7, and 18.0 GHz) to the optimal resolution (i.e., the 37 GHz) is summarized in Fig. 2. This procedure has been applied to a set of noisy images. In one example additive white noise with an rms value of 4°K was added to both the 37 GHz and 6.6 GHz synthetic images. Within a few iterations, we obtained the restored image scan as shown in Fig. 3. The 6.6 GHz noisy image is nearly completely restored to the optimal resolution.

IV. Conclusions

The constrainted iterative restoration procedure has been demonstrated through computer simulations to be effective in restoring the spatial resolution of all of the SMMR channels to a 27 x 18 km footprint. An obvious next step is to apply this procedure to real SMMR data.

References

[1] Wilheit, T. T., A. T. C. Chang, M. S. V. Rao, E. B. Rodgers, J. S. Theon, "A Satellite Technique for Quantitatively Mapping Rainfall Rates over the Ocean," J. Appl. Meteo., 16, pp. 551-560, 1977.

[2] Olson, W. S., "Estimation of Rainfall Rates in Tropical Cyclones by Passive Microwave Radiometry," Ph.D. thesis, (in progress) Dept. of Meteorology, University of Wisconsin-Madison.

[3] Smith, W. L. and H. M. Woolf, "The Use of Eigenvectors of Statistical Convariance Matrices for Interpreting Satellite Sounding Radiometer Observations," J. Atms. Sci., 33, pp. 1127-1140, 1976.

[4] Gerchberg, R. W., "Super-Resolution through Error Energy Reduction," Optica Acta, 21, pp. 709-720, 1974.

[5] Papoulis, A., "A New Algorighm in Spectral Analysis and Bandlimited Extrapolation," IEEE Trans. Circuits and Systems, CAS-22,9, pp. 735-742, 1975.

[6] Papoulis, A. and C. Chamzas, "Detection of Hidden Periodicities by Adaptive Extrapolation," IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-27, 5, pp. 492-500, 1979.

[7] Howard, S. J., "Method for Continuing Fourier Spectra given by the Fast Fourier Transform," J. Opt. Soc., Am., 71, pp. 95-98, 1981.

[8] Rushforth, C. K., A. E. Crawford, and T. Zhou, "Least-Squares Reconstruction of Objects with Missing High-Frequency Components," J. Opt. Soc. Am., 72, pp. 204-211, 1982.

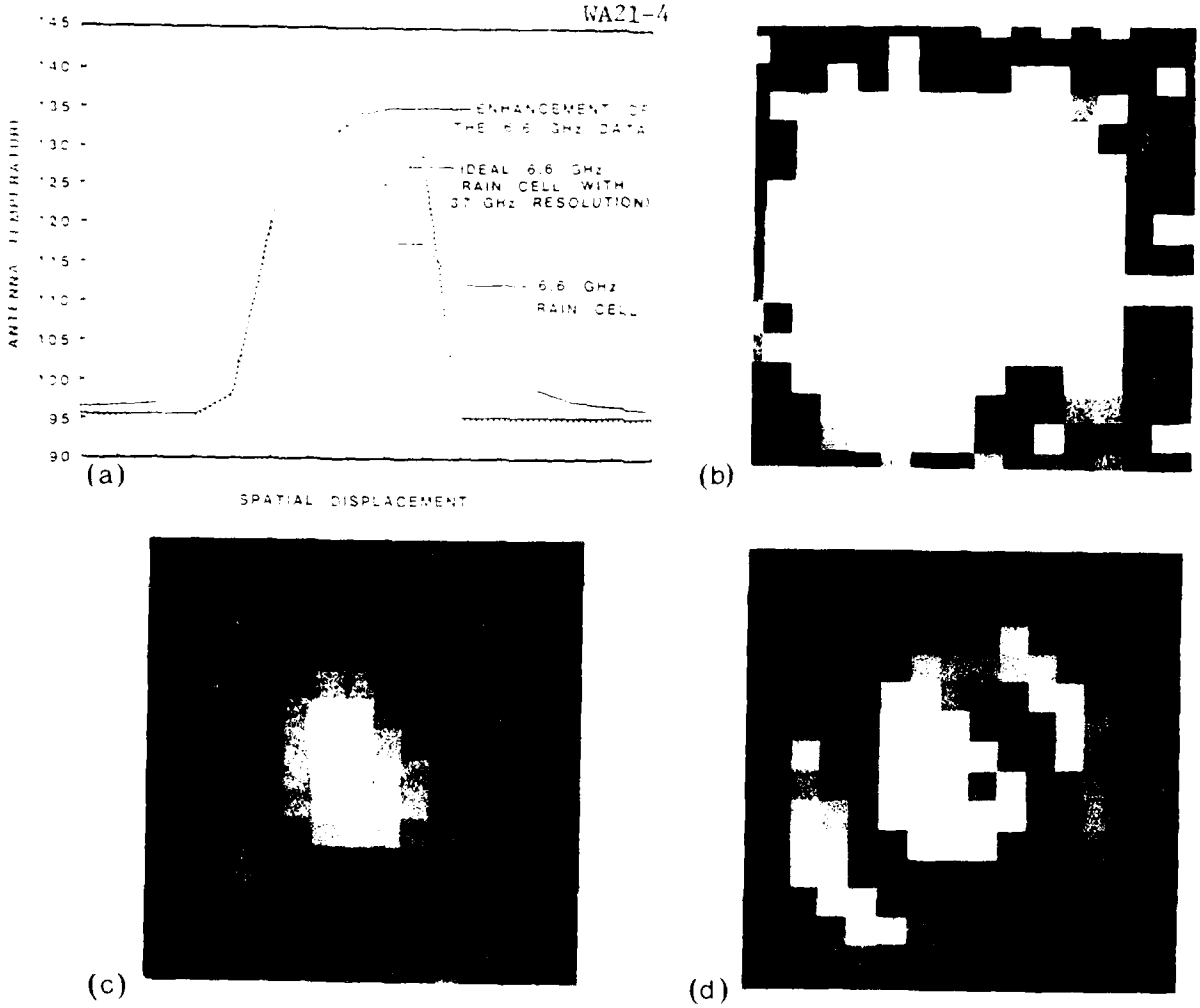


Figure 1: a) Enhancement of the resolution of a synthetic 6.6 GHz noise free image of a hurricane. This enhancement algorithm uses the known extent of the image shown in b) to enhance the degraded 6.6 GHz image shown in c). The derived enhanced image is shown in d).

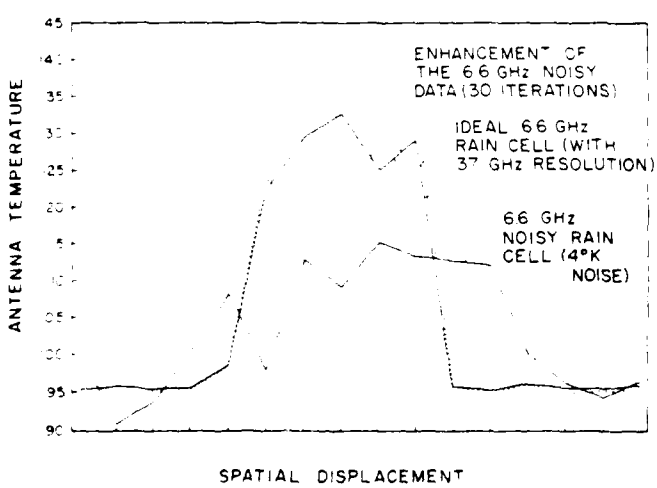
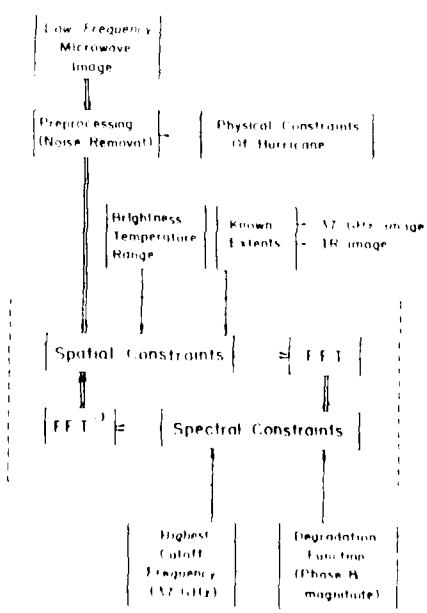


Figure 3: Restoration of Noisy Images.

Block Diagram of the Restoration Algorithm: Figure 2



Electromagnetic image reconstruction techniques in inhomogeneous media satisfying the Born-Rytov approximation

Wolfgang-M. Boerner
Electromagnetic Imaging Division, Communications Laboratory
Department of Information Engineering, University of Illinois at Chicago
P.O. Box 4348, SEO-1141
Chicago, Illinois 60680

Introduction

The basis for developing projection tomographic reconstruction algorithms has been the assumption of straight-line ray-path propagation. But, in the case in which propagation occurs within discretely inhomogeneous media at wavelengths of the order of the size of the scatterer, phenomena such as refraction, reflection and diffraction can no longer be neglected and a straight-line projection tomographic approach fails. This is especially evident when a large difference in refractive index occurs, such as that encountered with dm-to-mm-wave propagation in inhomogeneous atmospheric media, representing hydrometeorite distributions, the marine ocean boundary layer, the ground surface underburden, or bone and soft layers within soft tissue. An exact solution for the general vector scattering case which strictly requires a polarimetric radiative transfer approach is not available, and in this research, the assumption is made that the media are weakly diffracting so that the Born and Rytov approximations are valid. Based on this assumption, various diffraction imaging methods were developed most recently, and we are basing our studies on Devaney's back-propagation tomographic approach which was developed upon the scalar wave theory. It is the main objective of this research to extend this work to the realm of electromagnetic vector wave theory for the improved diffraction-corrected imaging of radar targets embedded in clutter within the dm-to-mm-wavelength region of the electromagnetic spectrum.

The mathematical theory of Radon Transformations is the basis of modern imaging techniques in computerized tomography. Hounsfield's invention of the computerized tomography (CT) scanner for bio-medical imaging started the avalanche of techniques of imaging in both electromagnetics and ultrasound as described in Hounsfield[1].

Basically, three types of CT imaging methods are in use: The x-ray CT, the ECT (Emission Computerized Tomography) and more recently in ultrasound CT as in Brooks and Dichiov[2].

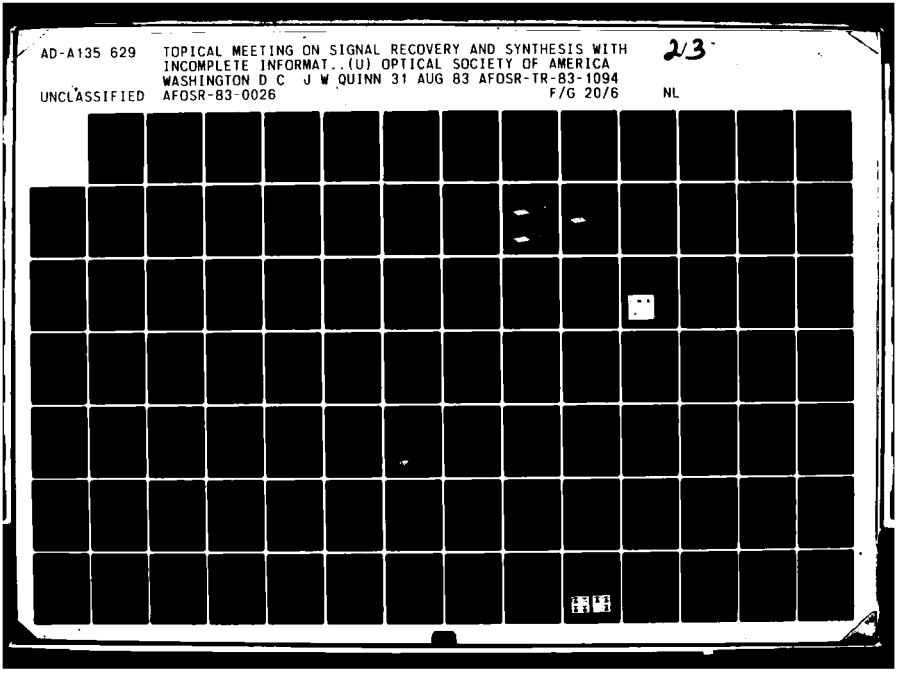
There exist two distinct imaging reconstruction methods for the electromagnetic case:

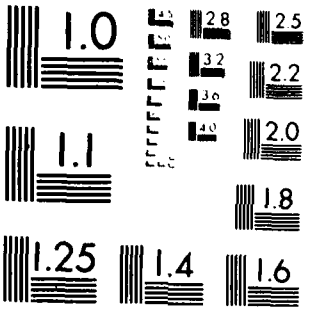
1) Complete polarimetric image reconstruction from cross-range signatures as, for example, encountered in Microwave Synthetic Aperture Radar Imagery in which the final reconstructed image is obtained by incoherently superimposing the images reconstructed from four (4) independent SAR-images

AD-A135 629 TOPICAL MEETING ON SIGNAL RECOVERY AND SYNTHESIS WITH INCOMPLETE INFORMAT..(U) OPTICAL SOCIETY OF AMERICA WASHINGTON D C J W QUINN 31 AUG 83 AFOSR-TR-83-1094

UNCLASSIFIED AFOSR-83-0026 F/G 20/6 NL

2/3





extracted from $|S_{AA}|^2$, $|S_{BA}|^2$, $|S_{AB}|^2$ and $|S_{BB}|^2$, where in the case of a reciprocal propagation space $|S_{AB}|^2 = |S_{BA}|^2$.

2) Complete polarimetric image reconstruction from slant-range signatures as, for example, in the case of radar target shape reconstruction using ramp response echoloting or linear chirp downrange scatterometry in which the final best-recoverable image is obtained from applications of Radon's projection theory to Kennaugh's optimal target echo area. This optimal target echo area needs to be recovered from complete polarimetric relative phase scattering matrix data, either using the polarization transformation invariant,

$$|S_{opt}|^2 = \text{Span}\ \{[S]\} = |S_{AA}|^2 + 2|S_{AB}|^2 + |S_{BB}|^2,$$

for the reciprocal target sensing case, or utilizing a geometrical reconstruction method applied directly to the optimal null location chart derived from the Poincare sphere.

If the medium is homogeneous the ray paths are along rectilinear geodesics and this simplifies the shape reconstruction process using the "parallel beam" projection algorithms. These projection tomographic principles rely on a g.o. limit description of the imaging process that may be (not necessarily is) adequate at x-ray wavelengths, but that has quite questionable validity in optical and ultrasonic, and very specifically, in electromagnetic dm-to-mm-wave applications in which diffractions and scattering effects do become important as described in Keller[3], Bates et al[4], Devaney[5]. (Also H.P. Baltes[14])

Radon projection theory

The problem of determining the size, shape and electromagnetic properties of a scatterer, given the incident and scattered electromagnetic fields, is the electromagnetic inverse problem. The problem of interest here is the reconstruction of the image of a convex scatterer from knowledge of the high frequency far field scattered from the object in response to a known incident field.

The size and shape of an object can be obtained from its area functions, and the problem can be reduced to the classical Radon problem as shown in Das and Boerner[7]. Since the area functions can be estimated from the scattered field when the incident field is a ramp, the remaining problem is thus essentially that of reconstruction from projections, as shown in Kennaugh and Cosgriff[6]. The problem of reconstruction has found wide application as of late in computer-assisted tomography.

Born and Rytov approximations

For the case where $(n^2 - 1)$ is small in the wave equation

$$\nabla^2\psi + (n^2)/(c^2)\psi = 0,\ n = n(\gamma, k),\ \Psi = \Psi\ (\gamma, k),$$

it is possible to recover $n(\gamma)$ using the Born-Rytov approximations given

measurements outside the domain of support at a fixed frequency and for different angles of plane wave incidence. The Rytov approximation provides a significant improvement of the Born approximation providing that scattering is mainly in a forward direction as was shown by Keller[3]. In a recent study by Bates et al[4], we have extended the Rytov approximation to smoothly varying media with the assumption that refraction predominates over reflection and its significance for acoustic remote sensing was demonstrated by Dunlop[15] for the case of slight ray bending. Inverse problems making use of the Rytov approximation are encountered frequently in passive atmospheric sounding, in optical applications and in ultrasonic computerized tomography. In most cases, measurements of the emitted terrestrial or reflected solar radiation is carried out as a function of wavelength, look angle, and atmospheric optical depth using ground-based, air-borne, or space-borne systems. The data, thus obtained, are inverted to yield the vertical structure of several atmospheric parameters. Like other inversion problems, the difficulties are both mathematically and numerically oriented (ill-posedness, non-linearity) and the ultimate choice of the best method for a particular application may not be based on accuracy, but is normally restricted by available resources, quantity of experimental data, and a priori knowledge.

Most recently, a succinct assessment of this inverse problem was made in A.J. Devaney[8], in which he overlooked our paper that contains, in parts, identical results, proving that the small wavelength extension of the Rytov approximation will pave the basis for diffraction tomography in weakly diffracting media.

Ray path bending problem

The basis for developing projection tomographic reconstruction algorithms has been the assumption of parallel beam transmitted wave propagation, whereby, the received signal at the observation point was related to a particular parameter of the irradiated media strictly along a straight line path between the receiver and the anticipated location of wave/media interrogation, i.e., rectilinear (forward and/or backward projection tomography). But, in the case in which wavelengths are of the order of the size of the scatterer, phenomena such as refraction, reflection and diffraction can no longer be neglected, and a straight-line projection tomographic approach fails. This is especially evident when viewing a reconstructed image of a region within which a large difference in refractive index (or for the case of x-ray, strong Bragg diffraction) occurs, such as that encountered with dm-to-mm wave propagation in inhomogeneous atmospheric media representing hydrometeorite distributions, the marine ocean boundary layer, the ground surface underburden, or bone and soft layers within living tissue. In light of this non-rectilinear projection problem of having to account for the non-straight path nature of wave propagation, a varied number of different algorithms have been developed in many different disciplines while a synthetic focusing operative system for reflective tomographic reconstruction has been developed by Greenleaf, Johnson et al[9], the analytical studies by Bates et al[4], and the experimental verification by Dunlop[15] discusses improvements to the Rytov approach and its applications, where we show that the Rytov approximation provides a significant improvement on the Born approximation

provided that the scattering is mainly in a forward direction, which is consistent with the independent findings of Keller[3] and Devaney[10].

Vector diffraction tomography

The electromagnetic inverse problem of reconstructing material parameters (ξ, μ, σ) profiles of inhomogeneous distributed regions is applicable to radar clutter and distributed (underburden) target description, non-destructive material testing, mine detection, and to microwave sub-acquatic as well as scalar acoustical media imaging. Due to the fact that electromagnetic waves are vector in nature, i.e., their complete polarimetric behavior for propagation in depolarizing inhomogeneous media need to be integrated into profile/image reconstruction algorithm, requiring Devaney's scalar diffraction tomography to be extended into "Vector Diffraction Tomography".

Work is presently under progress at the Communications Laboratory, University of Illinois at Chicago where we use the scalar equations of Devaney's and convert them to the vector ones for the electromagnetic case. We consider a vector a dyadic Green's function $\overline{G}$ (ξ, η) which is a propagator to map the filtered relative phases back into the image space. The construction of the dyadic Green's function involves the use of the invariance of the span of the scattering matrix. We parallel the work of Devaney to obtain the filtered relative phases as

$$\Pi_{\phi}(\xi, \eta) \int_{-\infty}^{\infty} d\xi \; Q_{\phi_o}(\xi') \; \overline{\overline{G}}(\xi - \xi', \eta - l_o)$$

to obtain an estimate on the profile as

$$\hat{O}(\underline{r}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\phi_o \; \Pi_{\phi_o}$$

References

1. G.N. Hounsfield, British J. Radiology, Vol. 46, pp. 1016-1022.
2. R.A. Brooks and G. Dichiov, Phy. Med. Bio., Vol. 21, pp. 689-732, 1976.
3. J.B. Keller, J. Opt. Soc. Am 59(8), pp. 1003-1004, 1969.
4. R.H.T. Bates, W-M. Boerner and G. Dunlop, Optics Comm. 18(4), pp. 421-423, 1976.
5. A.J. Devaney, IEEE Trans. Ultrasonic imaging, submitted March 18, 1982.
6. E.M. Kennaugh and R.L. Cosgriff, IRE National Convention Record, Part I, pp. 72-77, 1958.
7. Y. Das and W-M. Boerner, IEEE Trans. Antennas & Propagation, Vol. AP-26, No. 2, pp. 274-279, 1978a.
8. A.J. Devaney, Optics Letter, Vol. 6, 1981, pp. 374-376.
9. J.F. Greenleaf, S.A. Johnson and A. Lent, Ultrasound in Medicine and Biology 3, pp. 327-339, 1978.
10. A.J. Devaney, Submitted to IEEE Trans. Biomed. Eng., Nov. 1981.
11. W-M. Boerner, C-M. Ho and B-Y. Foo, IEEE Trans AP-29(2), March 1981.
12. W-M. Boerner, C-M. Ho, Wave Motion 3, 1981, pp. 311-333, North Holland Publishing.
13. W-M. Boerner, M.B. El-Arini, C-Y. Chan and P.M. Mastoris, IEEE Trans AP-29(2), March 1981.
14. H.P. Baltes, ed., Springer Verlag, 1978.
15. G.R. Dunlop, Ph.D. Thesis, Univ. of Canterbury, Christchurch, New Zealand, 1978 (unpublished).

SESSION V

RECONSTRUCTION FROM INTENSITY I

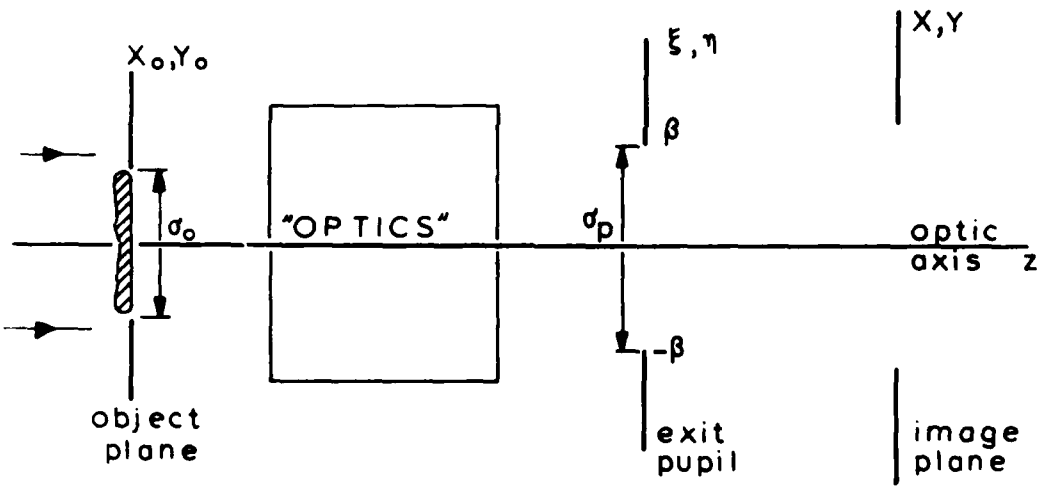
J. R. Fienup, *Presider*

The Phase Problem in Object Reconstruction and Interferometry.

H.A.Ferwerda, Department of Applied Physics, State University of Groningen, Nijenborgh 18, 9747 AG Groningen, The Netherlands.

1. Introduction. In this contribution I shall review phase problems from different fields of optics which can be handled with similar techniques. In all cases the problem is to reconstruct the phase of a function from its modulus. In object reconstruction we have to know the complex image wave function (w.f.) while the intensity distribution only gives its modulus. In speckle interferometry only the autocorrelation of the brightness distribution of the source (or equivalently the modulus squared of its Fourier transform) is measurable. In interference microscopy often only the visibility of the interference fringes formed in a Michelson interferometer can be observed, yielding the absolute value of the complex degree of temporal coherence. But we also need its phase for the calculation of the spectral distribution of the source.
Noise will not be taken into account in this review.

2. The phase problem of object reconstruction.



Consider an object, coherently illuminated by a quasi-monochromatic plane wave, described by the wave function exp(ikz) (we assume that the imaging process can be described in terms of scalar functions). After interaction with the object we obtain the w.f. $\psi_o(x_o,y_o)$ in the object plane, which is the plane perpendicular to the z-axis just touching the object. $\psi_o(x_o,y_o)$ will be the quantity to be reconstructed. The specific relation between ψ_o and the object structure is outside the scope of this contribution. In the exit pupil of the imaging system (in electron microscopy denoted as diffraction plane) we obtain on the "left-hand" side of the aperture the w.f.

$$\psi_{p,\ell}(\xi,\eta) = \int_{\sigma_o} \psi_o(x_o y_o)\exp[2\pi i(x_o\xi+y_o\eta)]\,dx_o dy_o \qquad (1)$$

Due to aberrations of the optical system we have to introduce the wave-aberration function $\Phi(\xi,\eta)$ (assuming isoplanatic imaging). This is done by writing the w.f. on the "right-hand" side of the aperture in the exit pupil as

$$\psi_{p,r}(\xi,\eta) = \exp[i\Phi(\xi,\eta)]\,\psi_{p,\ell}(\xi,\eta). \qquad (2)$$

Finally, the w.f. in the image plane is given by

$$\psi_i(x,y) = \int_{\sigma_p} \psi_{p,r}(\xi,\eta)\exp[2\pi i(\xi x+\eta y)]\,d\xi d\eta\ . \qquad (3)$$

x_o, y_o are measured in units λ(wavelength), ξ,η in units F (back focal distance) and x,y in units Mλ (M is the lateral magnification). The recorded quantity, the image intensity, is proportional to $|\psi_i(x,y)|^2$. For the determination of $\psi_o(x_o,y_o)$ we need the complex ψ_i. Walther [1] has already noticed that there is a relation between the phase and modulus of ψ_i because ψ_i is a band-limited function, as is evident from (3). In the following we shall make two important assumptions: i) we assume only one transverse spatial dimension. ii) we neglect noise. The first restriction is expedient because in that case we can draw from the theory of entire functions of one complex variable. The case of two complex variables is very hard on physicists! The second assumption will be briefly touched upon later on.

3. The relation between phase and modulus of a band-limited function. When $\psi_i(x)$ (from now on we shall drop the subscript "i") is continued in the complex plane we obtain an entire band-limited function (cf. eq.3). Two approaches have been presented: 1) a dispersion relation between $\arg\psi(x)$ and $|\psi(x)|$ can be derived, provided we know the zeros of $\psi(x)$ in the upper half plane [2]; 2) $\psi(x)$ is essentially specified by its zeros [3]. For both approaches it is necessary to determine the zeros of $\psi(x)$ in the complex plane.

4. Methods for determining the zeros of $\psi(z)$. If I(x) denotes the intensity distribution in the image plane, this quantity can be continued analytically:

$$I(z) = \psi(z)\psi^*(z^*). \qquad (4)$$

The zeros of I(z) can be determined by first determining the Fourier transform (FT) $i(\xi)$ of I(x). Then

$$I(z) = \int_{-\infty}^{\infty} i(\xi)\exp(-2\pi\xi y)\exp(2\pi i\xi x)\,d\xi, \qquad (5)$$

an operation to be performed on a computer [4]. The zeros of I(z), occurring in conjugate pairs, can thus be obtained. If the number of important zeros of $\psi(z)$ (which is approximately the number of degrees of freedom of the image) is N, there are 2^N possibilities for assigning the zeros to $\psi(z)$.
Several searches for zeros of $\psi(z)$ have been put forward to mention by Walker [5] and Brames & Dainty [6]. All these proposals use two measurements. In the first measurement a_n, a_n^* are determined. In the second measurement Walker places an exponential filter with transmission function $\exp(-2\pi a\xi)$ over the exit pupil; the analytic continuation of the corresponding image intensity distribution yields

displaced zeros; $a_n', a_n'^*$. If $a_n=\alpha_n+i\beta_n$, $a_n'=\alpha_n'+i\beta_n'$ we have the relationship: $\alpha_n'=\alpha_n$; $\beta_n'=\beta_n-a$. The zeros a_n and a_n^* of ψ and ψ^*, respectively, move in different directions, enabling us to establish which zeros belong to $\psi(z)$. This method can be extended to two spatial transverse dimensions simply by keeping one of the coordinates fixed.
This method works well in the case where the stochastic nature of the image may be ignored. If this is no longer acceptable as e.g. in low-dose electron microscopy or noisy images in general, other approaches may become useful as will be discussed in the next section.

5. Methods not based on analyticity arguments. The advantage of the approach by zeros of band-limited functions is that the solution of the problem belongs to the class of band-limited functions. Noise may spoil that advantage. Other approaches originally proposed by Gerchberg and Saxton [2] and Misell [2] have been studied in great detail in my group by Huiser and Van Toorn [2] in order to assess the significance of the iterative numerical procedures of Gerchberg and Saxton, and Misell. The Gerchberg-Saxton proposal tries to deduce the phase of the w.f. from the intensity distributions in image and diffraction plane. In their mathematical analysis Huiser and Van Toorn express $i(\xi+\beta)$ in terms of $\psi_{p,r}(\xi)$: (we drop the subscripts p,r in this section)

$$i(\xi+\beta) = \int_{\xi}^{\beta}\psi(\xi')\psi^*(\xi'-\xi-\beta)d\xi'. \qquad (6)$$

The phase problem is essentially solved when $\psi(\xi)$ is known. In the Gerchberg-Saxton approach ψ has to be solved from (6) using the fact that $|\psi|$ is known. Discretisizing the integral in (6) as a Riemann sum yields quadratic equations for the sampling values of ψ. As each quadratic equation usually has two solutions, we obtain a wealth of possible solutions. Evidently additional information (additional constraints) is necessary. A good example of such a case has been studied in a seminal paper by Fienup [7]. He reconstructs a brightness distribution from only the modulus of its Fourier transform, exploiting the constraint that the brightness is non-negative definite and that the support of the brightness function is finite. In the case of the Misell algorithm $\psi(\xi)$ has to be solved from the two coupled non-linear Volterra-like integral equations arising from two defocused images:

$$i_j(\xi+\beta) = \int_{\xi}^{\beta}\psi(\xi')\psi^*(\xi'-\xi-\beta)\exp[i\Delta_j\{\xi'^2-(\xi'-\xi-\beta)^2\}]d\xi'; \qquad j=1,2, \qquad (7)$$

where Δ_j is a measure of the amount of defocus of the exposure. Writing (7) as Riemann sums we get for every sampling value two linear equations in two unknowns, leading to a unique solution. Numerical investigations have revealed that the procedure is rather sensitive to noise which makes a statistical study imperative, in particular in the case of low intensity illumination. Such a study is in progress by C.H.Slump of my department [8].
The approach described in this section can be extended to two dimensions.

6. Phase problem of coherence theory and interferometry. The complex degree of temporal coherence $\gamma(\tau)$ and the spectral density of a light source $g(\omega)$ are related by ([9]):

$$\gamma(\tau) = \int_{o}^{\infty}g(\omega)\exp(-i\omega\tau)d\omega, \qquad (8)$$

where $g(\omega) \geq o$. In interference microscopy $\mathrm{Re}\gamma(\tau)$ is the measured quantity. Because $\mathrm{Re}\gamma(\tau)$ is often a very rapidly oscillating function we have to content ourselves with the envelope function of the oscillations in the interference pattern, the visibility curve, yielding $|\gamma(\tau)|^2$. For the reconstruction of $g(\omega)$ we need the complex $\gamma(\tau)$. Wolf [10] noticed that analyticity properties might yield the phase, the role of the zeros of $\gamma(\tau)$ in the complex τ-plane being important but obscure. Nussenzveig [9] has studied these zeros in some explicit cases. Intensity interferometric arrangements for the determination of the phase of the coherence functions have been discussed by Gamo [11]. Brames and Dainty [6] have independently used the technique of determining the zeros of an entire band-limited function to solve the phase problem in stellar speckle interferometry.

7. Phase problem in two spatial dimensions. The methods based on analyticity (zeros of band-limited functions, dispersion relations) cannot straightforwardly be extended to the case of two lateral spatial dimensions. It is believed [7,12] that the phase problem in two dimensions is less ambiguous than in one dimension. Solid mathematical treatment is scarce. Important progress has been made by Bruck and Sodin [12] whose results seem to indicate that uniqueness is the rule rather than an exception: the criterium for uniqueness is whether a polynomial in two variables is factorizable or not: if the polynomial is factorizable there is ambiguity, if not the result is unique. Huiser and Van Toorn [13] have constructed examples of ambiguities. The role of noise remains to be clarified.

Acknowledgement. Part of this research was supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

References.

[1] A.Walther, Opt.Acta 10 (1963) 41
[2] H.A.Ferwerda in Ch.2 of Inverse Source Problems, H.P.Baltes (Ed.), Springer Berlin 1978, where further references can be found.
[3] M.A.Fiddy et al, Opt.Acta 29 (1982) 23
[4] N.Nakajima and T.Asakura, Optik 60 (1982) 289
[5] J.G.Walker, Opt.Acta 28 (1981) 735
[6] B.J.Brames and J.C.Dainty, J.Opt.Soc.Am. 71 (1981) 1542
[7] J.R.Fienup, Opt.Lett. 3 (1978) 27
[8] C.H.Slump and H.A.Ferwerda, Optik (to be published)
[9] H.M.Nussenzveig, J.Math.Phys. 8 (1967) 561 and references cited there
[10] E.Wolf, Proc.Phys.Soc.(London) 80 (1962) 1269
[11] H.Gamo, "Symposium on Electromagnetic Theory and Antennas", Copenhagen 1962; Pergamon; "Advances in Quantum Electronics", J.R.Singer (Ed.) p.251, Columbia University Press
[12] Y.M.Bruck and L.G.Sodin, Opt.Commun. 30 (1979) 304
[13] A.M.J.Huiser and P.van Toorn, Opt.Letters 5 (1980) 499

Phase Retrieval from Intensity Data degraded by Shot-Noise.

C.H. Slump and H.A. Ferwerda
Department of Applied Physics, State University at Groningen,
Nijenborgh 18, 9747 AG Groningen, The Netherlands.

1. Introduction. In the paper by one of us (H.A.F.) [1], the phase problem was reviewed for circumstances under which the stochastic character of the image could be ignored. A statistical description becomes imperative for images obtained under low-intensity illumination as occurs e.g. in low-dose electron microscopy. Under these circumstances the images are heavily degraded by shot-noise. Often(e.g. [2,3])images are described as a signal plus signal-independent noise. Such a treatment does not apply to shot-noise, the quanta arriving in non-overlapping image cells are statistically independent Poisson-distributed random variables [4]. The variances of the data-counts also contain information about the signal. In this contribution we discuss the retrieval of the object wave function (w.f.) from shot-noise degraded images. Due to the stochastic imaging process, the reconstructed object w.f. is also stochastic. The purpose of this contribution is to establish the statistical characterization of the object w.f. We assume axial coherent quasi-monochromatic illumination. The analysis is presented for one lateral dimension only, the extension to two dimensions is obvious.

2. Statistical description of the image. The image plane is divided into N non-overlapping cells (pixels) with N the Shannon number (N=4β2d). The number of electrons arriving in the k-th cell is denoted by $\hat{n}_k$ where the hat "^" denotes that we deal with a random variable. $\hat{n}_k$ has a Poisson distribution with mean λ_k, where λ_k is proportional to the integral of the squared *modulus of the image* w.f. over the k-th cell. In our approach this noisy image is the input for the reconstruction procedure of the object w.f.. We do not smooth or filter by whatever method the data to be processed!

3. Reconstruction of the object wave function. The relation between the Fourier transform (F.T.) of the image and the w.f. $\psi_p(.)$ in the exit pupil (see [1] for details)is given by:

$$\hat{\imath}_k = \lambda_o N^{-1} \sum_{\ell=-\frac{1}{2}N+k}^{\frac{1}{2}N-1} \psi_p(\ell)\psi_p^*(\ell-k), \qquad k=0,1,\ldots \quad , \tag{1}$$

where $$\hat{\imath}_k = \sum_{\ell=-\frac{1}{2}N}^{\frac{1}{2}N-1} \hat{n}_\ell \exp(2\pi i k \ell N^{-1}) \ . \tag{2}$$

$\psi_p(\ell)$ stands for $\psi_p(\ell/2d)$ and λ_o denotes the mean number of quanta incident per pixel and 2d is the size of the object. The phase problem is essentially solved when $\psi_p(.)$ is determined. As $\hat{\imath}_k$ is a stochastic quantity, the resulting $\psi_p(.)$ also acquires stochastic properties. From (2) we derive that $\hat{\imath}_k$, $k=-\frac{1}{2}N,\ldots,\frac{1}{2}N-1$ are correlated complex random variables; the autocorrelation function is:

$$R(k,k') = E(\hat{\imath}_k \hat{\imath}_{k'}^*) = E(\hat{\imath}_k)E(\hat{\imath}_{k'}^*) + \sum_{\ell=-\frac{1}{2}N}^{\frac{1}{2}N-1} \lambda_\ell \exp(2\pi i \ell (k-k')N^{-1}), \tag{3}$$

the symbol E(.) denotes the mathematical expectation value. The non-linear equation (1) is not simply solvable and to characterize the obtained solution statistically is even more difficult. In the next section we shall therefore treat the more tractable case of weak-object imaging.

4. Weak-object reconstruction. In this case the non-linear terms in the squared modulus of the image w.f. are neglected and λ_k depends linearly on the object w.f.. This allows us to enlarge the pixel area, the signal-to-noise ratio of the image data is increased. We have now $\frac{1}{4}N$ pixels. λ_k is written $\lambda_k=\lambda_o(1+s_k)$ where s_k is the contrast due to the object ($s_k<<1$). $\hat{s}_\ell$ is estimated from $\hat{n}_\ell$ by:

$$\hat{s}_\ell = \lambda_o^{-1}(\hat{n}_\ell-\lambda_o) \tag{4}$$

Calculating $\hat{j}(.)$, the F.T. of the new random variables we obtain for the autocorrelation function:

$$E(\hat{j}_k\hat{j}^*_{k'}) = E(\hat{j}_k)E(\hat{j}^*_{k'}) + \sum_{\ell=-\frac{1}{4}N}^{\frac{1}{4}N-1} \mathrm{var}(\hat{s}_\ell)\exp(4\pi i\ell(k-k')N^{-1})$$

$$\simeq E(\hat{j}_k)E(\hat{j}^*_{k'})+\tfrac{1}{2}\lambda_o^{-1}\,\delta_{k,k'} \tag{5}$$

$$\text{with: } \hat{j}_k = \sum_{\ell=-\frac{1}{4}N}^{\frac{1}{4}N-1} \hat{s}_\ell\,\exp(4\pi ik\ell N^{-1}) \tag{6}$$

In the derivation of (5) we used that the variance of $\hat{s}_\ell$ is equal to $\lambda_o^{-1}(1+s_\ell)\simeq\lambda_o^{-1}$. It can be shown that the probability distribution of $\hat{s}_\ell$ is in good approximation Gaussian with mean s_ℓ and variance equal to λ_o^{-1}. The real and imaginary part of $\hat{j}_k$ are therefore also Gaussian distributed with mean the true deterministic values and with variances: $\lambda_o^{-1}\sum_\ell(1+s_\ell)\cos^2(4\pi k\ell N^{-1})$ and $\lambda_o^{-1}\sum_\ell(1+s_\ell)x$ x $\sin^2(4\pi k\ell N^{-1})$, respectively. The relation between $\hat{j}_k$ and the object w.f. is given by: (see e.g. [5])

$$\hat{j}_k = 2\,a_k\sin\phi(k)-2\,b_k\cos\phi(k), \tag{7}$$

$\phi(k)$ denotes the wave aberrations of the optical system and a_k and b_k are the F.T. of $\alpha(.)$ and $\beta(.)$, respectively. $\alpha(.)$ and $\beta(.)$ denote the phase and amplitude part of the object w.f.: $\psi_o(x_o)=1+i\alpha(x_o)-\beta(x_o)$. The aberration function $\phi(.)$ depends on the spherical aberration coefficient C_s and the defocusing δ_z:

$$\phi(k)=\phi(k/2d)=(2\pi/\lambda)\,[\tfrac{1}{4}C_s(k/2d)^4-\tfrac{1}{2}\delta_z(k/2d)^2]. \tag{8}$$

Taking two images with different defocus allows us to calculate a_k and b_k from two equations of the type (7). From (7) we see with (8) that Fourier components a_k for k in the neighborhood of zero do not contribute to the contrast in the image. Calculating these Fourier components would give rise to a very large noise-variance in the reconstructed object w.f.. We therefore exclude these components from the reconstruction procedure; a band-pass filtered

object phase function $\bar{\alpha}(.)$ is reconstructed. The reconstruction of $\beta(.)$ follows directly from the two equations of the type (7):

$$\hat{\beta}(\ell/2\beta) = N^{-1} \sum_{k=-\frac{1}{2}N}^{\frac{1}{2}N-1} \hat{b}_k \exp(-4\pi i k \ell N^{-1}) = \beta(\ell/2\beta)+N(o,\sigma_\beta^2), \quad \forall \ell \tag{9}$$

where $N(o,\sigma_\beta^2)$ denotes a zero-mean Gaussian variable with variance σ_β^2. This variance is signal independent and is only a function of the parameters of the imaging system:

$$\sigma_\beta^2 = (8N\lambda_o)^{-1} \sum_{k=-\frac{1}{2}N}^{\frac{1}{2}N-1} \frac{\sin^2\phi_2(k)+\sin^2\phi_1(k)}{\sin^2(\phi_1(k)-\phi_2(k))} \tag{10}$$

where the suffixes "1" and "2" symbolise the two different defocusings. The phase reconstruction proceeds along similar lines:

$$\hat{\bar{\alpha}}(\ell/2\beta) = \bar{\alpha}(\ell/2\beta) + N(o,\sigma_{\bar{\alpha}}^2) , \quad \forall \ell \tag{11}$$

were

$$\sigma_{\bar{\alpha}}^2 = (8N\lambda_o)^{-1} \sum_k' \frac{\cos^2\phi_2(k)+\cos^2\phi_1(k)}{\sin^2(\phi_1(k)-\phi_2(k))} \tag{12}$$

were the accent (",") denotes that values of k in the immediate neighborhood of zero (specified more precisely in [4]) have been omitted.

5. Discussion. The disadvantage of the present approach is that the low spatial frequencies of the phase structure of the object are practically beyond retrieval. This obstacle is circumvented by illuminating the object from different directions. The results of this research will soon be reported.

Acknowledgement. One of us (C.H.S.) acknowledges the support from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.)

References.

[1] H.A. Ferwerda, this conference.
[2] C.W. Helstrom, J.Opt.Soc.Am. 57 (1967), 297-303.
[3] D. Slepian, J.Opt.Soc.Am. 57 (1967), 918-922.
[4] C.H. Slump, H.A. Ferwerda, Optik, to be published.
[5] P.W. Hawkes: "Image Processing Based on the Linear Theory of Image Formation", in Computer Processing of Electron Microscope Images, P.W. Hawkes (Ed.), Springer, Berlin, 1980.

Phase Retrieval for Functions with Sufficiently Disconnected Support

T.R. Crimmins and J.R. Fienup

Environmental Research Institute of Michigan
P.O. Box 8618
Ann Arbor, Michigan 48107

1. INTRODUCTION

The problem of phase retrieval is to reconstruct a function, $f(x)$, from the modulus, $|F(u)|$, of its Fourier transform,

$$F(u) = \int_{-\infty}^{\infty} f(x)\, e^{-iux}\, dx.$$

This is equivalent to reconstructing the phase of $F(u)$ from $|F(u)|$ or to reconstructing $f(x)$ from its autocorrelation function, which is given by the inverse Fourier transform of $|F(u)|^2$. This problem arises in many fields, including astronomy, x-ray crystallography, wavefront sensing, pupil function determination, electron microscopy, and particle scattering. In this paper the function, f, is assumed to be a square-integrable, one-dimensional, complex-valued function. If f has disconnected compact support contained in the union of a sequence of intervals satisfying a certain separation condition, then it can be shown that f is almost always essentially the only solution with support contained in the union of those intervals. This holds no matter how many non-real zeroes F has.

2. EQUIVALENT SOLUTIONS

Let c be a real number and C be a complex number with $|C|=1$ and let $g(x) = Cf(x+c)$ and $h(x) = C\overline{f(-x+c)}$, where the overbar denotes complex conjugation. If F, G and H are the Fourier transforms of f, g and h, respectively, then

$$G(u) = C\, e^{icu}\, F(u), \text{ and } H(u) = C\, e^{-icu}\, \overline{F(u)}.$$

Thus

$$|G(u)| = |F(u)| = |H(u)|.$$

The solutions f, g and h are called <u>equivalent</u> or, in symbols,

$$g \approx f \approx h.$$

If all solutions are equivalent to f, then f is said to be <u>essentially</u> the only solution or the <u>unique</u> solution.

3. THE THEOREM

Let I_n=n=1,...,N be a sequence of nonoverlapping closed intervals. Define

$$I_m - I_n = \{x-y: x\varepsilon I_m \text{ and } y\varepsilon I_n\} .$$

We will assume that the following condition is satisfied.

Separation Condition: $(I_m-I_n)\cap(I_j-I_k) = \phi$ for $1\leq m,n,j,k\leq N$; $j\neq k$; and $(m,n)\neq(j,k)$, where (,) denotes ordered pair. (Note that m=n is allowed.)

For N=2, this condition is equivalent to the requirement that the lengths of the two intervals be less than the distance between them.

Returning to the general case, let

$$A = \bigcup_{n=1}^{N} I_n$$

and let f and g be two complex-valued square-integrable functions both of which are zero outside of A. For n=1,...,N, let

$$f_n(x) = \begin{cases} f(x), \text{for } x\varepsilon I_n \\ 0, \text{ otherwise} \end{cases}$$

and

$$g_n(x) = \begin{cases} g(x), \text{for } x\varepsilon I_n \\ 0, \text{ otherwide.} \end{cases}$$

Then

$$f(x) = \sum_{n=1}^{N} f_n(x), \text{ and } g(x) = \sum_{n=1}^{N} g_n(x).$$

It is assumed that $f_n \not\equiv 0$ [i.e., $f_n(x)$ is not identically zero], n=1,...,N.

Note that the autocorrelation of f(x) can be expressed as the sum of N^2 cross-correlation terms,

$$f(x)*\tilde{f}(x) = \sum_{m=1}^{N} \sum_{n=1}^{N} f_m(x)*\tilde{f}_n(x)$$

where * denotes convolution and

$$\tilde{f}(x) = \overline{f(-x)} .$$

The cross correlation term $f_m(x)*\tilde{f}_n(x)$ has support contained within I_m-I_n. Of these N^2 terms, N of them have m=n and are centered at the origin. The remaining N^2-N cross-correlation terms, for which $m\neq n$, are centered elsewhere. The separation condition is equivalent to requiring that none of those N^2-N cross-correlation terms overlap with each other or with the terms centered at the origin.

Let F, G, F_n, G_n be the Laplace transforms of f, g, f_n and g_n, respectively. Let Z(F) be the set of non-real zeroes of F and define Z(G), $Z(F_n)$ and $Z(G_n)$ similarly.

Let

$$B = \bigcap_{n=1}^{N} Z(F_n),$$

that is, B is the set of non-real zeroes common to all the F_n. Finally, let w=u+iv be a variable in the complex plane and define

$$F^*(w) = \overline{F(\bar{w})} .$$

The functions G^*, F_n^* and G_n^* are defined similarly.

Theorem: If the supports of f and g satisfy the same separation condition defined above and $f_n \not\equiv 0, n=1,\ldots,N$, and if $|F(u)|=|G(u)|$ for all real numbers u, then there exist a real number c, a complex number C with $|C|=1$ and a strictly positive integer-valued function α defined on a set $B_o \subseteq B$, such that for $N\neq 2$,

a) $G_n(w) = C\, e^{icw}\Phi(w)F_n(w)$, $n=1,\ldots,N$; and for N=2 either a) holds or

b) $G_1^*(w) = C\, e^{icw}\Phi(w)F_2(w)$ and $G_2^*(w) = C\, e^{icw}\Phi(w)F_1(w)$, where

$$\Phi(w) = \prod_{z\in B_o}\left(\frac{1-\frac{w}{\bar{z}}}{1-\frac{w}{z}}\right)^{\alpha(z)}$$

The integer $\alpha(z)$ determines how many zeroes at location z are being flipped. The proof of this theorem is in [1].

If $B=\phi$, i.e., there are no non-real zeroes common to all the F_n, then $\Phi\equiv 1$ and conclusion a) of the theorem becomes

a') $G_n(w) = Ce^{icw} F_n(w)$, $n=1,\ldots,N$

and conclusion b) becomes

b') $G_1^*(w) = C\, e^{icw} F_2(w)$ and $G_2^*(w) = C\, e^{icw}F_1(w)$.

In either case it follows that $f\approx g$ if $B=\phi$. This proves the following corollary.

Corollary: If $B=\phi$ and $|F(u)|=|G(u)|$ for all real numbers u, then $f \approx g$. We note that, since $|F(u)|^2$ and $|G(u)|^2$ are analytic functions, the condition that $|F(u)|=|G(u)|$ for all real u is implied by the condition that this equality hold for all u in some open interval.

4. CONCLUSIONS

If f has $N \geq 2$ separated parts contained within a set A satisfying the separation condition and f is gotten more or less randomly from the real world, then the set B will almost always be null. That is, it is unlikely that the Laplace transforms of the separated parts of f will have non-real zeroes common to all the parts. Thus, we may conclude in this case of functions with sufficiently separated parts that the phase retrieval problem almost always has a unique solution among functions having support contained within A.

Note, however, that our earlier counterexample [2] demonstrates that even when the separation condition is satisfied for f and $B=\phi$, there can be non-equivalent solutions having supports not contained in the set A. Only by specifying a stronger separation condition and requiring f to be real and non-negative can one insure that f is unique among all non-negative functions of compact support. Specifically, it can be shown [1] for N=2 that if [-d,d] is the smallest closed interval containing the support of the autocorrelation of f, which support is also contained within $[-d,-d/2] \cup (-d/3,d/3) \cup [d/2,d]$, and if $B=\phi$, then f is unique among non-negative functions.

It should also be noted that since a two-dimensional analog of the zero-flipping theorem of Hofstetter and Walther [3] does not presently exist, these results do not automatically extend to the 2-D case. However, from other considerations, both theoretical [4] and experimental [5], it appears that the probability of uniqueness is very high for 2-D functions of compact support, even when the support is not disconnected.

ACKNOWLEDGEMENT

This work was supported by the U.S. Air Force Office of Scientific Research.

REFERENCES

1. T.R. Crimmins and J.R. Fienup, "Phase retrieval for functions with disconnected support," in Appendix B of J.R. Fienup, "High resolution imaging of space objects," interim scientific report to Air Force Office of Scientific Research, ERIM Report No. 145400-7-P (Environmental Research Institute of Michigan, Ann Arbor, MI., March 1981).
2. T.R. Crimmins and J.R. Fienup, J. Opt. Soc. Am. 71, 1026-1028 (1981).
3. E.M. Hofstetter, IEEE Trans. Inf. Theory IT-10, 119-126 (1964); A. Walther, Opt. Acta 10, 41-49 (1963).
4. Yu. M. Bruck and L.G. Sodin, Opt. Commun. 30, 304-308 (1979).
5. J.R. Fienup, Opt. Eng. 18, 529-534 (1979).

Mathematical Results of the Phase Retrieval Problem for Bandlimited Functions of Several Variables

by

Wayne M. Lawton
Jet Propulsion Laboratory, Mail Stop 264-647
4800 Oak Grove Drive, Pasadena, California 91109

The purpose of this paper is to answer the following questions concerning bandlimited functions $F(x)$ and $G(x)$ of a vector variable $x = (x_1,\ldots,x_N) \in R^N$.

Question 1. If $|F(x)|$ has rotational symmetry, what properties of $F(x)$ are implied?

Question 2. If $|F(x)| = I_1(x_1)\ldots I_N(x_N)$, (separability) what properties of $F(x)$ are implied?

Question 3. If $|F(x)| = |G(x)|$, what relationship between the supports of the Fourier transforms of $F(x)$ and $G(x)$ is implied?

Question 4. What is a general condition on $F(x)$ such that if $|F(x)| = |G(x)|$ then $F(x) = d \exp(i < y, x>)G(x)$ or $F(x) = d \exp(i < y, x >) \bar{G}(x)$ for some $|d| = 1$, $y \in R^N$ where $<y, x>$ denotes the inner product of y and x?

Questions 1-3 were asked in reference [1]. Question 1 is completely answered by the following results from reference [2].

Answer 1. If $|F(x)|$ has rotational symmetry and $N \geq 2$ then $F(x)$ can be expressed as

$$F(x) = P(x) \exp(2\pi i < y, x >) \prod_{K=1}^{\infty} \left[1 - \frac{< x, x_K >}{\lambda_K^2} \right] \quad (1)$$

where $y \in R^N$ and λ_K are the complex roots of $F(\omega,\ldots,\omega)$ (bandlimited functions can be extended to be analytic functions on C^N) and $P(x)$ is a polynomial having either the form (a) if $N=2$, then

$$P(x_1,x_2) = A(x_1 + ix_2)^{m_1}(x_1 - ix_2)^{m_2}$$

for integers $m_1, m_2 > 0$ and $A \in C$ or (b) if $N \geq 3$, then $P(x) = A(<x, x>)^m$ for some integer $m > 0$ and some $A \in C$. If $N = 1$ then no 'symmetry property' for $F(x)$

can be implied as demonstrated by the following example

$$F(x)=\int_{-.5}^{.5} \exp(-2\pi y)\ \exp(-2\pi i x y)dy = \frac{\operatorname{Sin}\pi(x-i)}{\pi(x-i)} = F(o)\ \prod_{K=1}^{\infty}\left(1-\frac{x}{i+K}\right)\left(1-\frac{x}{i-K}\right) \quad (2)$$

Answer 2. If $|F(x)| = I_1(x_1)\ldots I(x_N)$ then $F(x) = F_1(x_1)\ldots F_N(x_N)$ where each $F_K(x_K)$ is bandlimited.

The proof of answer 2 like the proof of answer 1 depends on the result in reference [3] that the complex roots of a bandlimited function of one variable are determined up to conjugation by its modulus and the result in reference [4, pages 152-154] that the complex roots of an analytic function of one variable vary continuously with the analytic function. By translating F(x) we may assume $F(o,o,\ldots,o) \neq 0$. We prove the result for N=2. Let $\lambda_K(x_1)$ be the complex roots of the function $h(z) = F(x_1,z)$. Then

$$F(x_1,x_2) = F(x_1,o)\ \exp\left(x_2\ \frac{\partial}{\partial x_2}\ \ln\ F(x_1,o)\right)\ \prod_{K=1}^{\infty}\left[1-\frac{x_2}{\lambda_K(x_1)}\right]\exp\left(\frac{x_2}{\lambda_K(x_1)}\right) \quad (3)$$

Since the $\lambda_K(x_1)$ are continuous and each $\lambda_K(x_1)$ is a root of $I(x_2)$ whose set of roots must form a discrete sequence, each $\lambda_K(x_1)$ is constant, hence $\lambda_K(x_1) = \lambda_K$ and the factor on the right is a function $H_2(x_2)$ of x_2 only. A similar expression may be obtained using $H_1(x_1)$. Manipulating the resulting functional equation yields

$$\frac{\partial}{\partial x_2}\ \ln\ F(x_1,o) = \frac{\partial}{\partial x_2}\ \ln\ F(o,o) + dx_1 \text{ for some constant d.}$$

However, since $F(x_1, x_2)$ has exponential growth in both x_1 and x_2 (when extended to a function on C^2) d = o. Therefore, $F(x_1, x_2) = F(x_1,0)\ F(o,x_2)$ and we are done.

Answer 3. Let D(F) and D(G) be the smallest closed convex sets containing the closures of the inverse Fourier transforms of F and G. Then $|F(x)| = |G(x)|$ implies $D(F) - D(F) = D(G) - D(G)$. However, the following example proves that it may be the case that D(F) and D(G) need not have the same shape. Let $F_1(x)$ be the Fourier transform of the characteristic function of any triangular region in R^2. Let $F(x) = F_1(x)\ F_1(x)$ and let $G(x) = F_1(x)\ \overline{F_1(x)}$. Then D(F) is a triangular region but D(G) is a hexagonal region.

The proof of answer 3 for N=1 follows from the result in reference [5, Theorem 4] which states that if n(r) is the number of roots λ of a bandlimited function F(x) of one variable such that

$$|\lambda| \leq r \text{ then } \lim_{r \to \infty} \frac{n(r)}{r} = 2L$$

where L = length of the smallest closed interval containing the support of the inverse Fourier transform of F(x). The answer is extended for $N \geq 2$ using the Plancherel-Polya theorem in reference [6, page 353] together with the result in reference [7, page 7] that the support function of a closed convex set uniquely determines the set. An extensive treatment of the problem of reconstructing D(F) from D(F) - D(F) is found in reference [8].

Answer 4. If F(x) and G(x) are related as in question 4 then write $F(x) \sim G(x)$. Let V(F) be the zero set of F(Z) in C^N. Then the following condition on V(F) implies that $F(x) \sim G(x)$ whenever $|F(x)| = |G(x)|$:
1) there exists $V_o \subset V(F)$ such that V_o is dense in V(F) and if $Z, \omega \in V_o$ then there exits a continuous function g: $[0,1] \to V(F)$ with $g(o) = Z$, $g(1) = \omega$ and gradient $F(g(t)) \neq (o, o, \ldots, o)$ for every t. In particular, if

$$F(x_1,\ldots,x_N) = \prod_{K=1}^{N} (\text{Sinc } x_K)\, P(e^{2\pi i x_1},\ldots,e^{2\pi i x_N})$$

where P(Z) is an irreducible polynomial then F(x) is bandlimited and V(F) satisfies condition 1. Also, the condition holds if F(x) is the Fourier transform of any triangular planar region. The proof will be omitted for lack of space.

References

[1] Kiedron, P., "On the 2-D Solution Ambiguity of the Phase Recovery Problem", to appear in Optik.

[2] Lawton, W., "Uniqueness Results for the Phase-Retrieval Problem for Radial Functions", J. Opt. Soc. Am., Vol. 71, No. 12, pp 1519-1522, (Dec 1981).

[3] Walther, A., "The Question of Phase Retrieval in Optics", Opt. Acta, Vol. 10, pp 41-49, (1963).

[4] Ahlfors, L. Complex Analysis, McGraw-Hill, New York, (1979).

[5] Titchmarsh, E., "The Zeros of Certain Integral Functions", Proc. Land. Math. Soc., Vol. 25, pp 283-302, (1926).

[6] Fuks, B., Introduction to the Theory of Analytic Functions of Several Complex Variables, Vol. 8, Translations of Mathematical Monographs, American Mathematics Society, Providence, R.I., (1963).

[7] Ronkin, L., Introduction to the Theory of Entire Functions of Several Variables, Translations of Mathematical Monographs, Vol. 44, American Mathematical Society, Providence, R.I., (1974).

[8] Fienup, J., Crimmins, T., Holsztynski, W., "Reconstruction of the Support of an Object from the Support of its Autocorrelation", J. Opt. Soc. Am., Vol. 72, pp. 610-624, (May 1982).

Maximum Entropy Image Reconstruction from Phaseless Fourier Data

John Skilling

Dept. of Applied Mathematics and Theoretical Physics,
Silver Street, Cambridge, England

We investigate the reconstruction of a real and positive spatial pattern or "image"

$$(1) \qquad f_j = \sum_{k=0}^{N-1} F_k \exp(2\pi ijk/N) \quad , \quad j=0,1,2,\ldots,N-1$$

from incomplete phaseless Fourier data D_k with noise σ_k,

$$(2) \qquad D_k = |F_k|^2 \pm \sigma_k$$

The first step is to define the set of "feasible" images, any of which is consistent with the data. This involves comparing the actual data D_k with the simulated data $|F_k|^2$ which would be observed (apart from noise) from a trial image f. The simplest comparison measure is chisquared

$$(3) \qquad \chi^2(f) = \sum(|F_k|^2-D_k)^2/\sigma_k^2$$

Any trial image f for which $\chi^2 > M+3.3\sqrt{M}$ (M = number of data) is rejected with 99% confidence: the surviving images are feasible and only these need be considered further. In N-dimensional image-space, the feasible set forms a 2M-dimensional toroid, projected linearly to infinity in any unmeasured Fourier planes. Much of the difficulty encountered with phaseless data stems from the connected topology of this constraint.

Maximum entropy is a formidably powerful and general optimal way of reconstructing positive images from a wide variety of types of data. It selects that single feasible image which has greatest entropy

$$(4) \qquad S(f) = -\sum_j p_j \log p_j \quad , \quad p_j = f_j / \Sigma f$$

This corresponds (Gull & Skilling 1982) to a maximally non-committal answer to the question "Where would the next photon come from?". In selecting that image with minimal configurational information I=-S, it confers many advantages. For example (Gull & Daniell 1978), there must be evidence in the data for any structure which is seen. Noise is automatically suppressed in the reconstruction. Instrumental artefacts such as sidelobes are also suppressed. The resulting image is uniquely easy to comprehend.

An algorithm (Skilling 1982) has been constructed for phaseless data which reliably maximises the nonlinear entropy function S(f) over the feasible set for realistic 2-dimensional images. Its results are true local maxima of S, and appear to be the desired global maxima. Reconstructions of 120x120 images from two simulated datasets are presented here.
One starting image (Fig. 1) was a set of point sources. The other (Fig. 2) was a diffuse object containing some sharper structure(M87 galaxy).

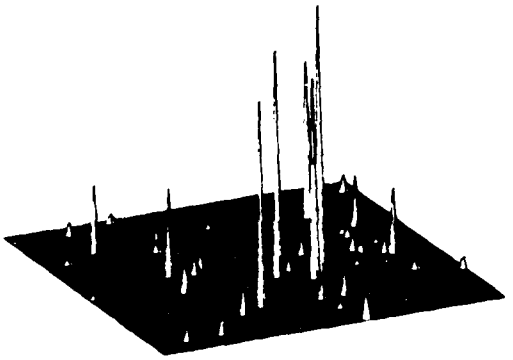


Fig. 1. Point sources

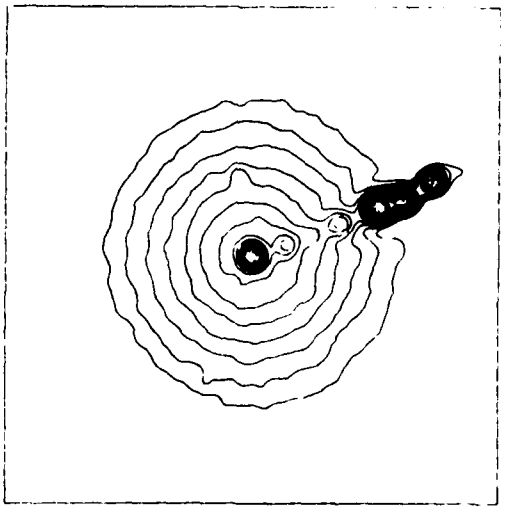


Fig. 2. Diffuse object (M87)

Each starting image was Fourier transformed to determine its amplitudes. All the high-frequency amplitudes were discarded and random errors were added to the remaining ones. Thus each dataset consisted of 3540 surviving low-frequency amplitudes, subject to noise at a typical level of 1-2%, together with the total flux Σf itself. Their phase-zero maps $f_{ij} = \Sigma D_k^{-1} \exp(2\pi ijk/N)$ (Figs 3 and 4) present these data as images.

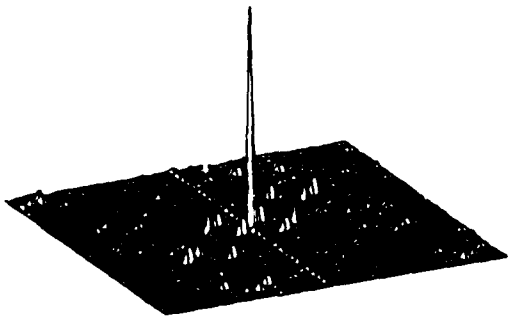


Fig. 3. Phase-zero map of amplitude data from Fig. 1

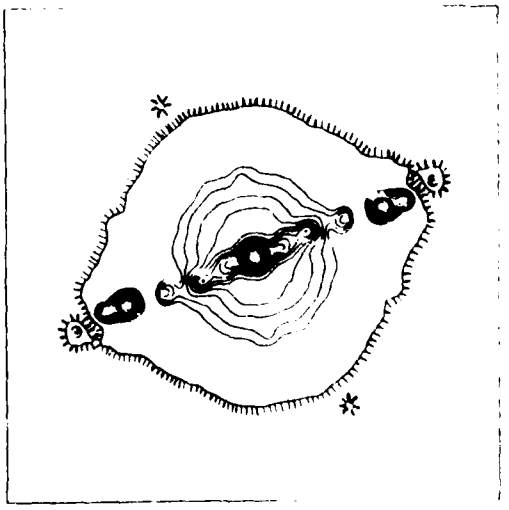


Fig. 4. Phase-zero map of amplitude data from Fig. 2

The maximum entropy reconstruction from the first dataset is shown in Fig. 5. Translated to fit the position of the original image Fig. 1, the agreement is excellent. All the brighter sources have been recovered in exactly the right places, and the fluxes are correct to about 0.1% of maximum. The only noticeable difference is that the faintest sources cannot be seen: there was no evidence for them in the data, so maximum entropy did not reproduce them. As expected, the reconstruction (Fig. 5) has an entropy S=3.227 slightly greater than the entropy S=3.178 of the original (Fig. 1), which of course also fits the data and is itself a feasible reconstruction. Since the reconstruction is so close to the original it seems reasonable to assume that the reconstruction is effectively unique: sparse point sources can be reproduced correctly.

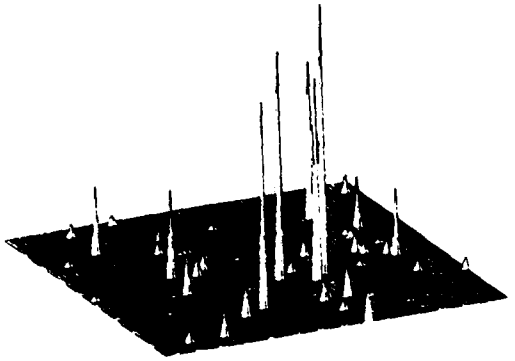


Fig. 5. Maximum entropy reconstruction of point sources

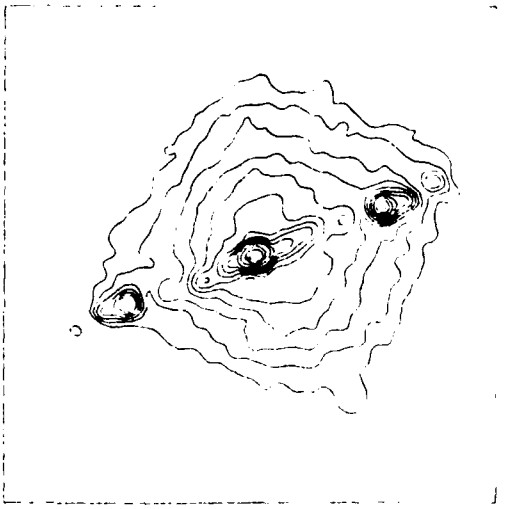


Fig. 6. Maximum entropy reconstruction of diffuse object

The maximum entropy reconstruction from the second dataset is shown in Fig. 6. This is not a good reconstruction of the original image (Fig. 2). The large-scale radial brightness distribution of the background galaxy is reproduced tolerably well. The bright spots in the jet of the galaxy are reproduced as such, but in symmetrised locations. Presumably the presence of high-frequency Fourier amplitudes has sufficed to show that the spots must be spots, but their low-frequency amplitudes have been confused with the background galaxy. Accordingly, the spots have been localised in a maximally non-committal fashion, symmetrised roughly as in the phase-zero image. Maximum entropy only reproduces sharp structure when there is good evidence for it in the data, and indeed the spots fail to reproduce quite as sharply as they are present in the original. Some of their high-frequency structure has been carried over into the background galaxy, giving it the somewhat irregular shape which is otherwise surprising in a maximum entropy image.

Nevertheless, the reconstruction (Fig. 6) is a true maximum entropy solution. Its entropy S=9.233 is greater than the entropy S=9.189 of the original. There is no evidence in the data for more structure than in Fig. 6. In particular, there is no evidence for significant asymmetry. There are, however, other local maxima. If the algorithm is started near the original (Fig. 2) instead of near a flat image, it converges to a local maximum entropy solution very similar to that original with correspondingly similar entropy.

An algorithm has been constructed which can reliably generate maximum entropy reconstructions from pure phaseless data, with no approximations and no prior assumptions. Its operation is undeniably expensive. The reconstruction Fig. 5 needed 5000 Fourier transforms, whereas the reconstruction of Fig. 6 needed 470 transforms.

Also, the user may not always be satisfied with the results. Fig. 6, for example, was suspiciously symmetrical. With experience of phaseless reconstructions to guide him, the user might well prefer to see an image in which the spots are distributed asymmetrically. The natural way of codifying this preference would be to construct a specific numerical model m_j, which could itself be refined by maximum entropy. The generalised entropy formula (Jaynes 1968)

$$S = -\sum_j p_j \log(p_j/m_j) \qquad (4)$$

can be used for this. Maximising the generalised entropy corresponds to seeking a maximally non-committal answer to the question "Where would the next photon come from, given a prior prejudice m_j about the radiation pattern?". It will produce the feasible positive image whose structural difference from the model is least. The algorithm can be generalised appropriately, and this may be a productive future development.

Another useful development will be the extension to three dimensions for crystallography (Collins 1982).

References

Collins D G, 1982, Nature 298 49

Gull S F, Daniell G J, 1978, Nature 272 686

Gull S F, Skilling J, 1982, submitted to J Opt Soc Am

Jaynes E T, 1968, IEEE Trans SCC-4 227

Skilling J, 1982, submitted to Mon Not R astr Soc

SPECKLE INTERFEROMETRY: ONE-DIMENSIONAL IMAGE RECONSTRUCTION FROM ZEROS OF COMPLEX SPECTRUM

Yuri M. Bruck and Leonid G. Sodin

Division of Radio Astronomy at the Institute of Radiophysics and Electronics, Academy of Sciences of the Ukrainian SSR, Kharkov, 310085 USSR

The method suggested by Labeyrie [1] allows reconstruction of the Fourier spectrum modulus to be performed for images scattered by a random atmosphere. Ultimately one is able to determine the autocorrelation function of the image, yet not the image itself. The method is based on the relation

$$\tilde{I}(\omega) = \tilde{O}(\omega)\,\tilde{S}(\omega) + \tilde{N}(\omega)$$

where $\tilde{I}(\omega)$, $\tilde{O}(\omega)$, $\tilde{S}(\omega)$ and $\tilde{N}(\omega)$ are, respectively, Fourier transforms of the speckle image $i(x)$, brightness distribution of the object $o(x)$, scattering function of the atmosphere $s(x)$ and additive noise $n(x)$. We will further assume that all the functions involved are specified by discrete sets of sample readings i_k, o_k, s_k and n_k, and go over to a polynomial representation of the spectra by setting $\exp(i\omega) = z$, $|z| = 1$.

$$I(z) = O(z)S(z) + N(z). \qquad (1)$$

Obviously enough, the object of our interest, i.e. $O(z)$ is completely defined by the complex set of its zeros, z_{ok}. Provided the noise level is low, the set of zeros of $I(z)$ (z_{ik}) will contain all the zeros of the object plus those of the atmosphere polynomial $S(z)$. If we dispose of a considerable number of speckle-images, then the random zeros z_s can be eliminated and the effect of noise greatly reduced by averaging moduli of the corresponding polynomials. In other words, we can expect the minima of

$$\overline{|I(z)|^2} = \overline{|O(z)S(z) + N(z)|^2} \qquad (2)$$

to rather closely coincide with the object's zeros z_{os}. Analysis of the impact of noise has shown that the minima of $|I(z)|^2$ would be closest to the zeros of $O(z)$ if the result of averaging of $|I(z)|^2$ were normalized, with the normalizing factor $\langle |N(z)|^2 \rangle$ (mathematical expectation) calculated on the assumption that

$$N(z) = \sum_{k=0}^{m-1} n_k z^k$$

where n_k are independent, normally distributed values with the variance σ^2. Then

$$\langle |N(z)|^2 \rangle = \sigma^2 \frac{|z|^{2m} - 1}{|z|^2 - 1} \qquad (3)$$

Thus, the procedure we suggest consists in using the measured results to determine the complex function

$$f(z) = \frac{\overline{|I(z)|^2}}{\langle |N(z)|^2 \rangle} - 1$$

which is, by definition, equal to

$$f(z) = |O(z)|^2 \frac{\overline{|S(z)|^2}}{\langle |N(z)|^2 \rangle} + \frac{\overline{|N(z)|^2}}{\langle |N(z)|^2 \rangle} - 1 . \qquad (4)$$

The displacement of $f(z)$ minima with respect to the zeros of $O(z)$ is rather small with even high levels of noise. The restoration accuracy of the z_{ok} positions depends substantially on their coordinates in the complex plane. The farther is $|z|$ from I and the nearer $\arg z$ to π, the greater is the effect of noise. Physically this is quite understandable as zeros with large or small magnitudes of $|z_o|$ correspond to images with highly contrasting details where smaller details can be easily "masked" by noise. The zeros with $\arg z$ close to π correspond to images whose details approach the resolution limit of the telescope, hence are suppressed by the scattering function $S(z)$.

The numerous model calculations that have been performed show the technique to provide a satisfactory restoration of $O(z)$ zeros within the complex sector $0.5 \le |z| \le 2$, $0 \le \arg z \le 5\pi/6$. Outside this sector the noise levels still allowing restoration are unrealistic.

The stages of function reconstruction are as follows.

I. The speckle images are covered with a rectangular mesh whose step size is matched with the resolution; the number of nodes, m, corresponds to the size of the scattered image.

2. From all the images readings i_k are taken at the mesh nodes ($0 \le k \le m-1$).

3. The polynomials are calculated

$$I(z) = \sum_{k=0}^{m-1} i_k z^k = \sum_{k=0}^{m-1} i_k |z|^k \exp(ik \arg z) .$$

It is convenient to specify $\arg z$ at equally spaced points $0, \pi/m, 2\pi/m, \ldots$ etc. taking $M=(2 \div 4)\,m$ and making use of the FFT algorithm.

4. The values of $|I(z)|^2$ obtained are averaged over the individual images, then the function $f(z)$ is calculated according to Eq.(4) and employed to restore the z_{ok} zeros. The multiplicity of a zero can be determined from the derivative of $f(z)$. The function readings corresponding to the object sought for are obtained through calculating coefficients of the $\prod(z - z_{ok})$ polynomial.

Consider an example where the initial object consisted of 5 readings, viz. I, 2, 4, 4 and 4. The zeros of $O(z)$ are $z_{1,2} = \pm i\sqrt{2}$ and $z_{3,4} = \sqrt{2}\exp(\pm i\,135°)$. The speckle-images after "scattering in the atmosphere" consisted of up to 32 readings (shown in Fig. 1 are two realizations and the doubled r.m.s. value of the additive noise). Generated were three groups of images, each of 200. Application of the above described procedure to each of the groups yielded the $z_{1,2} = \pm i\sqrt{2}$ zeros practically exactly. As for the $z_{3,4}$, we had

I.4I4 $\exp(\pm i\,135°)$ in the first group,
I.600 $\exp(\pm i\,135°)$ in the second group,
and I.385 $\exp(\pm i\,135°)$ in the third group.

Accordingly, the object readings obtained were

I; 2.0I; 4.03; 4.02 and 4.I5 in the first group,
I; 2.28; 4.56; 4.52 and 5.I2 in the second group
and I; I.96; 3.92; 3.92 and 3.84 in the third group.

If $|I(z)|^2$ were averaged over all the 600 realizations, the object could be reconstructed exactly.

The accuracy of $|I(z)|^2$ minimum localization is illustrated in Fig. 2 for a two-point object (the zeros are $\pm i$). Shown are two speckle-images and $|I(z)|^2$ "profiles" for $|z| = 1$ and $\arg z = 90°$; the noise level is also indicated.

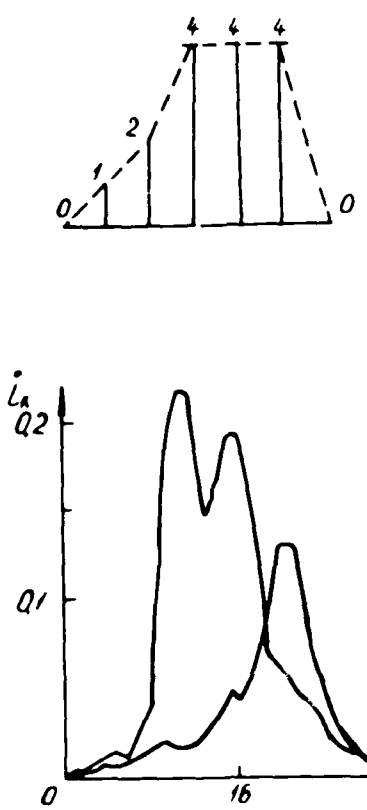


Fig. 1

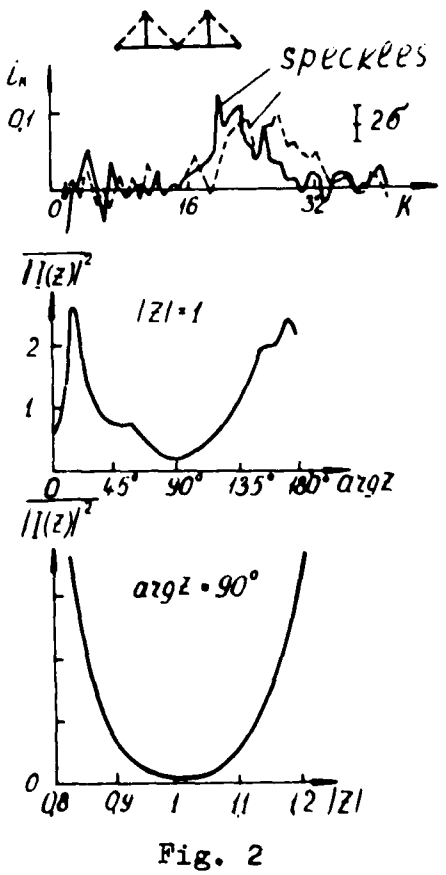


Fig. 2

A more detailed description of the method can be found in paper [2] . It might be worth noting that the technique can be regarded as a method of finding the greatest common divisor (GCD) for several realizations of $I(\vec{x})$. Note also that Euclid's algorithm permits restoration of an image from just two realizations of $I(\vec{x})$, however for very low noise.

REFERENCES

1 A. Labeyrie. Astron. and Astrophys., vol.6 (1970), 85.
2 Yu.M. Bruck and L.G. Sodin. Institute of Radio-Physics and Electronics, Acad.Sci. Ukr.SSR, Preprint No.170 (1981).

SPECKLE INTERFEROMETRY IMAGE RECONSTRUCTION TECHNIQUES PROCEEDING FROM THE PHASE OF THE FOURIER TRANSFORM

Yuri M. Bruck and Leonid G. Sodin

Division of Radio Astronomy at the Institute of Radiophysics and Electronics, Academy of Sciences of the Ukrainian SSR, Kharkov, 310085 USSR

The familiar methods of image reconstruction employed in the speckle interferometry require measurements of both the transfer function modulus [1] and phase [2] with the aid of a reference point object. In case the scattering is isotropic, the phase can be measured with a better accuracy than the modulus which makes suggestive image reconstruction from the phase alone. Hayes et al. [3] and Bruck and Sodin [4] have put forward techniques for reconstructing one- and two-dimensional images from either exact [3] or approximate [4] knowledge of the phase. In this paper we suggest reconstruction algorithms for those cases where the spectrum phase is known to a limited accuracy and no data exist as to the size and position of the image. Considered are the uniqueness and accuracy of the reconstruction, and image identification.

Thus, known from the experiment is assumed the spectrum phase $\theta(\omega)$ on the unit circle which can be written as

$$\theta(\omega) = \arg B(\omega) + l\omega + N(\omega), \quad 0 \le \omega \le \pi. \tag{I}$$

(for simplicity's sake we consider a one-dimensional case). Here $\arg B(\omega)$ is the true phase spectrum of the image sought for, l is an unknown factor arising from our lack of knowledge of the image position, and $N(\omega)$ a random phase error due to measurement inaccuracies, insufficient averaging according to [2], etc.

Similar as in our earlier work [5], we will assume the image to be specified by uniformly spaced readings b_κ taken on a discrete array. The spectrum has been extended to the entire complex plane as a polynominal $B(z)$. The spectrum phase is related to the readings through the equation

$$\arg B(z) = \arg \sum_{\kappa=0}^{n} b_\kappa z^\kappa \qquad \text{with} \quad z = r \exp i\omega. \tag{2}$$

The problem of image reconstruction can be formulated as that of finding such b_κ and l which would yield the best approximation of the right-hand side of Eq.(2) to Eq.(I). Combining the two relations one can obtain a linear set [3,4]

$$\sum_{\kappa=1}^{n} b_\kappa \sin(\kappa\omega - \theta(\omega)) = b_0 \sin\theta(\omega). \tag{3}$$

Along with these, reconstruction algorithms often make use of the relations

$$\sum_{k=1}^{n} b_{n-k} \sin(k\omega - n\omega + \theta(\omega)) = b_n \sin(n\omega - \theta(\omega)) \qquad (4)$$

and

$$\sum_{k=-n/2}^{n/2} b_{k+\frac{n}{2}} \sin(k\omega - \theta(\omega)) = b_{\frac{n}{2}} \sin\theta(\omega) \qquad (5)$$

which are different from Eq.(3) only in the position of the initial reading. Apparently, Eq.(4) corresponds to an image rotated by 180°. Eqs.(3) and (4) imply a relation $n \geq m$ between the order of the set, n, and the image size, $m+1$. The constraint of Eq.(5) is more rigid, viz. $m \leq n \leq 2m$.

If $\theta(\omega)$ is not known exactly but rather has been specified at N points, $N >> n$, the most expedient technique is the least-squares fit [4] leading to the pseudo-solution b_1:

$$S^T S b_1 = S^T y \qquad (6)$$

where S is the coefficient matrix of Eqs.(3) to (5), S^T the transposed S matrix and y the right-hand side column vector. The corresponding estimate error q is

$$q = (Sb_1 - y)^T (Sb_1 - y)/N.$$

The algorithm employs all the initially available data. The appearance of the matrices $S^T S$ and $S^T y$ allows judging on the image size without actually solving Eq.(6). The magnitude of q can serve as a measure of practical realizability of restoration. The non-uniqueness of reconstruction arises from the existence of various polynomials $B(z)$ with the same $\arg B(z)$.

As has been shown in papers [3] and [4], Eq.(6) has a unique solution if the phase $\theta(\omega)$ is known exactly, i.e. if $l = N(\omega) = 0$ and the set of zeroes of $B(z)$ does not contain mutually reciprocal values. Generally, the necessary and sufficient condition for the reconstruction uniqueness is that the polynomial of either one or two variables should not contain symmetrical factors with linear (or zero) phase.

From that point of view, employing the projection method for two-dimensional reconstruction [3] would not be a good choice because of the substantial non-uniqueness. In fact $N(\omega)$ is always non-zero and l cannot be known in advance, due to the possibility of symmetrical factors. As a result, the unique solution of Eq.(6) normally is rather far from the true one. Here we have three possibilities. The reconstructed image is closest to reality with some $l = l_0$ (where l_0 is not known!). With $l > l_0$ the image obtained through the reconstruction procedure is a convolution of an image close to the true one and a random, unstable symmetrical image. The case $l < l_0$ is the most "dangerous", as the result of reconstruction is a stable image which can yield a small net estimate error, and the negative sample readings that would inevitably be present might seem to be due to noise. Consider an illustrating example. Let an image consist of three points at 1; 3.7 and 2.1. The corresponding polynomial, i.e. $(1+3z)(1+0.7z)$ has one of its zeros inside and the other outside

the unit circle. Let us shift the polynomial leftwards by multiplying it by z^{-1} ($l=-1$). The result is $3(1+0.7z)(1+z/3)$. The power z^{-1} is not allowed by the constraints Eqs.(3) and (4). However, $\arg(1+z^{-1}/3)=\arg(1+z/3)^{-1}=\arg\left(1+\sum_{k=1}^{\infty}(-z/3)^k\right)$ with $|z/3|<1$. With $n=3$ Eq.(6) would yield a "solution" $1+0.37z-0.12z^2+0.04z^3$ characterized by the estimate error 10^{-4} . As can be seen, the negative reading is small while the "solution" obtained has nothing in common with the initial function.

Thus, with $N(\omega)\neq 0$ and l unknown each value of l would correspond to an image of its own, distinctly different from all the other . Yet for moderate noise levels algorithms to provide unique reconstruction of true images are conceivable. Here we will discuss two of such.

The first is based essentially on Eqs.(3) and (4). Before all, a linear phase shift is excluded from the phase specified in the form of Eq.(1),

$$\theta_1(\omega)=\theta(\omega)-d\cdot\omega. \tag{7}$$

This is done with the aim of having $\theta_1(\pi)=0$. Next, $\theta_2(\omega)=\theta_1(\omega)+l\omega$ is calculated for several values of n and for every l from the range $0\le l\le n$. Using Eqs.(3) and (7), all b_k/b_0 . are calculated. The procedure is repeated with Eq.(4) to obtain b_k/b_n . For all the n and l showing $q<0.5$, closest image pairs are selected either visually or by means of correlation analysis. The pair of true images is recognized with the aid of the criterium that $|\theta_2(\pi)|\to min$ and $|l|\to min$ for different n . Additional criteria for the correct choice would be the proximity of solutions obtained despite the difference in n .

The other procedure is based on using Eq.(5). It can be shown that with the condition

$$m+1\le n\le 2m \tag{8}$$

satisfied, one can always find $(n-m+1)/2$ pairs l_1 and l_2 of the linear shift parameter l for which the reconstructed images would be identical apart from a scale factor and the shift. One of such pairs would be true, corresponding to $|l_2|\to max$ and $|l_1|\to min$. In other words, if Eqs.(5) were used, true one-dimensional images would occur in pairs and two-dimensional in fours. The reconstruction algorithm using Eqs.(5) is similar to the preceding, with the constraints set by Eq.(8).

The two reconstruction procedures are illustrated by Fig.1 showing simple images that have been restored according to Eqs.(3), (4) and (5) for the case of low noise and different values of . The reconstruction of one-dimensional speckle images with different noise-to-signal ratios is shown in Fig. 2. The phase, $\theta(\omega)$ has been obtained from 200 realizations with the aid of the method described in papers [2] and [4] . As can be seen, identification of true images is no problem in either of these simple cases. The "quality" of

reconsiruction deteriorates with an increase in the noise level and for nearly symmetrical images.

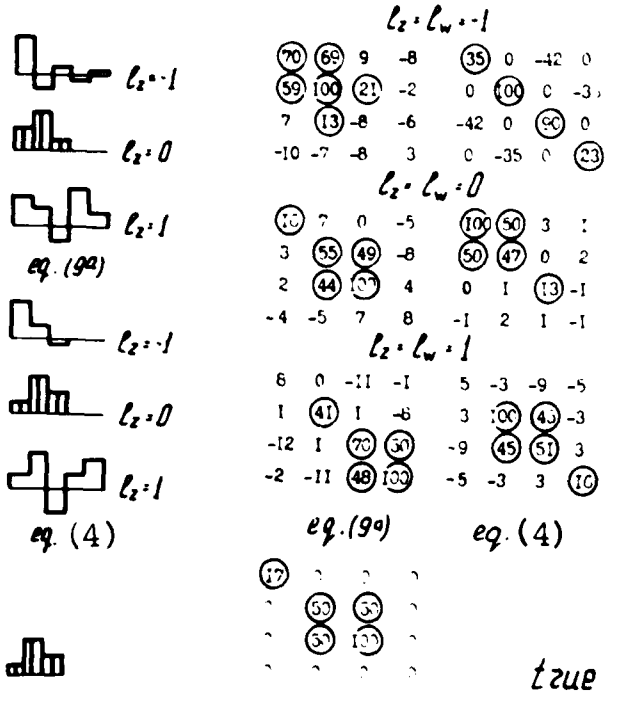


Fig. I

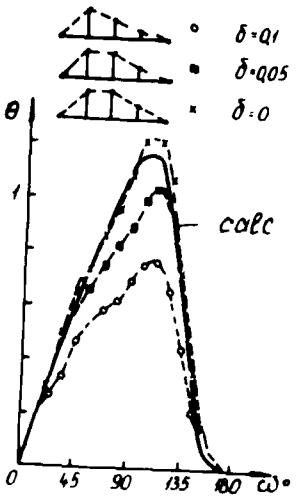


Fig.2

REFERENCES

I A.Labeyrie. Astron.and Astroph.,v.6(I970),85.
2 K.T.Knox and B.J.Thompson. Astroph.J.,v.I93(I974), pt.2, No.I, I45.
3 M.H.Hayes, J.S.Lim and A.V.Oppenheim. IEEE Trans. ASSP,v.28(I980),No.5, 472.
4 Yu.M.Bruck and L.G.Sodin. Institute of Radiophys. and Electron.,Acad.Sci.Ukr.SSR, Preprint No.I70(I98I).
5 Yu.M.Bruck and L.G.Sodin. Optics Communications,v.30 (I979), 304.

SESSION VI

RECONSTRUCTION FROM INTENSITY II

H. Ferwerda, *Presider*

ThA8-1

Phase Retrieval in Astronomy

J.R. Fienup

Environmental Research Institute of Michigan
P.O. Box 8618, Ann Arbor, MI 48107

1. Introduction

Atmospheric turbulence severely limits the resolution of images from large earth-bound optical telescopes to about one second of arc or worse, as compared with the theoretical diffraction-limited resolution of 0.02 seconds of arc for a 5-meter diameter telescope.

Several interferometric methods are capable of providing diffraction-limited information through atmospheric turbulence [1-3], the most promising of which is Labeyrie's stellar speckle interferometry [4-6]. These methods provide the modulus of the Fourier transform of the object; the phase of the Fourier transform is lost. Unfortunately, except for the very special case in which an unresolved star is very near the object of interest [7,8], the Fourier modulus can be used to directly compute only the autocorrelation of the object but not the object itself. The autocorrelation is ordinarily useful only for determining the diameter of the object or the separation of a binary star pair. The problem, then, is to reconstruct an object from its Fourier modulus (or, equivalently, from its autocorrelation function) or, equivalently, reconstruct the Fourier phase from the Fourier modulus.

This paper briefly reviews a number of proposed methods for reconstructing 2-D objects from stellar interferometry data. Emphasis will be placed on an iterative algorithm for reconstructing an object from its Fourier modulus, for which computer-simulation results will be shown.

2. Stellar Speckle Interferometry

Labeyrie's stellar speckle interferometry consists of averaging over the squared modulus of the Fourier transforms of a number of short-exposure images. Ignoring noise, let a short-exposure (frozen atmosphere) image at time t be

$$d_t(x) = f(x) * s_t(x) \tag{1}$$

where $f(x)$ is the object (x is a 2-D coordinate), $s_t(x)$ is the instantaneous point-spread function due to atmospheric turbulence and * denotes convolution. Let $D(u)$, $F(u)$ and $S_t(u)$ be the complex Fourier transforms of $d(x)$, $f(x)$ and $s_t(x)$, respectively, where $F(u) = |F(u)|\exp[i\psi(u)]$. Labeyrie's averaging yields

$$\sum_t |D_t(u)|^2 = \sum_t |F(u)\, S_t(u)|^2 = |F(u)|^2 \sum_t |S_t(u)|^2 \,. \tag{2}$$

By making measurements on an unresolved star through an atmosphere with the same statistics or by having an appropriate model for the effects of the atmosphere [9,10], one can estimate $\Sigma|S_t(u)|^2$ and divide it out from Eq. (2), then take the square root, leaving an estimate of the object's Fourier modulus, $|F(u)|$.

3. Reconstruction Methods Based on Speckle Imaging

If one utilizes all the short exposure images available in stellar speckle interferometry, then a number of reconstruction methods are possible. By the

Knox-Thompson method [11-13], one essentially averages over the phase differences between adjoining points in D(u) to arrive at an estimate of finite differences (the gradient) of $\phi(u)$. Then by averaging over the phase differences one arrives at an estimate of $\phi(u)$. Eq. (2) is used to determine the Fourier modulus. More recently log gradients have been used in a similar manner [14]. In the shift-and-add method [15,16], one averages over translated versions of $d_t(x)$, where the translations are chosen such that the brightest points in all the images are superimposed. If the object contains within it a very bright point (which must be present for the method to work) then the shift-and-add method will yield a diffraction-limited image with a fog-like bias term superimposed. An earlier version of shift-and-add was used to reconstruct an image of the giant star Betelgeuse [17].

4. Methods Based on Filtering

By detecting intensities both in an image plane and in the plane of a filtered image one can reconstruct the object function. The filtering may simply be a defocusing of the optical system. Then both the object and the atmospheric phase function can be reconstructed from a single pair of intensity measurements using an iterative method imbedded within an iterative method [18]. This unfortunately requires a bright object and is computationally burdensome. Another approach is to perform speckle interferometry in both an image plane and the plane of an exponentially filtered image [19,20]. Then the object can be reconstructed with an iterative method similar to the one described in Section 6.

5. Methods Based on Fourier Modulus

The image reconstruction methods described so far depend on processing the individual short-exposure images, but for other types of interferometry one may only have access to the Fourier modulus data. Furthermore, even when using the speckle imaging methods one would expect the estimate of $|F(u)|$ by Eq. (2) to be obtained with considerably greater signal-to-noise ratio than the reconstructed phase. For these reasons, methods relying only on $|F(u)|$ are of interest.

For special types of objects the reconstruction can be very simple. If the object has a companion unresolved star at least an object-width away but still within the same isoplanatic patch, then the autocorrelation includes a term equal to the object that can be separated out [8], analogous to holography [7]. If the object consists of a collection of unresolved stars with nonredundant spacings, then a simple reconstruction method involving the product of three translates of the autocorrelation function can be used [21]. Such objects could also be reconstructed by an iterative method [22] which was borrowed from the field of X-ray crystallography.

Some methods are applicable to any object of finite extent but are computationally difficult. One such method involves the recursive unfolding of the autocorrelation function, for which the possible solution set expands as additional equations are solved [23]. An iterative Newton-Raphson method has been shown to work [24], but it involves the inversion of a huge matrix. Another method relies on the method of tracking the complex zeroes of projections of the image [25].

A recently proposed method involves the use of "in-between" samples of the Fourier modulus to reconstruct the phase [26]. Although computationally very simple, this method remains to be demonstrated on complicated objects.

6. Iterative Fourier Transform Approach

Only one method of reconstructing an object from its Fourier modulus has proven to be practical for the case of general 2-D objects of moderate complexity. The method is related to the Gerchberg-Saxton algorithm [27,28] used for the phase retrieval problem in electron microscopy, for which one has a pair of intensity measurements. For the present problem one has only a single intensity measurement, $|F(u)|^2$, but one also has a nonnegativity constraint: $f(x) \geq 0$. In addition, from the knowledge of the autocorrelation function, which can be computed as the inverse Fourier transform of $|F(u)|^2$, one also has an upper bound on the diameter of the object [21]. The reconstruction algorithm [29-31] is depicted in Figure 1. One Fourier transforms back and forth between the object and Fourier domains, applying the measurements and a priori information available in each domain. Actually there are a few families of related algorithms that can be used. In all of them one Fourier transforms an input image $g(x)$, replaces the resulting Fourier modulus by the measured modulus, $|F(u)|$, and inverse Fourier transforms, resulting in an image $g'(x)$. At this point a number of different algorithms have been used to form a new input based on how $g'(x)$ violates the object domain's nonnegativity and diameter constraints [31]. The most obvious one is to set it equal to zero wherever it violates the constraints. One that works much better is to use

$$g_{k+1}(x) = \begin{cases} g_k'(x), & x \in OK \\ g_k(x) - \beta g_k'(x), & x \notin OK \end{cases}$$

where k denotes the kth iteration, OK denotes the set of points for which $g'_k(x)$ satisfies the object-domain constraints and β is a constant ($\beta = 0.5$ to 1.0 works well).

Figure 2 shows a computer experiment testing this reconstruction method on a realistic simulation of the Fourier modulus data provided by stellar speckle interferometry [32]. An undegraded object, shown in Figure 2(a), was convolved with 156 different point-spread functions to produce 156 different blurred images. Each point-spread function represented a different realization of the effects of the turbulent atmosphere. The blurred images were then subjected to a Poisson noise process to simulate the effects of photon noise. Two of the resulting 156 degraded images are shown in Figures 2(b) and 2(c). The degraded images were then processed by Labeyrie's method [4], and a photon-noise bias term [33] was subtracted, to arrive at the noisy Fourier modulus estimate shown in Figure 2(d). An image reconstructed from this data is shown in Figure 2(e). For objects of this complexity (in 128 x 128 arrays) for a complete reconstruction it takes about 100 iterations [31] at about one second per iteration using a Floating Point Systems AP-120B array processor.

7. Combinations of Methods

Some of the methods described above can be combined to produce a better result than what can be obtained by any one method alone. In particular, an image reconstructed with any of the other methods based on speckle interferometry or using the Fourier modulus information will usually have some negative values and may exceed half the diameter of the autocorrelation. If one then performs several iterations of the iterative Fourier transform algorithm, using the reconstructed images as the initial input, then a truer reconstruction will be obtained. One can either view the iterative Fourier transform algorithm as a post-

processing step to "clean up" the image reconstructed by another method, or view the other reconstruction method as a means for providing an initial estimate to be used by the iterative Fourier transform algorithm.

8. Uniqueness

There has been considerable controversy over the question of whether the solution to this problem is likely to be unique in the 2-D case [34-38]. It appears that although the 1-D problem is usually not unique [39], the 2-D problem usually is unique [34]; and this has been borne out by experimental results using the iterative Fourier transform algorithm [29-32,40].

9. Conclusions

Both from a theoretical viewpoint and from computer simulation results it appears that the reconstruction of diffraction-limited images of astronomical objects should be feasible, despite the turbulent atmosphere, using stellar speckle interferometry combined with the iterative Fourier transform algorithm.

Acknowledgment

This work was supported by the U.S. Air Force Office of Scientific Research.

References

1. A.A. Michelson and F.G. Pease, Astrophys. J. 53, 249-259 (1921).
2. R. Hanbury Brown and R.Q. Twiss, Nature 177, 27-29 (1956).
3. D.G. Currie, S.L. Knapp, and K.M. Liewer, Astrophys. J. 187, 131-144 (1974).
4. A. Labeyrie, Astron. and Astrophys. 6, 85-87 (1970).
5. O.Y. Gezari, A. Labeyrie, and R.V. Stachnik, Astrophys. J. Lett. 173, L1-L5 (1972).
6. A. Labeyrie, in Progress in Optics, Vol. XIV, E. Wolf, ed. (North-Holland, Amsterdam, 1976), pp. 49-87.
7. J.W, Goodman, J. Opt. Soc. Am. 60, 506-509 (1970).
8. C.Y.C. Liu and A.W. Lohmann, Opt. Commun. 8, 372-377 (1973).
9. D. Korff, G. Dryden, and M.G. Miller, Opt. Commun. 5, 187-192 (1972).
10. C. Aime, S. Kadir, G. Ricort, C. Roddier, and J. Vernin, Optica Acta 26, 575-581 (1979).
11. K.T. Knox and B.J. Thompson, Astrophys. J. Letters 193, L45-L48 (1974).
12. K.T. Knox, J. Opt. Soc. Am. 66, 1236 (1976).
13. R.V. Stachnik, P. Nisenson, et al., Nature 266, 149-151 (1977).
14. H.W. Swan and J.W. Goodman, Proc. SPIE 358-14 (August 1982).
15. R.H.T. Bates and F.M. Cady, Opt. Commun. 32, 365-369 (1980).
16. F.M. Cady and R.H.T. Bates, Opt. Lett. 5, 438-440 (1980).
17. C.R. Lynds, S.P. Worden, and J.W. Harvey, Astrophys. J. 207, 174-180 (1976).
18. R.A. Gonsalves and R. Chidlaw, Proc. SPIE 207, 32-39 (1979).
19. J.G. Walker, Optica Acta 28, 1017-1019 (1981).
20. J.G. Walker, Appl. Opt. 21, 3132-3137 (1982).
21. J.R. Fienup, T.R. Crimmins, and W. Holsztynski, J. Opt. Soc. Am. 72, 610-624 (1982).
22. J.E. Baldwin and P.J. Warner, Mon. Not. R. Astr. Soc. 182, 411-422 (1978).
23. W.J. Dallas, Optik 44, 45 (1975).
24. B.R. Frieden and D.G. Currie, J. Opt. Soc. Am. 66, 1111 (1976) (Abstract).
25. P.J. Napier and R.H.T. Bates, Astron. Astrophys. Suppl. 15, 427-430 (1974).
26. R.H.T. Bates, Optik 61, 247-262 (1982).
27. R.W. Gerchberg and W.O. Saxton, Optik 35, 237-246 (1972).

28. W.O. Saxton, Computer Techniques for Image Processing in Electron Microscopy (Academic Press, New York, 1978).
29. J.R. Fienup, Opt. Lett. 3, 27-29 (1978).
30. J.R. Fienup, Opt. Eng. 18, 529-534 (1979).
31. J.R. Fienup, Appl. Opt. 21, 2758-2769 (1982).
32. G.B. Feldkamp and J.R. Fienup, Proc. SPIE 231, 84-93 (1980).
33. J.W. Goodman and J.F. Belsher, Proc. SPIE 75, 141-154 (1976).
34. Yu.M. Bruck and L.G. Sodin, Opt. Commun. 30, 304-308 (1979).
35. W. Lawton, Proc. SPIE 231, 94-98 (1980).
36. A.M.J. Huiser and P. Van Toorn, Opt. Lett. 5, 499-501 (1980).
37. T.R. Crimmins and J.R. Fienup, J. Opt. Soc. Am. 71, 1026-1028 (1981).
38. J.R. Fienup, in Current Trends in Optics, F.T. Arecchi and F.R. Aussenegg, eds. (Taylor and Francis, London), pp. 95-102.
39. A. Walther, Optica Acta 10, 41-49 (1963).
40. J.R. Fienup, Report RADC-TR-81-63 (1981).

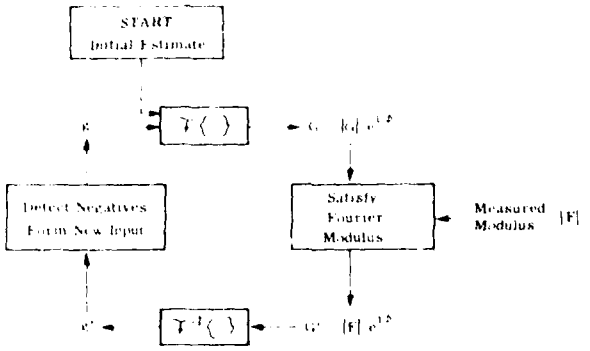


Figure 1. Iterative Fourier Transform Algorithm for Reconstructing a Nonnegative Object from the Modulus of Its Fourier Transform.

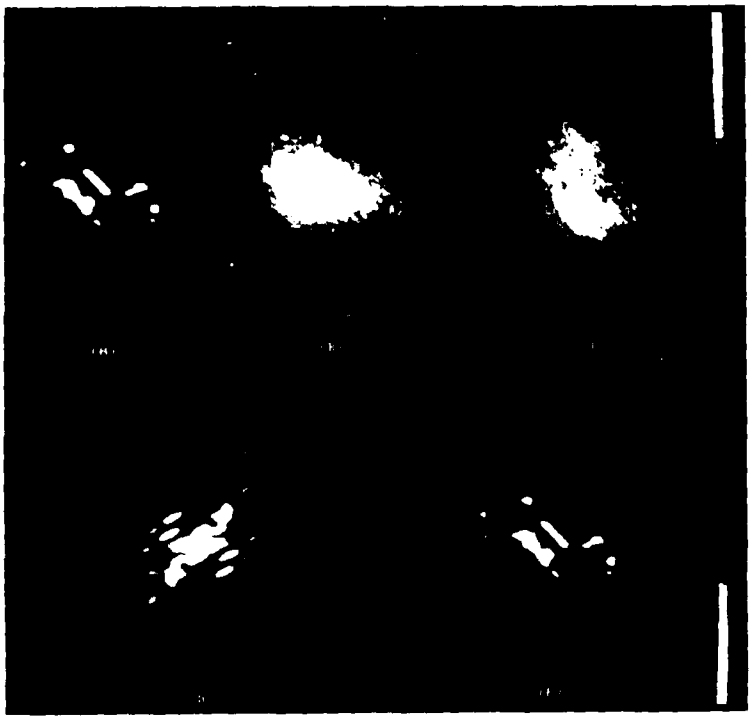


Figure 2. Reconstruction Experiment for Stellar Speckle Interferometry. (A) Undegraded object; (B),(C) simulated images degraded by atmospheric turbulence and photon noise; (D) Fourier modulus estimate; (E) reconstructed image.

Frequency Sampling of the Short-Time Fourier Transform Magnitude*

by

Thomas F. Quatieri**, S. Hamid Nawab**, Jae S. Lim***

Lincoln Laboratory
Research Laboratory of Electronics
Massachusetts Institute of Technology

Abstract

Under mild restrictions, a sequence x(n) is uniquely specified by its short-time Fourier transform magnitude at one or two frequencies for each n.

I. Introduction

The short-time Fourier transform (STFT) is a signal representation of considerable interest in a number of signal processing applications including speech processing. For a discrete time signal x(n), the STFT is defined [1] as

$$X_w(nL,\omega) = \sum_{m=-\infty}^{\infty} x(m)w(nL-m)e^{-j\omega m} \qquad (1)$$

where the subscript w in $X_w(nL,\omega)$ denotes the analysis window w(n), and where L denotes the number of samples w(n) is shifted in computing successive Fourier transforms. We will assume that both x(n) and w(n) are real, and that w(n) is N points long and non-zero for $0 \leq n \leq N-1$. Note that when L has minimum value 1, adjacent analysis window positions have maximum overlap.

---

* This work was supported in part by the Department of the Air Force and in part by the National Science Foundation under Grant #CS80-007102 and in part by the Advanced Research Projects Agency monitored by ONR under contract N0014-82-K-0742 NR-49-506. The U.S. Government assumes no responsibility for the information presented.

** T.F. Quatieri and S.H. Nawab are with the M.I.T. Lincoln Laboratory, Lexington, MA 02173.

*** J.S. Lim is with the M.I.T. Dept. of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Cambridge, MA 02139.

It has been recently established [2] that under mild conditions x(n) is uniquely specified by N+1 arbitrarily spaced samples of its STFT magnitude $|X_w(nL,\omega)|$ over the frequency interval $[0,\pi]$. These results are sufficiently general so that L ranges up to half the analysis window length. In this paper, we show that for the maximum overlap case, fewer frequency samples of $|X_w(n,\omega)|$ are sufficient for unique specification of x(n).

II. Theory

For a large class of practically important sequences two frequency samples of $|X_w(n,\omega)|$ for each n are sufficient to uniquely represent x(n). When x(n) is further constrained to be non-negative, only one frequency sample is needed. Specifically, the first more general result can be stated as:

> Let x(n) denote a real, right-sided sequence with no more than N-2 consecutive zero samples between any two non-zero samples. The sequence x(n) is uniquely specified within a sign factor by two appropriately chosen frequency samples of $|X_w(n,\omega)|$ in the interval $[0,\pi]$ for each n.

We outline a demonstration [3] of this statement by developing an algorithm that reconstructs x(n) from two appropriate frequency samples of $|X_w(n,\omega)|$. Consider the smallest value of n, say n_o, such that $x(n_o)$ is non-zero. Then from (1), $|X_w(n_o,0)| = |w(0)||x(n_o)|$. Thus, by finding the smallest n such that $|X_w(n,0)|$ is not zero, n_o can be determined. Then we have

$$x(n) = 0 \quad \text{for} \quad n < n_o \tag{2a}$$

and

$$x(n_o) = \pm \frac{|X_w(n_o,0)|}{w(0)} \neq 0 \tag{2b}$$

Note that $w(0)\neq 0$ in (2b) since we are constraining w(n) to be nonzero over its duration. The two solutions for $x(n_o)$ represent the sign ambiguity, and $x(n_o)$ can be chosen to be either of the two solutions. Assuming we choose the correct sign, we outline how x(n) can be recursively determined for $n>n_o$. Noting that w(n) is zero outside $0\leq n\leq N-1$, with L=1 (1) can be rewritten as

$$X_w(n,\omega) = Y(n,\omega) + x(n)w(0)e^{-j\omega n} \tag{3a}$$

where

$$Y(n,\omega) = \sum_{m=n-N+1}^{n-1} x(m)w(n-m)e^{-j\omega m} \tag{3b}$$

Taking the squared magnitude on both sides of (3a),

$$x^2(n) + b(n,\omega)x(n) + c(n,\omega) = 0 \tag{4a}$$

where

$$b(n,\omega) = 2\,\frac{\mathrm{Re}[Y(n,\omega)e^{j\omega n}]}{w(0)} \tag{4b}$$

$$c(n,\omega) = \frac{|Y(n,\omega)|^2 - |X_w(n,\omega)|^2}{w^2(0)} \tag{4c}$$

Note that both $b(n,\omega)$ and $c(n,\omega)$ can be determined from $x(m)$ for $m<n$. When (4) is solved for $x(n)$, there are two solutions for each value of ω. Consider two distinct values of ω, say ω_1 and ω_2 in the interval $[0,\pi]$. Since $x(n)$ is assumed not to contain more than N-2 consecutive zero samples between the first and last non-zero samples, it is possible to show [3] that one can always find an ω_1 and ω_2 for which $b(n,\omega)$ and $c(n,\omega)$ in (4) are not the same. From the properties of quadratic equations, then, the two solutions associated with ω_1 cannot be the same as the pair of solutions for ω_2. However, one of the solutions must be identical and that is the true value of $x(n)$. From the above, once $x(n_o)$ is determined from (2), $x(n)$ can be recursively determined by solving (4) at two appropriate frequencies for $n>n_o$. This allows us to determine $x(n)$ for all n. As we continue the above procedure, if we find $x(n)=0$ for N-1 consecutive points, $x(n)=0$ from that point on. If we choose the wrong sign in (2), the reconstruction procedure yields $-x(n)$. Thus we determine $x(n)$ within a sign factor.

When $x(m)w(n-m)$ in (3) is non-negative, it is easily seen that $|Y(n,0)|^2 \leq |X_w(n,0)|^2$ so that $c(n,0)\leq 0$ in (4). Consequently, the two solutions of (4) for $\omega=0$ have different signs (except when $|Y(n,0)|^2 = |X_w(n,0)|^2$ where it can be shown that $x(n)=0$) and $x(n)$ is uniquely determined from only one frequency sample. With this observation, it is straightforward to derive conditions on $x(n)$ and $w(n)$ to uniquely represent a non-negative $x(n)$ with $|X_w(\omega,0)|$. For example, we can state the following:

> Let $x(n)$ denote a real, non-negative, right-sided sequence. Assume $w(n)$ is positive for $0\leq n\leq N-1$. Then the sequence $x(n)$ is uniquely specified by $|X_w(n,0)|$.

This result can be alternately demonstrated through a recursive reconstruction procedure which is linear, and avoids solving (4). As before, $x(n_o)$ is solved for using (2b). The sign ambiguity is resolved since $x(n)$ is known to be non-negative. To determine $x(n)$ for $n>n_o$, we evaluate (3) at $\omega=0$ to obtain

$$X_w(n,0) = \sum_{m=n-N+1}^{n-1} x(m)w(n-m) + x(n)w(0) \qquad (5)$$

Noting that x(n) is non-negative and w(n) is positive, (5) can be rewritten as

$$x(n) = \frac{|X_w(n,0)| - \sum_{m=n-N+1}^{n-1} x(m)\, w(n-m)}{w(0)} \qquad (6)$$

Clearly, this equation can be evaluated recursively to yield x(n) for $n>n_o$.

In our conditions, we have assumed that x(n) is a right-sided sequence. It should be clear however, that similar statements can be made for a left-sided sequence.

III. Implementation

In practice, the short-time Fourier transform magnitude is often computed at uniform frequency samples through the discrete Fourier transform (DFT). Therefore, from the point of view of implementing algorithms for signal recovery, it is of interest that our theory remains valid when $|X_w(n,\omega)|$ is replaced by $|X_w(n,2\pi r/M)|$ where M is the DFT length.

In particular, suppose that x(n) is a right-sided sequence with no more than N-2 consecutive zeros over its duration. Then in a style similar to our demonstration in the previous section, we can show that x(n) is uniquely specified within a sign factor by two appropriate frequency samples of the discrete-frequency function $|X_w(n,2\pi r/M)|$ provided that $M \geq 2N-2$ [3] . That is, with this condition on M, we can always find two values of $2\pi r/M$ which yield two distinct quadratics in (4).

References

[1] L.R. Rabiner and R.W. Schafer, Chapter 6, Digital Processing of Speech Signals. Prentice-Hall, 1978.

[2] S.H. Nawab, T.F. Quatieri, J.S. Lim, "Signal Reconstruction from Short-Time Fourier Transform Magnitude", to be published in IEEE Trans. Acoustics, Speech, and Signal Processing.

[3] S.H. Nawab, "Signal Estimation from Short-Time Spectral Magnitude", Ph.D. Thesis, M.I.T., June 1982.

Averaging the Fourier Phase Information in a Signal Ensemble without Calculating Phase

Herbert W. Swan†
Joseph W. Goodman

Stanford Electronics Laboratory
Information Systems
Stanford, Ca. 94305

1. Introduction A length-M discrete stochastic process, $s(k)$, and a length-N deterministic sequence, $h(k)$, are convolved to yield the discrete stochastic process, $x(k)$. If we take the length-L discrete Fourier transform of $s(k)$, $h(k)$, and $x(k)$, where L is some integer greater than $N+M-2$, and pad with zeros as necessary, then

$$X(n) = S(n)H(n) \tag{1}$$

for $0 \leq n < L$. Here $H(n)$ is deterministic while $X(n)$ and $S(n)$ are stochastic. Given an ensemble of the process $x(k)$, and sufficient knowledge of the self-statistics of $s(k)$, we wish to recover $h(k)$. Problems such as this arise in the fields of geophysics, radar signal processing, and space object imaging. Although it is easy to estimate the Fourier magnitude of $h(k)$ from

$$|H(n)| = \left[\overline{|X(n)|^2}/\overline{|S(n)|^2}\right]^{1/2}, \tag{2}$$

estimating the phase of $H(n)$ can be difficult, due to the phase unwrapping problem[1]. The difficulty is worsened by observational noise and by extending the problem to 2-dimensional images.

2. Cepstral Averaging One method of solving this problem is by averaging the logarithms of Eq. (1). Then

$$H(n) = \exp\left\{\langle \log X(n) \rangle - \langle \log S(n) \rangle\right\}. \tag{3}$$

Although this technique is simple in concept, it must be remembered that $X(n)$ and $S(n)$ are complex quantities, and that the phases of $X(n)$ must be properly unwrapped to the correct multiple of 2π. An adaptive algorithm exists to do this[2], but it is cumbersome to extend to two dimensions, and as we shall see, not really necessary.

The alternative to using Eq. (3) is to use the so-called "ramp cepstrum," defined by

$$\mathbf{C}\{x(k)\} = \frac{jL}{2\pi}\mathbf{F}^{-1}\left\{\mathbf{D}[X(n)]/X(n)\right\}, \tag{4}$$

where $\mathbf{F}$ is the discrete Fourier transform and $\mathbf{D}$ is a modified differentiation operator satisfying

$$\mathbf{D}\left\{e^{j2\pi n/L}\right\} = \frac{j2\pi n}{L} e^{j2\pi n/L}. \tag{5}$$

Since an arbitrary length-L sequence, $X(n)$, can be decomposed into a sum of complex exponentials via the discrete Fourier transform, we may use the linearity property of differentiation to write that

†This work was performed while the first author was with Sandia National Laboratories, Livermore, Ca. The second author is with Stanford University. The work was sponsored by the United States Department of Energy, and by the Air Force Office of Scientific Research.

$$\mathbf{D}\{X(n)\}=\frac{2\pi}{jL}\left[x(1)W_L^n+2x(2)W_L^{2n}+\cdots+(L-1)x(L-1)W_L^{(L-1)n}\right], \tag{6}$$

where $W_L=e^{-j2\pi/L}$, and

$$x(k)=\frac{1}{L}\sum_{n=0}^{L-1}X(n)W_L^{-nk}=\mathbf{F}^{-1}\{X(n)\}. \tag{7}$$

It is shown in [3] that this differentiation operator satisfies the product rule

$$\mathbf{D}\{S(n)H(n)\}=\mathbf{D}\{S(n)\}H(n)+S(n)\mathbf{D}\{H(n)\}, \tag{8}$$

provided the length of $s(k)*h(k)$ is less than or equal to L. Substitution of Eq. (8) into Eq. (4) yields

$$\begin{aligned}\mathbf{C}\{s(k)*h(k)\}&=\frac{jL}{2\pi}\mathbf{F}^{-1}\left\{\frac{\mathbf{D}[S(n)]}{S(n)}+\frac{\mathbf{D}[H(n)]}{H(n)}\right\}\\&=\mathbf{C}\{s(k)\}+\mathbf{C}\{h(k)\}.\end{aligned} \tag{9}$$

Consequently, we have an exact mapping of convolution into addition, which was the original motivation behind the introduction of the complex cepstrum[4]. Although the ramp cepstrum has appeared in the literature previously[2] it has never been applied to this class of problems before, nor has it been realized that it represents an *exact* mapping into finite field addition, without the subsequent integration and phase unwrapping step.

Next, we notice that Eq. (4) may be written in the form

$$x(k)\circledast\dot{x}(k)=k\,x(k), \tag{10}$$

where $\circledast$ denotes circular convolution and $\dot{x}(k)=\mathbf{C}\{x(k)\}$. This may further be written either as

$$[X_c]\hat{\mathbf{x}}=R\mathbf{x} \tag{11a}$$

or

$$[\dot{X}_c]\mathbf{x}=R\mathbf{x}, \tag{11b}$$

where $[X_c]$ is the $L\times L$ circulant representation of $x(k)$; $[\dot{X}_c]$ is the similar representation of $\dot{x}(k)$, and $\mathbf{x}$ and $\hat{\mathbf{x}}$ are $L\times 1$ vector representations of $x(k)$ and $\dot{x}(k)$, respectively. The first column of $[X_c]$ is $\mathbf{x}$, while the first column of $[\dot{X}_c]$ is $\hat{\mathbf{x}}$. R is the $L\times L$ ramp matrix, $R=\mathrm{diag}(0,1,\ldots,L-1)$. To find the *forward* ramp cepstrum of $\mathbf{x}$ we solve $\hat{\mathbf{x}}=[X_c]^{-1}R\mathbf{x}$. If $[X_c]$ is singular, then $X(n)=0$ for some n. In this case we instead use $\hat{\mathbf{x}}=[X_c]^{+}R\mathbf{x}$, where $[X_c]^{+}$ is the Moore-Penrose pseudo-inverse. To find the *inverse* ramp cepstrum of $\hat{\mathbf{x}}$, we solve for the vector $\mathbf{x}$ which minimizes

$$\lambda=\frac{\left|\left|\left([\dot{X}_c]-R\right)\mathbf{x}\right|\right|^2}{||\mathbf{x}||^2}. \tag{12}$$

The required $\mathbf{x}$ is an eigenvector corresponding to the minimum eigenvalue of $([\dot{X}_c]-R)^H([\dot{X}_c]-R)$. Note that a ramp cepstrum may be recovered only to within a constant d.c. gain factor. Fast approximate and iterative techniques to solve this problem are given in [3].

Eq. (3) now is changed to become

$$h(k)=\mathbf{C}^{-1}\left\{<\mathbf{C}\{x(k)\}>-<\mathbf{C}\{s(k)\}>\right\}. \tag{13}$$

$<\mathbf{C}\{x(k)\}>$ may be obtained from the given ensemble of data. $<\mathbf{C}\{s(k)\}>$ in general depends upon the joint statistics of $S(n)$ and is more difficult to obtain. However, if $S(n)$ is a circular complex Gaussian r.v. with nonzero variance for all n, then it may be argued[3] that

$$<\mathbf{C}\{s(k)\}> \approx \frac{jL}{2\pi}\mathbf{D}\{<|S(n)|>\}/<|S(n)|>. \qquad (14)$$

3. Experimental Results This procedure was applied to a series of 1-dimensional data. The unknown deterministic function, $h(k)$, appears in Figure (a), along with a typical realization for $s(k)$, designed to simulate a cross section of an atmospheric turbulence point spread function. A typical realization for $x(k)$, plotted with truncated edges, is shown in Figure (b). Finally, Figure (c) shows signal recovery after 25, 100 and 1000 averages of Eq. (13). In these simulations, $<\mathbf{C}\{s(k)\}>$ was obtained empirically by averaging the cepstra of signals from the same ensemble of $s(k)$. Examples of 2-dimensional signal recovery appear in [5].

References

[1]. M. A. Rodriguez, R. H. Williams, and T. J. Carlow, "Signal Delay and Waveform Estimation Using Unwrapped Phase Averaging," *IEEE Trans. Acoust., Speech, Signal Proc.* **ASSP-29**(3) p. 508 (1981).

[2]. J. M. Tribolet, "A New Phase Unwrapping Algorithm," *IEEE Trans. Acoust., Speech, Sig. Proc.* **ASSP-25**(2) pp. 170-177 (1977).

[3]. H. W. Swan, *Imaging through Atmospheric Turbulence via Cepstral Averaging*, Stanford University (1982). Department of Electrical Engineering Ph.d. Dissertation

[4]. A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, Jr., "Nonlinear filtering of Multiplied and Convolved Signals," *Proc. IEEE* **56** pp. 1264-1291 (1968).

[5]. H. W. Swan and J. W. Goodman, "Imaging through Atmospheric Turbulence via Modified Log Gradients," *Proc. SPIE* **358**(1982).

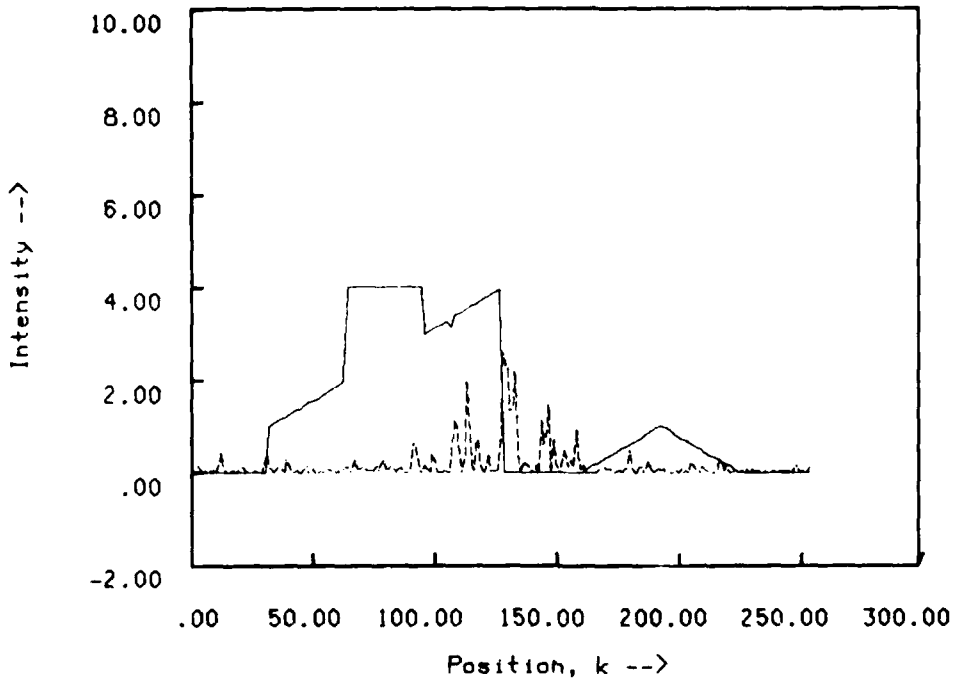


Figure (a) - h(k), solid; Typical realization for s(k), dashed

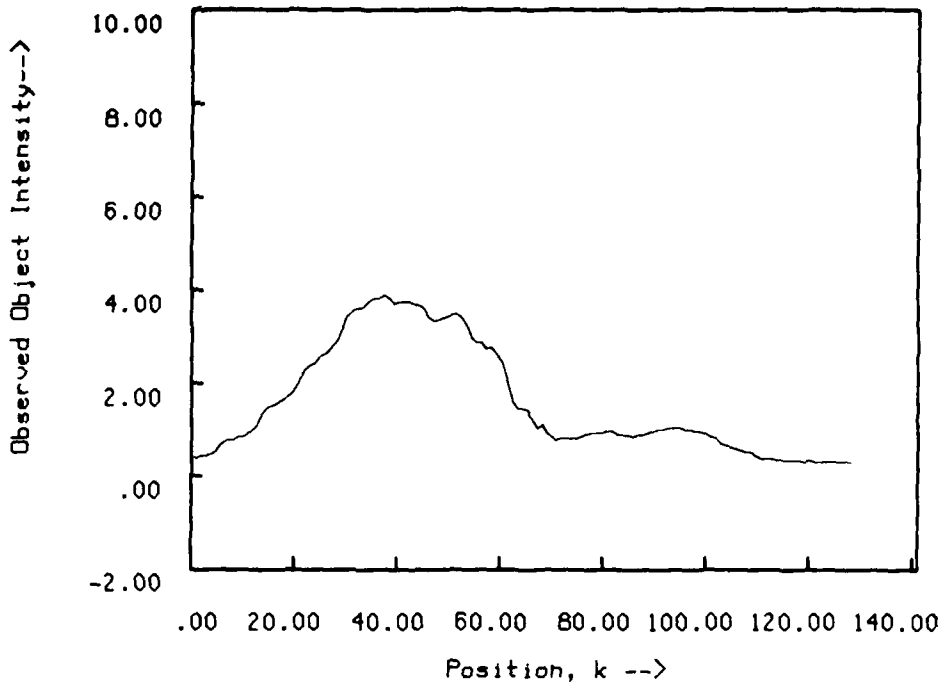


Figure (b) - Typical Realization of x(k)

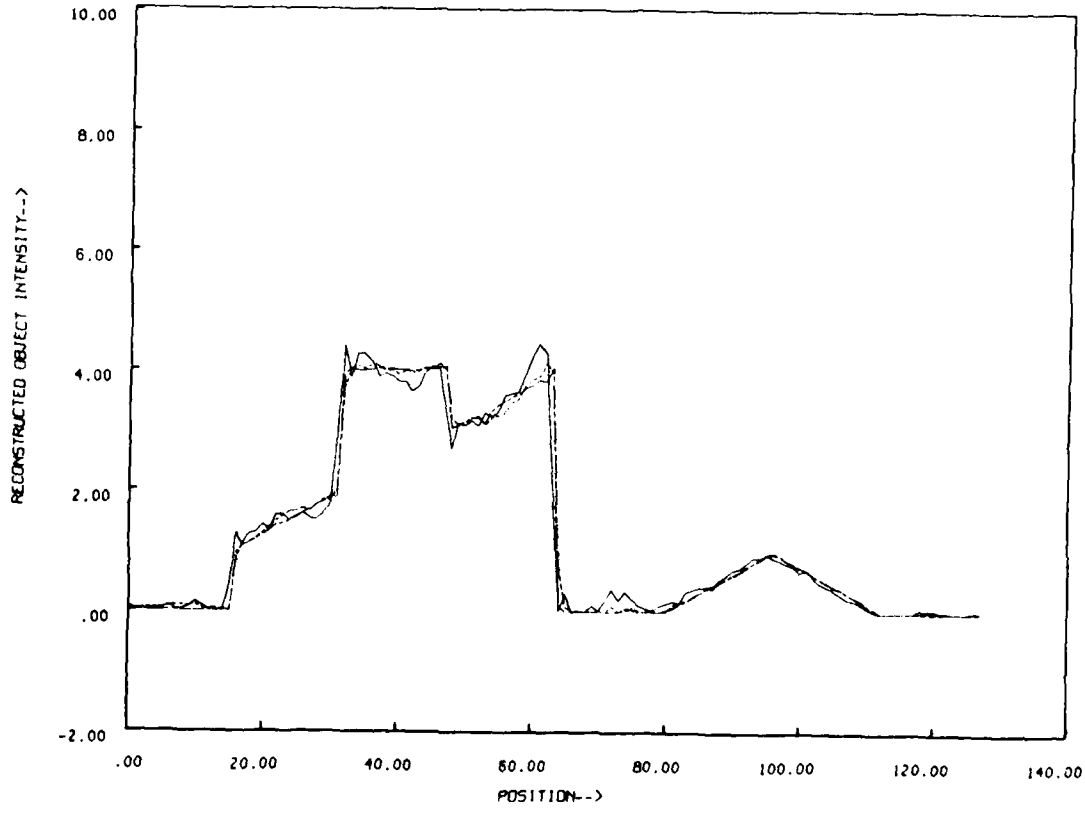


Figure (c) - Retrieved object after 25, 100, and 100 iterations

A SUFFICIENT CONDITION FOR PHASE RETRIEVAL IN TWO DIMENSIONS

M. A. Fiddy — Physics Department, Queen Elizabeth College, Campden Hill Road, London W8 7AH

B. J. Brames, J. C. Dainty — The Institute of Optics, The University of Rochester, Rochester, NY 14627.

Abstract

Eisenstein's criterion for irreducibility is used to modify an object function, thus ensuring uniqueness of phase retrieval in two dimensions.

INTRODUCTION

Phase recovery methods considered to date include the two defocus method, the Gerchberg-Saxton algorithm and Fienup's algorithm[1,2]. These iterative techniques have been compared in detail[3], but for their convergence rather than uniqueness characteristics with real sampled data. Uniqueness is guaranteed for the two dimensional phase problem if the observable is an irreducible entire function. Although it has been shown[4] that the set of reducible polynomial functions of more than one variable is a set of measure zero, this does not mean that one can always assume a priori

that a unique phase exists in two dimensions. To ensure uniqueness we should restrict ourselves to classes of objects for which it is known that their Fourier transforms are irreducible polynomials of degree determined by the number of data points. We present a modification to the object distribution which ensures that its Fourier transform is an irreducible polynomial. The comparison is then made, using Fienup's algorithm, of object reconstruction with and without the modification. The particular modification selected from many which would ensure irreducibility has similarities to holography.

IRREDUCIBLE POLYNOMIALS

A polynomial of total degree N in two variables will require $(N+1)(N+2)/2$ coefficients and thus this number of data points to represent it uniquely. If we have N^2 data points, then we can assume that the maximum degree in each variable is $N-1$, and thus find N^2 of the coefficients which are associated with this polynomial. We wish to ensure irreducibility of this polynomial. The example given by Bruck and Sodin[5] in which a reference point is placed to one side of a one dimensional object is very restricted.

If the object support is not known, then the simplest step is to assume a constraint for irreducibility outside a simply-shaped region within which the object is known to be confined. If the object support is known, a more appropriate and specific constraint can be introduced and, in addition, it may be possible to model the object by a polynomial of higher degree and thus achieve

higher resolution. The following is a sufficient condition for irreducibility of $F(z_1,z_2)$, z complex : EISENSTEIN'S CRITERION[6]. Consider $F(z_1,z_2)$ as a polynomial in z_1 , i.e.,

$$F(z_1,z_2) = a_0(z_2) + a_1(z_2)\, z_1 \cdots a_{N-1}(z_2)\, z_1^{N-1}$$

Thus the coefficients are polynomials in z_2 . If there exists a prime (irreducible) factor $p(z_2)$ which divides $a_0, a_1, \cdots a_{N-2}$, but not a_{N-1} , and if $p^2(z_2)$ does not divide a_0, then $F(z_1,z_2)$ is irreducible. In C the only prime is of the form $z_2 + b$, where b is complex.

Consider the general form of a polynomial in two variables having maximum powers J and K in z_1 and z_2 :

$$F(z_1,z_2) = \sum_{0}^{J} \sum_{0}^{K} f(j,k)\, z_1^{j}\, z_2^{k}$$

The coefficients of the polynomial are samples of the object. We can construct an irreducible polynomial in, for example, the following way. Assume the region containing the object support is a rectangle defined by $0 \leq j \leq L-1$ and $1 \leq k \leq M$. A reference point at $(L,0)$ ensures irreducibility provided that the point at $(0,1)$ is non-zero. The simplest prime, z_2, divides all coefficients except that of the z_1^{L} term, and z_2^{2} does not divide the z_1^{0} coefficient.

The reference function introduced can be arbitrarily close to the object support, provided that the Eisenstein criterion is satisfied. The method has similarities to off-axis holography, a holographic reconstruction failing because of the overlap of the auto- and cross-correlation terms.

IMPLEMENTATION

It has yet to be proved that, when only one possible phase function exists, the Fienup method converges to the correct phase. However, we have compared object reconstructions using this algorithm with and without the reference using the same initial phase guess. The results are very encouraging, showing a rapid convergence with the reference when there is no sign of convergence to the correct solution without it.

CONCLUSIONS

Despite the likelihood of irreducibility for functions of more than one variable, the lack of consistent success of phase algorithms suggests that irreducibility should be guaranteed a priori. The Eisenstein criterion is one particular sufficient condition for irreducibility. Having imposed the irreducibility criterion, it was found that the Fienup algorithm converged quickly to the correct missing phase.

ACKNOWLEDGEMENTS

This work was supported by the Air Force Office of Scientific Research under grant AFOSR 81 0003.

REFERENCES

1. Ferwerda, H.A. in AIP Conf. Proc. 65 (1980), ed. by M.A. Machado & L.M. Narducci.
2. Fienup, J.R. Opt. Lett. 3 (1978) 27.
3. Fienup, J.R. Appl. Opt. 21 (1982) 2758.
4. Hayes,M.H. & McClellan,J.H. Proc. IEEE 70 (1982) 197.
5. Bruck,Yu.M. & Sodin,L.G. Opt. Commun. 30 (1979) 304 .
6. van der Waerden,B.L. "Algebra,Vol 1", F. Ungar Pub. Co. (1970), 2nd. ed.,p94.

SESSION VII

SYNTHESIS AND RECONSTRUCTION

M. H. Hayes, *Presider*

SIGNAL RECONSTRUCTION FROM PARTIAL FOURIER DOMAIN INFORMATION *

Alan V. Oppenheim and Jae S. Lim

Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
Research Laboratory of Electronics
Room 36-615
Cambridge, Mass. 02139

I. INTRODUCTION

There are a variety of practical problems in which only the phase or magnitude of the Fourier transform of a signal is known and it is desired to reconstruct the signal. In this talk, a number of results developed in the Digital Signal Processing Group at M.I.T. over the past several years will be described. The work discussed began initially with an exploration of the intelligibility of phase-only signals, that is ones for which the correct Fourier transform phase is combined with a constant or characteristic Fourier transform magnitude function. Motivated by the importance of Fourier transform phase in relation to Fourier transform magnitude, a theory and associated algorithms were then developed for the exact reconstruction of finite length signals from phase information alone.

It is generally recognized that there is an asymmetry in the results for signal reconstruction from phase alone and from Fourier transform magnitude alone. In particular, whereas for finite length signals, exact reconstruction to within a scale factor is possible from phase for one-dimensional or multi-dimensional signals, a corresponding result is only true for multi-dimensional signals for reconstruccion from Fourier transform magnitude. Furthermore, in the multi-dimensional case, there is considerably more difficulty with robustness and convergence of the algorithms for reconstruction from magnitude than for reconstruction from phase. Recently, we have developed a theory and associated algorithms for exact reconstruction of finite length signals from Fourier transform "amplitude" defined as the Fourier transform magnitude augmented by one bit of phase information. Both the theory and algorithms parallel very closely those for reconstruction from phase. Furthermore, a theory and associated algorithms have been developed for the exact reconstruction of signals from the magnitude of the short-time, or sliding, Fourier transform. The algorithms for this reconstruction appear to be relatively robust and a number of applications are currently being pursued.

* This work has been supported in part by the National Science Foundation under Grant ECS80-07102 and in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506.

II. REPRESENTATION OF A SIGNAL BY ITS FOURIER TRANSFORM PHASE

Apparently independently, and in a number of different contexts, including x-ray crystallography, image processing, and acoustical and optical holography, it has been recognized that many features of a signal are retained in a phase-only Fourier synthesis, but not in a magnitude-only Fourier synthesis [1]. There have been a variety of (not totally satisfactory) analytical attempts at explaining the reason for the intelligibility of phase-only signals. The explanations more or less center around the fact that edge and position information in signals is strongly reflected in the phase and to the extent that this information is important for intelligibility, as it tends to be for speech and images, intelligibility can be retained in phase-only reconstructions.

The reasonably high intelligiblity of phase-only signals demonstrates the fact that much of the important information resides in the phase and raises the question as to whether some or perhaps all of the magnitude information can be extracted or inferred from the phase. Although, in general, a signal is not uniquely defined by its Fourier transform phase, it may be under certain conditions or constraints. One well known set of conditions under which a signal may be uniquely recovered to within a scale factor from its phase is the minimum phase or maximum phase condition. Under these conditions, the log magnitude of the Fourier transform is the Hilbert transform of the phase. For many signals of interest, the minimum phase or maximum phase condition does not generally apply. There are, however, other sets of conditions unrelated to the minimum phase or maximum phase conditions under which signal recovery to within a scale factor is possible from the phase. In particular, for one-dimensional discrete time signals, it has been shown that if the signal is finite length and has a z-transform with no zeros in conjugate reciprocal pairs, then phase information alone is sufficient for signal reconstruction [2]. This result has also been extended in a number of ways to multi-dimensional signals [3,4].

A variety of algorithms have been developed for implementing exact signal reconstruction from phase. One algorithm consists of solving a set of simultaneous linear equations. A second, more robust algorithm, is an iterative procedure which alternately imposes the finite length constraint in the time domain and the known phase information in the frequency domain [5].

III. REPRESENTATION OF A SIGNAL BY ITS FOURIER TRANSFORM AMPLITUDE

It is well known that the above conditions for signal reconstruction from Fourier transform phase do not also apply to reconstruction from Fourier transform magnitude. While theoretically a two or higher dimensional finite length signal can be recovered from Fourier transform magnitude [6] the procedure does not appear to be robust and practical algorithms have not been developed. However, a theory paralleling that in Section II has recently been developed which demonstrates that for real-valued, causal and finite extent

signals in either one dimension or higher dimensions, a signal is exactly represented by and reconstructable from the Fourier transform magnitude together with the sign of the real part of the Fourier transform, corresponding to a one-bit representation of phase [7,8]. The Fourier transform magnitude, together with one bit of phase information, has been referred to as the Fourier amplitude, and although in its present form the theory requires knowledge of the amplitude at all frequencies, a practical algorithm has been developed utilizing the discrete Fourier transform. This algorithm is an iterative algorithm similar to that used for reconstruction from the Fourier transform phase whereby the time domain and frequency domain constraints are alternately imposed. While convergence of the algorithm has not been demonstrated theoretically, it has effectively converged to the correct answer for all cases on which it has been tried.

IV. REPRESENTATION OF A SIGNAL FROM THE SHORT TIME FOURIER TRANSFORM MAGNITUDE

In many application areas, signal processing is carried out on the basis of a short time Fourier analysis. In speech processing in particular, the short time Fourier transform is used as the basis for both speech analysis and speech synthesis in a wide variety of applications. Often it is the Fourier transform magnitude that is recorded or processed under the assumption that the associated loss of information associated with discarding the phase is acceptable. Recently a theory has been developed which demonstrates that under very mild conditions, the short time Fourier transform magnitude is sufficient for exact representation of the signal [9,10]. In essence the requirement is that the analysis window be known and that the short time Fourier transform magnitude be available at time increments which are less then one-half of the length of the analysis window. Based on these conditions, it has been shown that the original signal can be exactly recovered to within a multiplication by plus or minus unity. Furthermore, a variety of algorithms implementing this reconstruction have been developed and implemented.

The importance of this theory relates not only to reconstruction when the exact short time Fourier transform magnitude is known but also to applications in which it has been purposely or inadvertently modified. This arises, for example, in speed rate changes of speech for which the time scale of the short time Fourier transform is purposely altered. In such cases, the resulting function of time and frequency is no longer a valid short time Fourier transform. Nevertheless, reconstruction using the algorithms based on short time Fourier magnitude alone provide a phase consistency in the reconstructed signal, which is highly desirable. A similar issue and corresponding application lies in the use of these results in the context of signal enhancement based on the short time Fourier transform.

V. REFERENCES

1. A. V. Oppenheim and J. S. Lim, "The Importance of Phase in Signals," Proc. IEEE, vol. 69, pp. 529-541, May 1981.

2. M. H. Hayes, J. S. Lim and A. V. Oppenheim, "Signal Reconstruction from Phase or Magnitude," IEEE Trans. on Acoust., Speech and Sig. Proc., vol. ASSP-28, pp. 672-680, Dec. 1980.

3. M. H. Hayes, "Signal Reconstruction from Phase or Magnitude," Sc.D. Thesis, M.I.T., Dept. of EECS, June 1981.

4. M. H. Hayes, "The Reconstruction of a Multidimensional Sequence from the Phase or Magnitude of its Fourier Transform," IEEE Trans. on Acoust., Speech and Sig. Proc., vol. ASSP-30, pp. 140-154, April 1982.

5. A. V. Oppenheim, M. H. Hayes and J. S. Lim, "Iterative Procedures for Signal Reconstruction from Fourier Transform Phase," Optical Eng., vol. 21, pp. 122-127, Jan./Feb. 1982.

6. Y. M. Bruck and L. G. Sodin, "On the Ambiguity of the Image Reconstruction Problem," Optics Communications, vol. 30, pp. 304-308, 1979.

7. P. L. Van Hove, "Signal Reconstruction from Fourier Transform Amplitude," S.M. Thesis, M.I.T., Dept. of EECS, Sept. 1982.

8. P. L. Van Hove, J. S. Lim and A. V. Oppenheim, "Signal Reconstruction from Fourier Transform Amplitude," IEEE Trans. on Acoust., Speech and Sig. Proc., submitted August 1982.

9. S. H. Nawab, "Signal Estimation from Short-time Spectral Magnitude," Ph.D. Thesis, M.I.T., Dept. of EECS, May 1982.

10. S. H. Nawab, T. F. Quatieri and J. S. Lim, "Signal Reconstruction from Short-Time Fourier Transform Magnitude," IEEE Trans. on Acoust., Speech and Sig. Proc., submitted July 1982.

A Phase-retrieval Technique in Pupil Synthesis

Piotr Kledron

The Ohmart Corporation, 4241 Allendorf Drive

Cincinnati, Ohio 45209

The pupil function of an incoherent system is uniquely determined uniquely by the incoherent transfer function [1]. This fact creates a phase-retrieval problem when the spectral function has to be determined from an intensity measurement. The phase and amplitude of the pupil function can be found from the point spread function only if additional information about the optical system is available. The information has a mathematical character if some constraints upon the class of admissible solutions are imposed. (See [2] and [3] where the class of phase functions is reduced to finite polynomials.) Another method assuring the uniqueness of the solution requires additional information of physical characteristics. (See [4], [5], [6], and [7] where the second measurement is postulated.)

The mathematical structure of the phase problem can be applied to another aspect of the pupil function determination. The nonunique relation between the pupil function and the incoherent transfer function is an advantage in pupil synthesis for incoherent processing. One needs to construct the pupil function H(u,v). To accomplish this the following integral equation

$$P(u,v) \boxtimes P^*(-u,-v) = H(u,v) \quad , \tag{1}$$

where "*" denotes complex conjugation and "$\boxtimes$" denotes the convolution operator, has to be solved.

Eq. (1) has a unique solution if and only if the Fourier transform of H(u,v,) has real zeros only. Usually eq. (1) has

an infinite number of solutions. This fact gives some freedom in selecting that solution which has an additional property. In a paradoxical example one can construct an optical system which is diffraction-limited (i.e. has no aberrations) if illuminated with incoherent light and on the other hand has aberrations if illuminated with coherent light.

The pupil function of a diffraction-limited system is constant within, and zero outside, the exit pupil. If the right-hand side of eq. (1) is a transfer function of the diffraction-limited system with either ractangular or circular aperture, then eq. (1) has only one solution. This can be derived from results in [9] and [10]. The recent work [11] should be helpful in generalization of this result for an arbitrary convex pupil.

The following example shows that the case of the exit pupil consisting of disjoint sets is different. Then eq. (1) may have many solutions.

Let us define the diffraction-limited system with the pupil function

$$P(u) = \mathrm{rect}(u + 1.5) + \mathrm{rect}(u-0.5) \qquad (2)$$

One may check that the Fourier transform

$$p(x) = \mathrm{F.T.}\,[\,P(u)\,] \qquad (3)$$

has essentially complex zeros. According to the theorem from [8], the following function is the solution of eq. (1) with the right-hand side equal to the autocorrelation of (2):

$$P_o(u) = \mathrm{F.T.}^{-1}\,[\,p(x)(x-z^*)/(x-z)\,]\,, \qquad (4)$$

where z is one of the zeros of p(x) and is of the form

$$z = 0.409064 - 0.0608354j \qquad (5)$$

One may check that $P_o(u)$ is not real, i.e., the system with a pupil function equal to $P_o(u)$ has aberrations and works like the diffraction-limited system under incoherent illumination because $P_o(u)$ generates the same incoherent transfer function as

Eval. Additionally, in terms of a coherent imaging, $P_0(u)$ is superior because it does not vanish for any spatial frequency $-1 \leq u \leq 1$ while the pupil function $P(u)$ does not transmit information encoded in frequencies $-0.5 \leq u \leq 0.5$.

The above example shows that in some cases a non-negative apodizer can completely compensate the effect of aberrations. Let us consider the optical system with a pupil function given by

$$P'(u) = |P'(u)| \exp[i \cdot \arg(P_0(u))] \quad ,$$

Then the positive apodizer

$$A(u) = |P_0(u)| / |P'(u)| \quad ,$$

reduces the system $P'(u)$ to the system $P_0(u)$ which in fact is diffraction-limited.

The question whether any system with aberrations can be reduced by means of a non-negative apodizer to some diffraction-limited system arises. We anticipate that the above description can be applied to a limited class of aberration functions; however, for this class which needs to be specified, the correction of aberrations seems to be promising.

1. H.A.Ferwerda, Inverse Source Problems in Optics, Vol.9, ed.H.P.Baltes (Springer-Verlag, Berlin 1978)
2. W.H.Southwel, J.Opt.Soc.Am. 67,396 (1977)
3. J.T.Foley, R.R.Butts, J.Opt.Soc.Am. 71,1008 (1981)
4. R.W.Gerchberg, W.O.Saxton, Optik 35,237 (1972)
5. A.J.J.Drenth, A.M.J.Huiser, H.A.Ferwerda, Opt.Acta 22, 615 (1975)
6. D.L.Misell, R.E.Burge, A.H.Greenaway, J.Phys.D:Appl. Phys. 7,127 (1974)
7. P.Kiedron, Opt.Acta ,254 (1980)
8. A.Walther, Opt.Acta 10,41 (1963)
9. P.Kiedron, Optik 53,303 (1981)
10. W.Lawton, J.Opt.Soc.Am. 71,1519 (1981)
11. J.R.Fienup, T.R.Crimmins, W.Holsztynski, J.Opt.Soc.Am. 72,610 (1982)

Dependent and Independent Constraints for a Multiple Objective Iterative Algorithm

Joseph N. Mait and William T. Rhodes

Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

Constrained iterative algorithms have been applied primarily to single objective applications,[1] where by objective we mean that distribution that is to be reconstructed from partial information or synthesized with desired characteristics. In a recent work by the authors,[2] Fienup's error-reduction algorithm was extended to multiple objectives, as shown in Fig. 1, and was applied to a specific synthesis problem with two objectives. In this paper, we illustrate an important characteristic of multiple objective iterative algorithms, dependent constraints.

Our problem is the synthesis of two pupil functions $P_1(u)$ and $P_2(u)$, for use in the incoherent optical spatial filtering system in Fig. 2, such that a desired bipolar spatial impulse response or point spread function (PSF) results. The effective pupil function $P(u;A_1,A_2,\Phi)$ of the optical system in Fig. 2 is given by[3]

$$P(u;A_1,A_2,\Phi) = A_1P_1(u)e^{j\Phi} + A_2P_2(u), \tag{1}$$

and the corresponding PSF $f(x;A_1,A_2,\Phi)$ by

$$\begin{aligned} f(x;A_1,A_2,\Phi) &= |p(x;A_1,A_2,\Phi)|^2 \\ &= A_1^2|p_1(x)|^2 + A_2^2|p_2(x)|^2 \\ &\quad + A_1A_2[p_1(x)p_2^*(x)e^{j\Phi} + p_1^*(x)p_2(x)e^{-j\Phi}] \end{aligned} \tag{2}$$

where the pupil function $P(u)$ and the coherent spread function (CSF) $p(x)$ form a Fourier transform pair. A desired bipolar PSF $f(x)$ may be synthesized through control of transmittance factors A_1 and A_2 and phase Φ.

Lohmann and Rhodes identify two distinct regimes for implementing bipolar PSFs in this way, pupil interaction and pupil noninteraction.[3] The synthesized PSF $f_s(x)$ resulting from pupil noninteraction is given by

$$f_s(x) = A_1^2|p_1(x)|^2 - A_2^2|p_2(x)|^2, \tag{3}$$

and for pupil interaction by (where $\phi_1(x) = \arg\{p_1(x)\}$, $1=1,2$)

$$\begin{aligned} f_s(x) = 2A_1A_2|p_1(x)p_2(x)|\{&\cos[\Phi_a + \Phi_1(x) - \Phi_2(x)] \\ &- \cos[\Phi_b + \Phi_1(x) - \Phi_2(x)]\} \end{aligned} \tag{4}$$

where Φ_a and Φ_b are two different, but fixed, values of phase Φ in Fig. 2.

Since it is the pupil functions $P_1(u)$ and $P_2(u)$ that describe the system, they are our objective functions. From a practical optical standpoint, the pupil functions must be of finite extent; thus we desire the following of our system:

I. synthesis of a bipolar PSF

$f_s(x) = f(x)$, where $f(x)$ is the desired bipolar PSF.

II. finite extent pupil functions

$P_{1,2}(u) = P_{1,2}(u)$ rect $(\frac{u}{w})$, where w is the extent of the pupil.

With respect to Eqs. (3) and (4), Condition I is a dependent, or mutual constraint, since both $p_1(x)$ and $p_2(x)$ must jointly satisfy the constraint. This is in contrast to Condition II, where the constraint on $P_1(u)$ in no way determines or affects the constraint on $P_2(u)$. Understandably, greater freedom exists, and more ingenuity may be required, to satisfy a dependent constraint as opposed to an independent constraint because there are no explicit constraints on the objectives.

By definition of the error-reduction algorithm, a point not satisfying the domain constraints is replaced by a point that satisfies the constraint and is a minimum distance from the original point.[1] Figure 3 is a vector diagram depicting the minimum changes necessary to assure Condition I. It is assumed that the desired PSF $f(x)$ is dependent equally upon $p_1(x)$ and $p_2(x)$; thus modifications to one are equal and opposite to modifications of the other as long as Condition I is maintained. Condition I is therefore satisfied for the pupil noninteraction regime, assuming $A_1 = A_2 = 1$, by

$$|p_1'(x)| = \sqrt{\text{POS}\left[\frac{|p_1(x)|^2 + |p_2(x)|^2 + f(x)}{2}\right]}, \tag{5a}$$

$$|p_2'(x)| = \sqrt{\text{POS}\left[\frac{|p_1(x)|^2 + |p_2(x)|^2 - f(x)}{2}\right]}, \tag{5b}$$

where POS[g(x)] is a half-wave rectification of g(x). The phase is undisturbed. For the interaction regime, assuming the modified point is $\alpha e^{j\theta}$, $2A_1A_2 = 1/2$, $\Phi_a = 0$, and $\Phi_b = \pi$, Condition I is satisfied by

$$|p_1'(x)| = \sqrt{\alpha \frac{|p_1(x)|}{|p_2(x)|}}, \qquad \Phi_1'(x) = \frac{\Phi_1(x) + \Phi_2(x) + \theta}{2}, \tag{6a}$$

$$|p_2'(x)| = \sqrt{\alpha \frac{|p_2(x)|}{|p_1(x)|}}, \qquad \Phi_2'(x) = \frac{\Phi_1(x) + \Phi_2(x) - \theta}{2}. \qquad (6b)$$

Algorithms implemented using Eqs. (5) and (6) were tested experimentally. With the bandpass filter in Fig. 4 as the desired bipolar PSF, Figs. 5 and 6 represent pupil noninteractive and pupil interactive synthesis of the PSF after 100 iterations using Eqs. (5) and (6), respectively. The normalized squared error is 0.0875 for the noninteractive regime and 0.5137 for the interactive regime.

The high error for the interactive regime may result from the great amount of freedom the algorithm presents; although the error is reduced with each iteration, the reduction is slight. For this reason, the algorithm was modified to force the synthesized point equal to the desired, the addition of Δ_2 to f_s in Fig. 3b, as opposed to altering its projection onto the real axis, the addition of Δ_1. In the limit of a large number of iterations, Δ_1 and Δ_2 should be equal. The results of this algorithm are presented in Fig. 7. The normalized error is 0.0276.

References

1. J.R. Fienup, "Reconstruction and synthesis applications of an iterative algorithm," in Transformations in Optical Signal Processing, W.T. Rhodes, J.R. Fienup, B.E.A. Saleh, eds. (SPIE, Bellingham, 1983).
2. J.N. Mait and W.T. Rhodes, "Iterative design of pupil functions for bipolar incoherent spatial filtering, "Processing of Images and Data from Optical Sensors, W.H. Carter, ed. (Proc. SPIE, vol. 292, 1981) pp. 66-72.
3. A.W. Lohmann and W.T. Rhodes, "Two-pupil synthesis of optical transfer functions," Appl. Opt. 17 (1978) 1141-1151.

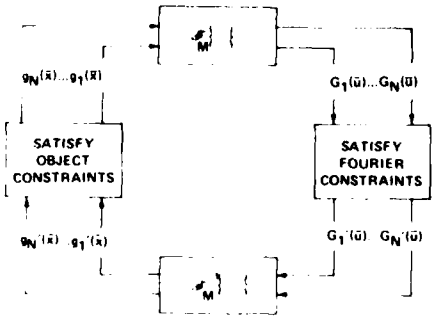


Figure 1. Generalization of Fienup's error-reduction algorithm to multi-dimensions and multiple objectives. (From Ref. 2)

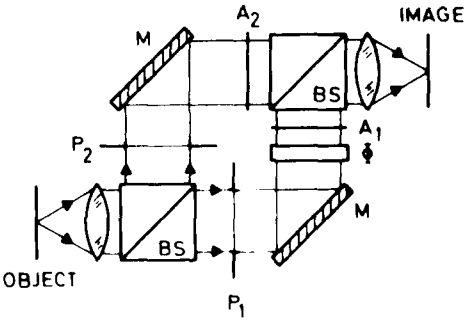
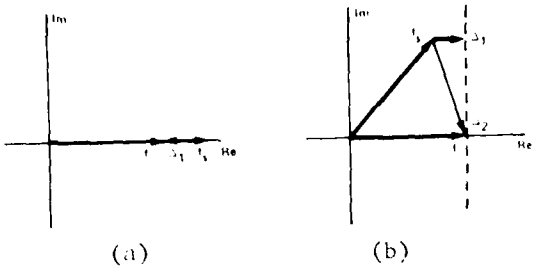


Figure 2. Incoherent optical system for synthesizing bipolar PSFs. P_1 and P_2 denote pupil transparencies, A_1 and A_2, attenuators; Φ a phase shift, BS, a beam splitter, and M, a mirror. (From Ref. 1)



(a) (b)

Figure 3.
The vector Δ_1 represents the minimum change necessary to f_k to satisfy Condition 1.
a. pupil noninteractive regime
b. pupil interactive regime; Δ_2 is a much harsher constraint

Figure 4.
a. Desired bipolar PSF $f(x,y)$
b. Associated optical transfer function (OTF) $F(u,v)$

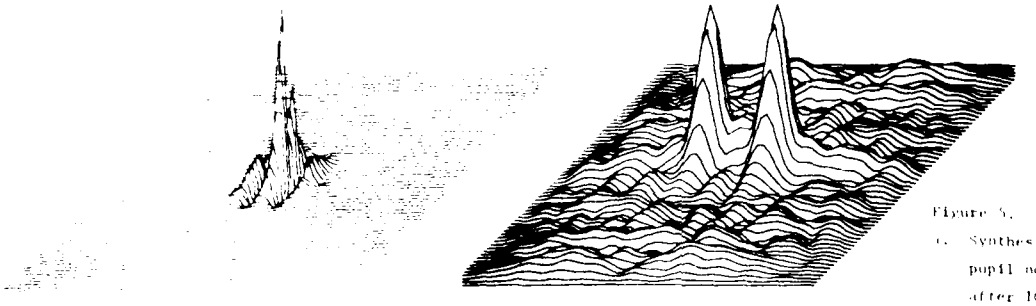


Figure 5.
a. Synthesized PSF $f_k(x,y)$ in the pupil noninteractive regime after 100 iterations
b. Associated OTF $F_k(u,v)$

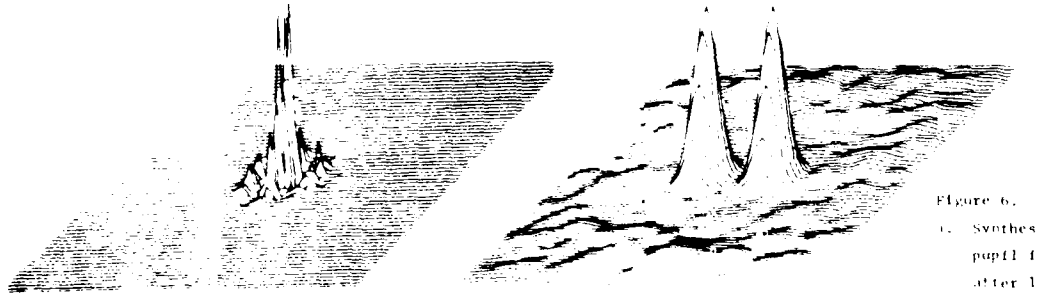
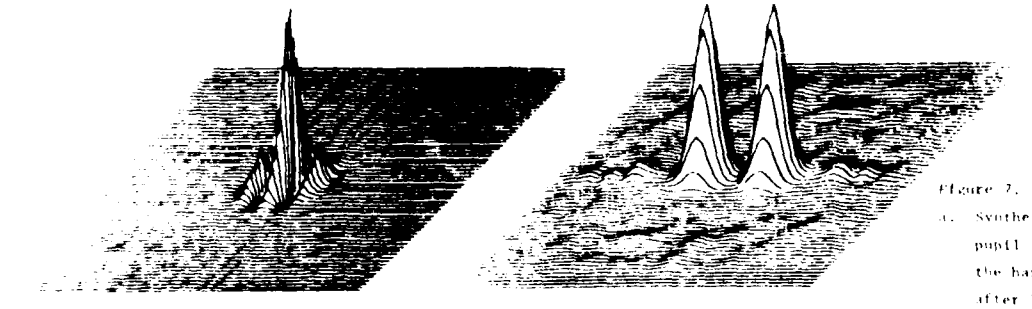


Figure 6.
a. Synthesized PSF $f_k(x,y)$ in the pupil interactive regime after 100 iterations
b. Associated OTF $F_k(u,v)$



(a) (b)

Figure 7.
a. Synthesized PSF $f_k(x,y)$ in the pupil interactive regime with the harsher equality constraint after 100 iterations
b. Associated OTF $F_k(u,v)$

Signal Reconstruction From Fourier Transform Amplitude*

Patrick L. Van Hove, Jae S. Lim, and Alan V. Oppenheim
Massachusetts Institute of Technology
Research Laboratory of Electronics
Department of Electrical Engineering and Computer Science
Cambridge, Massachusetts 02139

I. Introduction

In a variety of contexts, it is desirable to reconstruct a sequence from partial Fourier domain information. As a consequence, considerable attention has been paid to, and some significant results have been developed in, this area. For example, it has been established [1,2] that under very mild restrictions a finite extent one-dimensional (1-D) or multi-dimensional (M-D) sequence is uniquely specified to within a scale factor by its Fourier transform (FT) phase, and algorithms for implementing the reconstruction have been developed. It is well known that in contrast, the FT magnitude does not uniquely specify a 1-D sequence. Even for M-D sequences, the FT magnitude specifies a sequence only to within a translation and a central symmetry [2], and reconstruction algorithms developed so far have been successful [2] for only a very restricted class of M-D sequences.

In this paper we summarize new results on the reconstruction of 1-D and M-D sequences when the FT magnitude and one bit of phase information is known. In particular, these results show that under very mild restrictions, this is sufficient to uniquely specify the sequence.

II. Theory

Before we present a summary of the theoretical results, we define the notation that will be used. Let $x(n)$ denote a 1-D sequence which is causal and finite extent so that $x(n)$ is zero outside $0 \leq n \leq L-1$. Furthermore, we restrict $x(n)$ to be real-valued. Let $X(z)$ and $X(\omega)$ represent the z-transform and Fourier transform of $x(n)$, so that

*This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

$$X(z) = \sum_{n=0}^{L-1} x(n)z^{-n} \tag{1}$$

$$X(\omega) = X(z)\Big|_{z=e^{j\omega}} = \sum_{n=0}^{L-1} x(n)e^{-j\omega n} \tag{2}$$

The FT $X(\omega)$ can be represented in terms of its magnitude $|X(\omega)|$ and phase $\theta_x(\omega)$ as follows:

$$X(\omega) = |X(\omega)|e^{j\,\theta_x(\omega)} \tag{3}$$

To ensure that $\theta_x(\omega)$ is well defined at all ω, we assume that $X(z)$ has no zeros on the unit circle. The phase function $\theta_x(\omega)$ in Equation (3) represents the principal value of the phase so that

$$-\pi < \theta_x(\omega) \leq \pi \tag{4}$$

The one-bit FT phase information will be represented by the function $S_x^{\alpha}(\omega)$ defined as

$$S_x^{\alpha}(\omega) = \begin{cases} +1 & \text{when } \alpha-\pi \leq \theta_x(\omega) \leq \alpha \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

where α is a known constant in the range of $0<\alpha\leq\pi$. Thus, the complex plane is divided into two regions separated by a straight line passing through the origin and at an angle α with the real axis. For example, for $\alpha=\frac{\pi}{2}$, $S_x^{\pi/2}(\omega)$ represents the algebraic sign of the real part of $X(\omega)$. The function $G_x^{\alpha}(\omega)$ defined as

$$G_x^{\alpha}(\omega) = S_x^{\alpha}(\omega)\ |X(\omega)| \tag{6}$$

will be reffered to as the FT <u>amplitude</u> since it contains both magnitude and sign information.

In our recent research, we have shown [3] the following statement:

<u>Statement 1</u>

Let $x(n)$ and $y(n)$ be two real, causal (or anti-causal), and finite extent sequences, with z-transforms which have no zeros on the unit circle. If $G_x^{\alpha}(\omega)=G_y^{\alpha}(\omega)$ for all ω and $0<\alpha<\pi$, then $x(n)=y(n)$. When $\alpha=\pi$, if $G_x^{\pi}(\omega)=G_y^{\pi}(\omega)$ and

$x(0)=y(0)=0$, then $x(n)=y(n)$.

Statement 1 shows that under mild restrictions, a 1-D sequence is uniquely specified by its FT amplitude.

Statement 1 has been extended to M-D sequences. Let $x(\underline{n})$ denote an M-D sequence $x(n_1,n_2,\ldots,n_M)$, and let $G_x^{\alpha}(\underline{\omega})$ denote the FT amplitude of $x(\underline{n})$, where $G_x^{\alpha}(\underline{\omega})$ represents $G_x^{\alpha}(\omega_1,\omega_2,\ldots,\omega_M)$ and is given by $S_x^{\alpha}(\underline{\omega})|X(\underline{\omega})|$. We define an M-D sequence $x(\underline{n})$ to have a one-sided region of support in the M-dimensional space $n_1,n_2,\ldots,n_M$ if it only has non-zero values for one polarity on each index n_i. For example, for a two-dimensional sequence there are four possible regions of support which are consistent with the sequence being one-sided, corresponding to the four quadrants. Statement 2, which follows, represents a generalization of Statement 1 to encompass M-D sequences.

Statement 2

Let $x(\underline{n})$ and $y(\underline{n})$ be two real, finite extent sequences with one-sided support and with z-transforms which have no zeros at $|\underline{z}|=1$. If $G_x(\underline{\omega})=G_y(\underline{\omega})$ for all $\underline{\omega}$ and $0<\alpha<\pi$, then $x(\underline{n})=y(\underline{n})$. When $\alpha=\pi$, if $G_x^{\pi}(\underline{\omega})=G_y^{\pi}(\underline{\omega})$ and $x(\underline{0})=y(\underline{0})=0$, then $x(\underline{n})=y(\underline{n})$.

The theoretical result in Statement 2 differs from that of Hayes [2] in several respects. In Hayes' result, only samples of the FT magnitude are required, but the sequence is restricted to have a non-factorizable z-transform and the unique specification of the sequence is only to within a sign, a translation, and a central symmetry. In Statement 2, the FT amplitude is required, but the sequence may have a factorizable z-transform and is uniquely specified in the strict sense.

III. Algorithm

The results in Section II show that a 1-D or M-D sequence which satisfies certain conditions is uniquely specified by its FT amplitude. To reconstruct a sequence that satisfies the conditions of Statement 2 from its FT amplitude,

we nave developed an iterative procedure which is similar in style to other iterative procedures studied by Gerchberg-Saxton [4] and Fienup [5]. In the iterative algorithm, the "time" domain constraint that $x(\underline{n})$ be real and finite extent with a one-sided region of support, and the frequency domain constraint that tne FT amplitude of $x(\underline{n})$ be given by $G_x^{\alpha}(\underline{\omega})$, are imposed separately in each iteration.

To implement the algorithm, the Fourier and inverse Fourier transform operations are approximated by discrete Fourier transform (DFT) and inverse DFT (IDFT) operations. Although the uniqueness is not guaranteed in terms of the FT amplitude samples, we have empirically observed that the algorithm reconstructs the desired sequence provided that the FT amplitude is densely sampled in the frequency domain, so that the FT magnitude is completely specified and the discontinuities of $S_x^{\alpha}(\underline{\omega})$ are individually resolved by the samples of $S_x^{\alpha}(\underline{\omega})$. This observation is based on both 1-D and 2-D examples derived from speech signals and images.

When the results shown in this paper are combined with previous work [2] on the problem of reconstructing a 1-D or M-D sequence from its FT phase, we obtain a simple general result. Specifically, a 1-D or M-D sequence is uniquely specified under mild restrictions by its FT phase or its FT amplitude. In addition, an iterative algorithm which is similar in style can be used to reconstruct a 1-D or M-D sequence from its FT phase or amplitude.

References

[1] M. H. Hayes, J. S. Lim, A. V. Oppenheim, "Signal Reconstruction From Phase or Magnitude," IEEE Trans. ASSP ASSP-28, 672-680, Dec. 1980.

[2] M. H. Hayes, "Signal Reconstruction From Phase or Magnitude," Sc.D. thesis, MIT, June 1981.

[3] P. L. Van Hove, "Signal Reconstruction From Fourier Transform Amplitude," S.M. thesis, MIT, Sept. 1982.

[4] R. W. Gerchberg, W. O. Saxton, "A Practical Algorithm for the Determination of Phase From Image and Diffraction Plane Pictures," Optik 35, 237-246, 1972.

[5] J. R. Fienup, "Reconstruction of an Object From the Modulus of its Fourier Transform," Optics Letters, 27-29, July 1978.

SESSION VIII

PROBLEMS AND METHODS

N. C. Gallagher, *Presider*

Reconstruction in Electron Microscopy

W. O. Saxton

Cavendish Laboratory, England

ANALYSIS OF TIME-SEQUENTIAL SAMPLING WITH A SPATIALLY HEXAGONAL LATTICE

Robert M. Cramblitt
Jan P. Allebach
Department of Electrical Engineering
University of Delaware
Newark, Delaware 19711

Hexagonal sampling of 2-D signals has been considered a useful alternative to rectangular systems. Mersereau [1] analyzed hexagonal sampling and showed that when images are circularly or elliptically bandlimited the sampling density can be 13.6% less than that required for a rectangular sampling system. Murphy and Gallagher [2] studied hexagonal sampling in the context of optical systems. Here we investigate spatially hexagonal sampling of spatiotemporal signals. Since obtaining samples at every point in space at the same time instant is impractical for many applications, we constrain the sampling to be time-sequential. We examine the performance of various sampling patterns as we sample below the temporal Nyquist rate, and we compare their performance with that of the corresponding rectangular case that was analyzed previously [3].

Figure 1 shows a typical hexagonal grid and some relevant parameters. In time-sequential sampling, we take a sample at point $(a(\ell)\cdot X, b(\ell)\cdot Y)$ at time $\ell\cdot T$ where T is the temporal sampling interval. The ordered pairs $\{a(\ell), b(\ell)\}$ define a sampling pattern that is repeated from frame to frame. The sum $a(\ell)+b(\ell)$ must be even for hexagonal sampling.

In order to take a non-deterministic approach, we assume the signal g is a spatially and temporally wide sense stationary random process with zero mean. The power spectral density of the hexagonally sampled signal h is given by:

$$S_{hh}(u,v,f) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{MN/2-1} |Q_{mnp}|^2 S_{gg}\left(u - \frac{m}{A}, v - \frac{n}{C}, f - \frac{p}{B}\right)$$

where Q_{mnp} is the 3-D DFT of the sampling pattern:

$$Q_{mnp} = \frac{2}{MN} \sum_{\ell=0}^{MN/2-1} \exp\left\{ - j2\pi \left(\frac{a(\ell)m}{M} + \frac{b(\ell)n}{N} + \frac{2p\ell}{MN} \right) \right\}$$

Thus, the power spectral density consists of replications of the baseband spectrum weighted by the DFT of the sampling pattern. Assuming the signal to be spatially bandlimited to frequency U and temporally limited to W, we can show the noise power due to aliasing to simply be

$$\sigma_e^2 = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{MN/2-1} |Q_{mnp}|^2 V_{mnp}$$

where V_{mnp} is the integral over that part of the replication at (m,n,p) that overlaps the baseband spectrum.

Nyquist sampling, implying zero noise power, will be attained for any pattern when:

$$U \leq = \min \begin{cases} 1/2X = M/2A \\ \\ 1/2Y = N/2C\sqrt{3} \end{cases}$$

$$W \leq = 1/2B = 1/(MNT)$$

The overall sampling rate is then

$$F_s = 2\sqrt{3}\ CUW\ \sqrt{4(AU)^2 - 1}\ , \quad A < C$$

If we are going to reduce F_s below the Nyquist rate, it is preferable [3] to increase the time-bandwidth product BW rather than reduce M or N.

The signal to noise power ratios (SNR's) as BW increases from the Nyquist value of .5 are shown in Fig. 2 for several sampling patterns. The pseudorandom case (PSR) is an average over the entire ensemble of possible sampling patterns and thus represents the performance we might expect if we picked a pattern at random. The lexicographic (LEX) pattern is simply discrete line-by-line scanning, and is seen to perform slightly worse than average. The bit-reversed pattern (BRV) was generated by repeating a rectangular bit-reversed pattern on a diagonally shifted grid to obtain a hexagonal pattern. The rapid degradation in the SNR is an unexpected result since this pattern is uniformly better than average in rectangular sampling.

In order to make comparisons with the rectangular patterns, the SNR's are plotted in Fig. 3 versus the sampling rate normalized by ACU^2W with units of samples per cycle^3. This compensates for differences in M,N and image size between the rectangular and hexagonal lattices. In both the PSR and LEX cases, the hexagonal patterns closely retain their 13.6% advantage in sampling rate at a given SNR. The hexagonal BRV pattern loses its 13.6% advantage as BW increases.

The patterns considered so far were chosen primarily because they are easy to generate; but they are not optimal in the sense of maximizing the SNR at a given BW. Two methods were used for finding optimal sampling patterns. The first involves mapping the indices a,b, and ℓ into sub-indices using the Chinese Remainder Theorem. A search is then made for the best ordering of the sub-indices. The second method produces a closed form expression for the sampling pattern. Here we note that the energy in any p-plane is

$$\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |Q_{mnp}|^2 = \begin{cases} 1, \text{ rectangular sampling,} \\ \\ 2, \text{ hexagonal sampling.} \end{cases}$$

We attempt to concentrate this energy at indices (m,n) where V_{mnp} is small. For the rectangular case, it can be shown that the resulting pattern is

optimal when $BW \leq = 1$. The search and closed form methods generate identical results, strongly suggesting that both methods produce an optimal pattern for both the hexagonal and rectangular cases.

Figure 4 shows the SNR's for both the rectangular and hexagonal optimal patterns. Comparison of these patterns surprisingly reveals that the hexagonal optimal pattern does not retain a 13.6% savings in sampling rate. This is in contrast to the PSR case, also shown in Fig. 4, in which the 13.6% savings is approximately maintained. The SNR's in the figure also demonstrate that an optimal pattern has the lowest temporal Nyquist rate.

We conclude that certain hexagonal time-sequential patterns are decidely better than others under conditions of temporal aliasing, and that patterns that optimize the SNR at a given time-bandwidth product can easily be found. Some patterns show a 13.6% sampling rate advantage over equivalent rectangular patterns, but this advantage will not be present when optimal patterns are used.

ACKNOWLEDGMENT

This research is supported by the National Science Foundation under grant ECS-3102759. One of the authors (J.P.A.) wishes to acknowledge a seminal discussion with H. Stark on the topic of this paper.

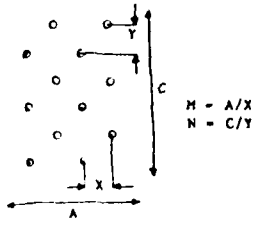
REFERENCES

[1] R. M. Mersereau, "The Processing of Hexagonally Sampled Two-Dimensional Signals," Proc. IEEE 67, 930-949, June 1979.

[2] N. Gallagher and R. Murphy, "Hexagonal Sampling Techniques Applied to Fourier and Fresnel Digital Holograms," J. Opt. Soc. Am. 72, 929-937, July 1982.

[3] J. P. Allebach, "Analysis of Sampling-Pattern Dependence in Time-Sequential Sampling of Spatio-Temporal Signals," J. Opt. Soc. Am. 71, 99-105, January 1981.

[4] J. P. Allebach, "Design of Anti-Aliasing Patterns for the Time-Sequential Sampling of Spatio-Temporal Signals," Proceedings of the 15th Annual Conference on Information Science and Systems, Baltimore, Md., 25-27 March 1981, 446-450.



T = Time between samples
B = Frame Period = MNT/2
AxC = Image Size

Fig. 1. A 4x6 hexagonal grid. Note that there are MN/2 points in one frame. The frame period is the time between successive samples at the same point.

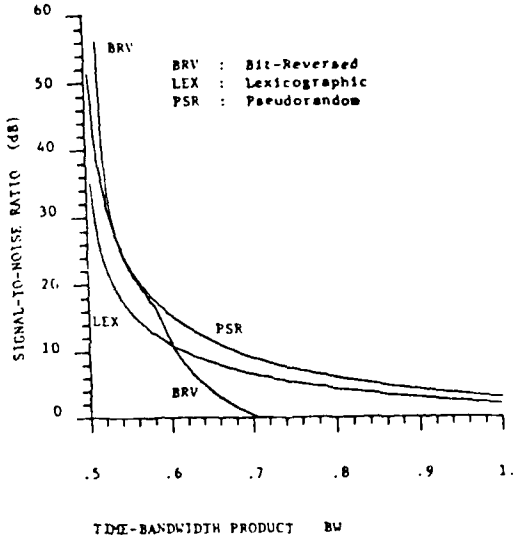


Fig. 2. Signal-to-noise ratios for three hexagonal sampling patterns. The curves are nearly independent of M and AU when M/AU is constant.

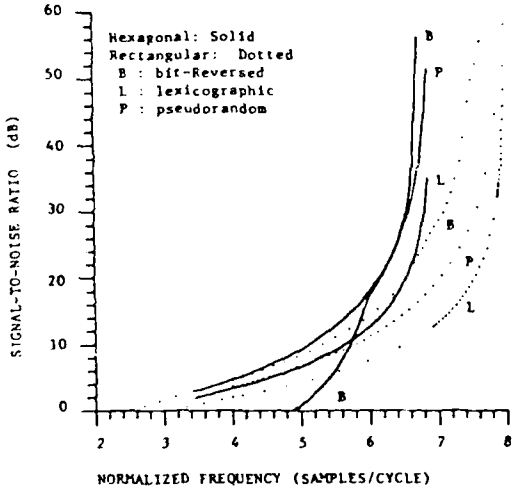


Fig. 3. SNR for hexagonal and rectangular sampling patterns.

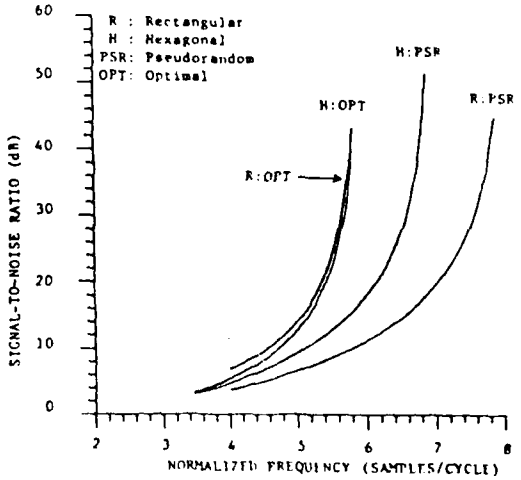


Fig. 4 SNR for optimal patterns. The hexagonal pattern size is 40x42. The rectangular pattern size is 44x45.

IMAGE REGISTRATION; THE UNDERSAMPLED CASE

P. E. Barry, M. Klop, J. D. Hulsmann
Research and Development Center, Grumman Aerospace Corporation
Mail Stop A08-35, Bethpage, New York 11714

The registration of a multi-frame sequence of images has been a classic problem in image processing whose solution has found wide application in such diverse areas as astronomy, earth resource analysis, electron microscopy, computed- aided medical imaging, military surveillance and photo-reconnaissance. In the most general sense, precise image registration allows such subsequent processing as signal-to-noise improvement, pattern enhancement and characterization, and the automated detection of spatio-temporally varying patterns, to be carried out at a level that would be impossible if the image sequence were not accurately registered.

To date, the vast majority of image registration techniques have depended upon cross correlation measures to determine the direction and amount of relative scene displacement. The applicability of this approach rests squarely on the assumption that the scene sampling is dense compared to the frequency content of the image. This paper addresses the registration problem when the image is undersampled and the relative displacements are small fractions of the basic pixel size. Under this assumption, the output of the i^{th} detector, is given by

$$O_i(t) = \int_0^L \int_0^L I(\eta - x(t), \gamma - y(t))d\eta d\gamma \qquad (1)$$

where $x(t)$ and $y(t)$ represent the instantaneous detector position and $I(x,y)$, the spatial scene intensity. A one- dimensional depiction of the displacement effect is shown in Fig. 1. For small displacements $x(t)$, the integrated intensity change is linear with respect to $x(t)$ and can be expressed as

$$O_i(t_1) = O_i(0) + (I(\alpha + L) - I(\alpha))\ x(t_1) \qquad (2)$$

The basic problem with undersampled data is, that as seen from the dotted line in Fig. 1, the integrated intensity on the i^{th} detector can change without changing the coefficient through which the displacement $x(t)$ effects $O_i(t_1)$. It is for this reason that cross-correlation techniques cannot yield relative frame displacements in undersampled data sets. For two-dimensional arrays, equation (2) generalizes to

$$O_i(t) = O_i(0) + f_i\ x(t) + g_i y(t) + n_i(t) \qquad (3)$$

with f_i and g_i representing the local scene gradients "seen" by the i^{th} detector, and $n_i(t)$, the measurement noise at the i^{th} detector.

The basic problem then is, given $\{O_i(t_j);\ i = 1, \ldots, N,\ j = 1, \ldots n\}$ (N = number of detectors in array), no knowledge of the actual scene $I(x,y)$, no knowledge of the random scene displacements $x(t)$ and $y(t)$, register the image sequence.

The image data is arranged in n-vector format as follows:

$$\underline{z}_i = \begin{bmatrix} z_i(t_1) \\ z_i(t_2) \\ \vdots \\ z_i(t_n) \end{bmatrix} = f_i \begin{bmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_n) \end{bmatrix} + g_i \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_n) \end{bmatrix} + \begin{bmatrix} n(t_1) \\ n(t_2) \\ \vdots \\ n(t_n) \end{bmatrix} \qquad (4)$$

$$i = 1, 2, \ldots, N$$

or more concisely, as

$$\underline{z}_i = \begin{bmatrix} | & | \\ \underline{x} & \underline{y} \\ | & | \end{bmatrix} \begin{bmatrix} f_i \\ g_i \end{bmatrix} + \begin{bmatrix} | \\ \underline{n}_i \\ | \end{bmatrix}, \qquad i = 1, 2, \ldots, N, \qquad (5)$$

where $z_i(t_j)$ represents the measured output of detector i at time t_j. It is clear from equation (5) that, were it not for the measurement noise vectors n_i, the set of N, n-dimensional measurement vectors $\{\underline{z}_i\}_{i=1}^{N}$ would all lie in a two-dimensional subspace determined by the displacement vectors $\underline{x}$ and $\underline{y}$. The noise vectors n_i pull the measurement vectors z_i slightly out of the two-dimensional subspace C_J (displacement subspace) as depicted geometrically in Figure 2.

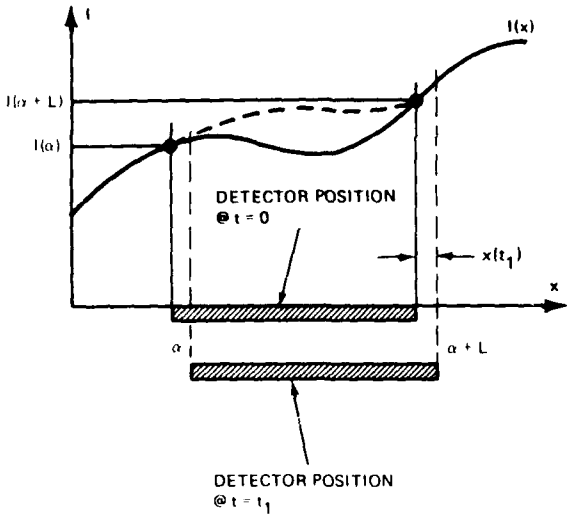


Figure 1. One-dimensional example of displacement effect.

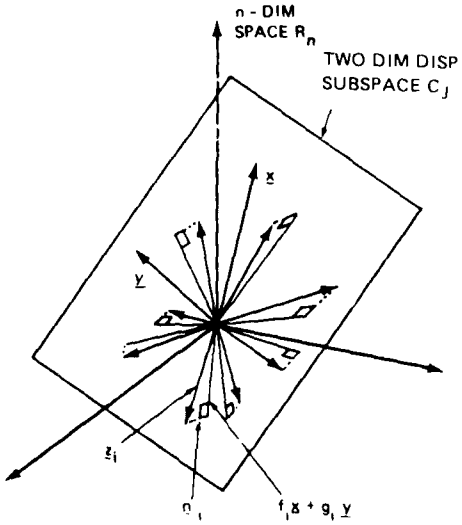


Figure 2. Geometrical representation of the displacement subspace.

It can be easily shown that the best (in the minimum mean square error sense) approximation to C_J is found by first computing the n x n (n = number of measurements) matrix $Z_N \equiv \frac{1}{N} \sum_{i=1}^{N} \underline{z}_i \underline{z}_i^T$. If this matrix is then spectrally factored into a set of rank one matrices

$$Z_N = \sum_{i=1}^{n} \lambda_i^2 \underline{e}_i \underline{e}_i^T, \qquad (6)$$

where λ_i^2 and $\underline{e}_i$ are the ordered ($\lambda_i^2 > \lambda_j^2$, $j > i$) eigenvalues and orthonormal eigenvectors of Z_N, then the approximating subspace $\hat{C}_J$ is determined by the span of the vector set ($\underline{e}_1$, $\underline{e}_2$). The best approximation of the displacement content of each $\underline{z}_i$ is then obtained by projecting $\underline{z}_i$ onto the space $\hat{C}_J$; i.e.,

$$\underline{\hat{z}}_i = \left\{ \begin{bmatrix} | & | \\ e_1 & e_2 \\ | & | \end{bmatrix} \begin{bmatrix} — e_1^T — \\ — e_2^T — \end{bmatrix} \right\} z_i \qquad (7)$$

data vector

projection operator P onto C_J

best approximation of displacement effects vector

This procedure is represented geometrically in Figure 3, as are the types of errors which will occur due to measurement noise effects. One portion of the error ($\underline{z}_i - \underline{f}_i x - \underline{g}_i y$) is caused by the error between the subspaces C_J and $\hat{C}_J$ and another is caused by the component of $\underline{n}_i$ which lies within the subspace C_J.

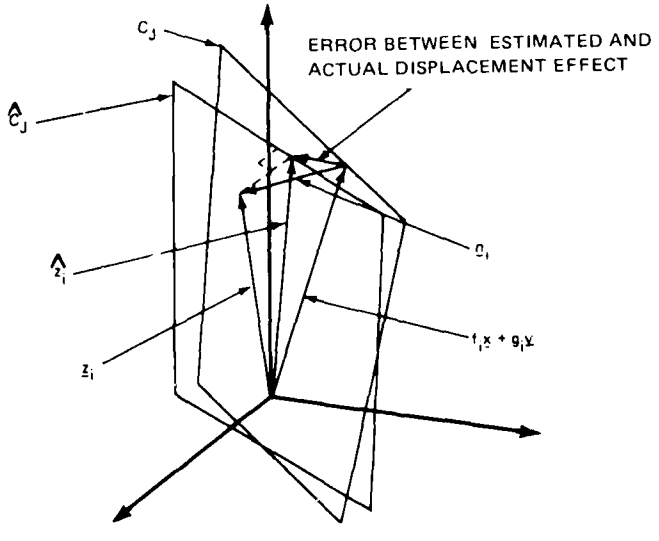


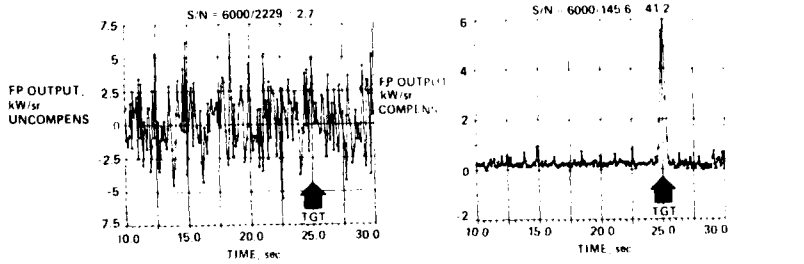
Figure 3. Displacement effect estimation technique.

At this point one can conclude that the actual displacements $\underline{x}$ and $\underline{y}$ are impossible to determine as well as are the local gradients f_i and g_i. The best that can be done is to determine a reasonable approximation to the two dimensional subspace in which the displacement effect vector is contained. This, however, is usually very good indeed, since by projecting the data

vectors (with DC suppressed) onto this subspace one can estimate and remove (i.e. register) the displacement effect vector from the data with very little effect on other temporally changing "patterns of interest."

It can be shown that as the number of detectors and time samples tend toward infinity perfect registration of the image sequence can be obtained without any attenuation of other temporally varying "patterns of interest." The remarkable aspect is that this can be accomplished without ever determining the actual displacement now with any use of high fidelity scene information. This is to be contrasted with the classical cross-correlation techniques in which the displacments are determined and the scenes actually displaced by that amount.

These techniques were applied to simulated data generated by a staring space based surveillance system with a 500 m pixel groundprint situated over the Santa Cruz area of California. The sensor line-of-sight was randomly jittered with an in-band (1 - 10 Hz) rms value of .1 arcsec and image frames collected every .1 second. Typical results of image sequence registration are shown in Fig. 4, for a typical detector through which a target aircraft flew at t = 25 seconds. The event is totally obscured when unregistered data is examined and easily visible when registered data is used. The interested reader is referred to Klop and Barry[1] for a more detailed discussion of this particular application.



(a) Without displacement effects compensation (b) With displacement effects compensation

Figure 4. Performance over Santa Cruz.

References

1. "Rejection of Jitter-Induced Noise in Staring Sensors," M. Klop and P.E. Barry, presented at 8th DARPA Strategic Space Symposium, 22-25 June 1982.

Partial Shape Recognition using Fourier-Mellin Transform Methods

Timothy A. Grogan and O. Robert Mitchell
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907

I. Introduction

In recognizing shapes automatically by computer, a problem often arises when the unknown object is partially obscured or poorly segmented. Most algorithms described in the literature use syntactic methods. A major problem is describing a proper grammar for several objects. This problem is even further complicated when each object may be imaged from many possible aspect angles. The algorithm in this paper avoids this problem. It uses a global method to normalize scale and starting point, even when part of the contour is missing or incorrect.

If the range to the object is unknown, the fraction of its contour that corresponds to the prototype object contour is unknown. The Mellin transform [1] of the Fourier transform magnitude of the curvature function for both contours is calculated. These are multiplied and the Fourier transform applied yielding a crosscorrelation independent of starting point and scale. The shift of the peak of this cross correlation is the logarithm of the "best" scaling needed to match the unknown and prototype curvature functions.

Once this information is available, the unknown is scaled and crosscorrelated to determine the "best" starting point shift. The unknown and prototype curvature functions now have commensurate scales and corresponding starting points and thus, can be directly compared.

II. Curvature Function

The feature processed by the algorithm is the curvature function. The curvature function is used as the feature because it is invariant to rotation and to scale changes resulting from an unknown range. Thus, only missing contour segments or additional segments will result in disparate scales between the unknown and prototype. This assumes that the unknown and the prototype have the same length. This is insured by the preprocessing [2]. Given a silhouette of an object, the contour is traced. Each contour is processed by smoothing it as a function of arc-length between critical points. Critical points are those points on the contour where the angle changes by more than 90degree. The points between these critical points are used to perform a least-squares polynomial fit. The polynomial approximation is used to over sample the contour between the critical points. Each contour is then resampled to a standard number of points. This preprocessing helps to insure that a smoothed representation of the contour is obtained while preserving important angle information and near uniform tracing of the contour. Let the periodic curvature function be $z(t) = x(t) + j\,y(t)$. The contour function is then used to calculate the angle and its derivative, the curvature function. The curvature function is $c(t)=\frac{d}{dt}\tan^{-1}\frac{y(t)}{x(t)}$. The curvature function is then smoothed to facilitate the correlation process to follow.

This research supported by the U. S. Army Research Office.

III. Mellin Transform

The Mellin transform is to be used because of its scale invariance property. The definition of the Mellin transform is

$$M_F(ju) = \int_0^\infty F(x)\, x^{-ju-1} dx$$

If $G(x) = F(ax)$, then $M_G(-ju) = a^{-ju}\, M_F(-ju) = e^{-ju \ln a}\, M_F(-ju)$.

Since the signal is discrete, the Direct Mellin Transform (DMT) is calculated at N-1 points [3]. The DMT is defined as

$$M_F(-ju_i) = \sum_k \Big[\cos(u_i \ln k) - j \sin(u_i \ln k)\Big]\, \Delta_k / -ju_i$$

where $\Delta_k = f(k) - f(k+1)$ and $u_i = \frac{2\pi}{N} i$, $i=1,\ldots,N-1$.

IV. Fourier-Mellin Technique

In order to match an unknown curvature function to a prototype, it is necessary to eliminate first the unknown *shift (or starting point)* ambiguity. To do this the magnitude of the Fourier transform is calculated. That is, we know that if $g(t) = f(t-t_0)$, then $\mathbf{F}\{g(t)\} = \mathbf{F}\{f(t)\} \times e^{-j\omega t_0}$. So, $\left|\mathbf{F}\{g(t)\}\right| = \left|\mathbf{F}\{f(t)\}\right|$.

Now, using the Mellin transform, a scale estimate is obtained. Let the unknown have scale 'a' and shift 't_0', i.e. let

$$g(t) = f\Big[a(t-t_0)\Big].$$

Then

$$\left|\mathbf{F}\{g(t)\}\right| = \left|\frac{1}{a} F(\frac{\omega}{a})\right| \text{ , where } \quad F(\omega) = \mathbf{F}\{f(t)\}.$$

Thus, the Mellin transform of the magnitude of the Fourier transform of the scaled version is related to the Mellin transform of the magnitude of the Fourier transform of the original by

$$M_G(-ju) = \frac{1}{a} a^{ju} M_F(-ju), \text{ where } \quad M_F(-ju) = \mathbf{M}\{\left|\mathbf{F}\{f(t)\}\right|\}.$$

So, if the Fourier transform of M_F times M_G^* is performed, a crosscorrelation is obtained whose shift is the logarithm of the scale, i.e.

$$\mathbf{F}\{M_F\, M_G^*\} = C_{FF}(r - \ln a), \text{ where } \quad C_{FF}(r) = \mathbf{F}\{M_F(-ju) \times M_F^*(-ju)\}.$$

V. Experimental Results

An experiment is performed by generating the contour for an F104 airplane using a computer graphics program (see Fig.1a). It is preprocessed as described in Section II to obtain the smoothed curvature function in Fig 2a. This curvature function is used as the prototype. The prototype curvature function is also processed by a program that chops out 10% of the contour and replaces it with a line segment (see Fig.1b). The smoothed curvature function for this contour is calculated. This smoothed curvature function is the unknown (see Fig.2b). The magnitude of the Fourier transform for both curvature functions is calculated using a 256 point FFT algorithm. After taking the magnitude of the Fourier transform only the positive frequency

components are needed. Also, the DC component is set to zero. This is necessary because the first point in the signal passed to the DMT has an overwhelming effect. Now the DMT of both are calculated and multiplied as previously described. Instead of calculating the Fourier transform of the multiplied functions over a large range in shifts, the Fourier transform is calculated in a narrow range about the origin (see Fig. 3). This calculation is justified if the logarithm of the scale is known to be small. The position of the peak of the crosscorrelation gives the logarithm of the scale. It is used to resample the unknown to a new scale. The scaled version of the unknown and the prototype are crosscorrelated (see Fig. 4). The peak in this signal provides the shift in starting point necessary to properly align the unknown and the prototype. Now the unknown and the prototype can be directly compared by taking the difference between the scaled & shifted unknown and the prototype (see Fig. 5 & Fig. 6). Simply thresholding the difference signal will provide the interval over which the unknown and prototype correspond.

VI. Conclusion

A general method for determining partial matches between an unknown and a prototypical shape has been described. At present the computational complexity is high, but it may compare favorably with the complex combinatorial methods previously reported. Further study is required to show if it is possible to reduce the required resolution of the Fourier and Mellin transforms thereby reducing the computational load. Another consideration would be to research the classes of functions having the same Fourier-Mellin crosscorrelation resulting in an ambiguity between shapes.

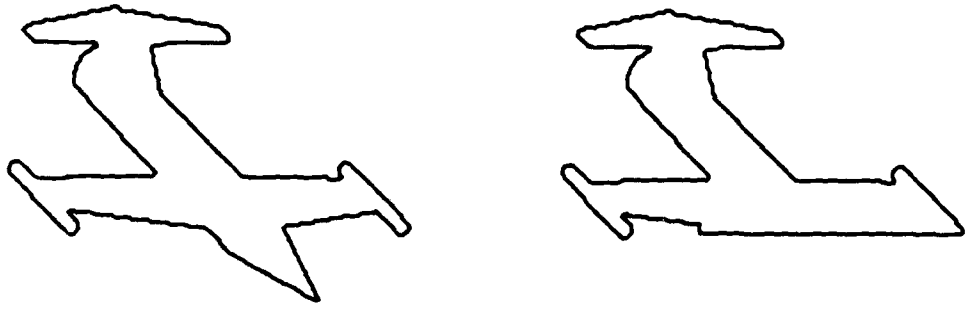


Fig. 1 (a) F104 airplane contour (prototype); (b) partial contour (unknown).

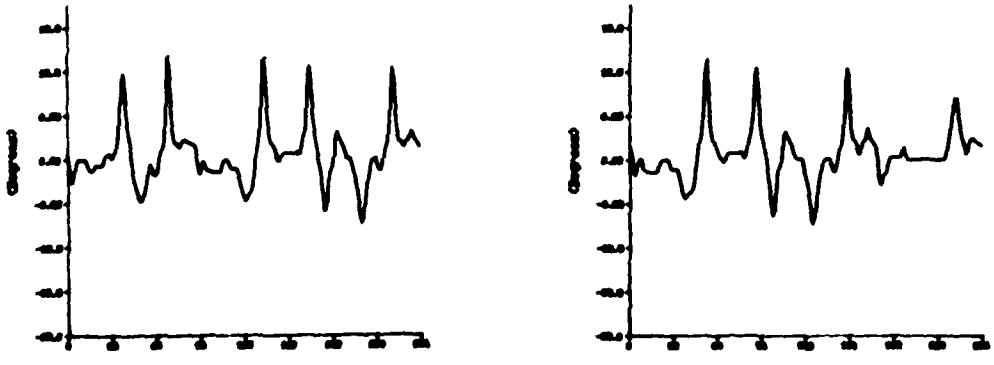


Fig. 2. Curvature functions for the two contours of Fig. 1.

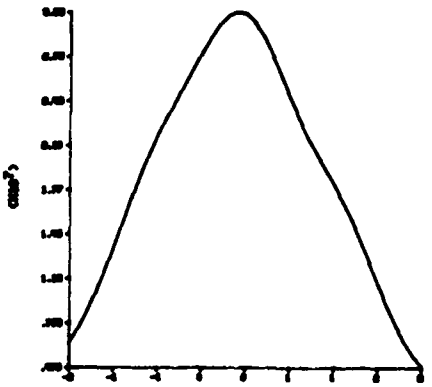


Fig. 3. Mellin crosscorrelation result. The offset of the peak corresponds to logarithm of the scale difference between the matching portions of the two curvature functions in Fig. 2. The scale factor is 0.90484.

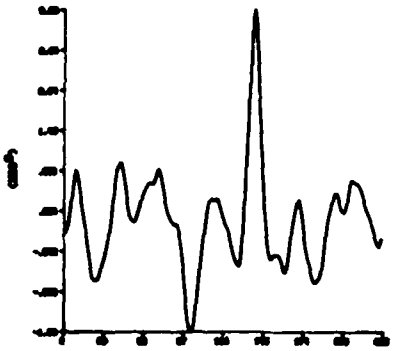


Fig. 4. The crosscorrelation between the scaled unknown and the prototype curvature functions. The offset of the peak corresponds to the shift in starting point. The circular shift is 140 samples.

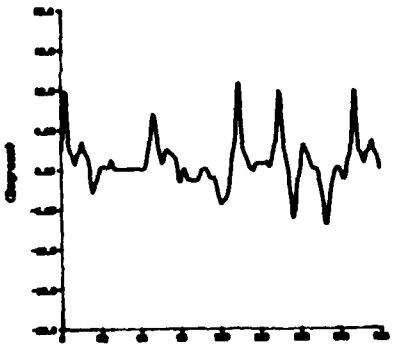


Fig. 5. The scaled and shifted curvature function of the unknown.

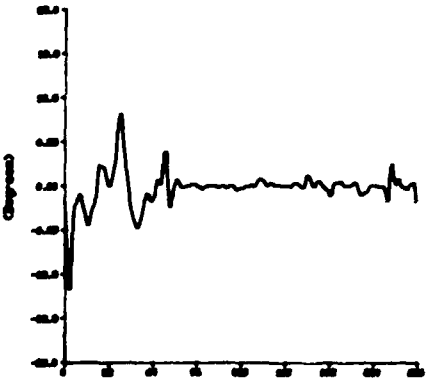


Fig. 6. The difference signal between Fig. 1(a) and Fig. 5. A 3 point median filter has been applied to remove impulsive differences due to subsample shifts.

VII. References

[1] Casasent, David and Psaltis, Demetri, "New Optical Transforms for Pattern Recognition", *Proc. IEEE*, vol. 65, pp. 77-84, Jan. 1977.

[2] Wallace, T. P., Mitchell, O. R., and Fukunaga, K., "Three-Dimensional Shape Analysis using Local Shape Descriptors," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. PAMI-3, pp. 310-323, May 1981.

[3] Zwicke, Philip E., and Kiss, Imre Jr., "A New Implementation of the Mellin Transform, and its Application to Radar Classification of Ships," to appear in *IEEE Trans. Pattern Analysis and Machine Intell.*

SINGULAR VALUE ANALYSES OF INVERSION OF LAPLACE AND OPTICAL IMAGING TRANSFORMS

M Bertero, Istituto di Scienze Fisiche
dell' Università and Istituto Nazionale di Fisica
Nucleare, Genoa, Italy

and

E R Pike, Physics Group
Royal Signals and Radar Establishment
Malvern, UK

The use of the eigenvalues and eigenfunctions of the first order Fredholm equations describing optical imaging has long been known. The eigenvalues and eigenfunctions of the first order Fredholm equation of the Laplace transform have only recently been discovered and similarly used (McWhirter and Pike 1978). Let us consider for simplicity the one dimensional case with magnification unity. For an eigenfunction to be defined the linear mapping $A: f \to g$ of the "object" f into its "image" g must define a compact bijective operator A of, say, $L^2(-1,+1)$ into itself.

$$(Af)(t) = \int_{-1}^{1} \frac{\sin c(t-s)}{\pi(t-s)} f(s)ds \qquad |t| \leq 1 \qquad (1)$$

We require the object and image to be defined over identical domains. In a real optical system the physical image g' will be continued outside this support, if finite, and technically the linear mapping $K: f \to g'$ defines a compact injective operator K of $L^2(-1,+1)$ into $L^2(\infty,\infty)$.

$$(Kf)(t) = \int_{-1}^{1} \frac{\sin c(t-s)}{\pi(t-s)} f(s)ds \qquad -\infty < t < +\infty \qquad (2)$$

As a consequence when the entire image is considered the eigenvalue analysis must be supplanted by a singular value analysis (Bertero and Pike 1982).

The eigenfunctions u_k and eigenvalues λ_k of A obey

$$Au_k = \lambda_k u_k \quad , \qquad (3)$$

where

$$u_k(t) = \lambda_k^{-\frac{1}{2}} \psi_k(ct) \tag{4}$$

and where ψ_k are the linear prolate spheroidal functions.

The adjoint operator K* obeys

$$(K^*g)(t) = \int_{-\infty}^{\infty} \frac{\sin c(t-s)}{\pi(t-s)} g(s)ds \quad , \qquad |t| \leq 1 \quad . \tag{5}$$

Bertero and Pike (1982) show that

$$K^*Ku_k = \lambda_k u_k \tag{6}$$

and that we have a singular system obeying the coupled equations

$$Ku_k = \lambda_k^{\frac{1}{2}} v_k \tag{7}$$

$$K^*v_k = \lambda_k^{\frac{1}{2}} u_k \tag{8}$$

where

$$v_k(t) = \psi_k(c,t) \qquad -\infty < t < \infty \quad . \tag{9}$$

Similarly the eigenfunctions and eigenvalues of the Laplace transform found by McWhirter and Pike are defined for a mapping from $L^2(0,\infty)$ into itself, while for an object defined on finite support the mapping $K:f \rightarrow g$ defines the following linear operator from $L^2(a,b)$ into $L^2(0,\infty)$:

$$(Kf)(p) = \int_a^b e^{-pt} f(t)dt \quad , \qquad 0 \leq p \leq \infty \quad . \tag{10}$$

It is easy to show that K is continuous and that Kf = 0 has only the trivial solution f = 0 and therefore K is injective. The adjoint operator K* is given by

$$(K^*g)(t) = \int_0^\infty e^{-tp} g(p)\,dp \qquad a \leq t \leq b \tag{11}$$

and it is a linear compact operator from $L^2(0,\infty)$ into $L^2(a,b)$. Since K is injective the range of K* is dense in $L^2(a,b)$. K* is also injective and therefore the range of K is dense in $L^2(0,\infty)$. From these properties it follows that the operator K admits a singular system $\{\alpha_k;\ u_k,\ v_k\}_{k=0}^{\infty}$ given by the solutions of the coupled equations

$$Ku_k = \alpha_k v_k \tag{12}$$

$$K^*v_k = \alpha_k u_k \quad . \tag{13}$$

Since the null spaces of K and K* contain only the null elements, the singular values α_k must be strictly positive and $\{u_k\}_{k=0}^{\infty}$ $\{v_k\}_{k=0}^{\infty}$ are bases, respectively, in $L^2(a,b)$ and $L^2(0,\infty)$. As is well known the singular functions u_k are the eigenfunctions of the operator KK* with eigenvalues α_k^2

$$KK^*u_k = \alpha_k^2 u_k \qquad k = 0,\ 1,\ 2 \ldots \quad . \tag{14}$$

From equations 10 and 11 one may show that

$$(KK^*f)(t) = \int_a^b \frac{f(s)}{t+s}\,ds \qquad a \leq t \leq b \tag{15}$$

and therefore KK* is of trace class

$$\text{trace } (KK^*) = \sum_{k=0}^{\infty} \alpha_k^2 = \int_0^\infty \frac{dt}{zt} = \frac{1}{2}\log\gamma \tag{16}$$

where

$$\gamma = b/a \quad . \tag{17}$$

The singular values α_k, in fact, depend only on the parameter γ. This may be seen by using equation (14) with a transformation to new variables

$$t = a + (b-a)x \quad , \qquad s = a + (b-a)y \tag{18}$$

to find

$$\int_0^1 \frac{\chi_k(y)}{x + y + 2/(\gamma - 1)}\, dy = \alpha_k^2 \chi_k(x) \tag{19}$$

where

$$u_k(t) = \frac{1}{\sqrt{b - a}}\, \chi_k\left(\frac{t - a}{b - a}\right) . \tag{20}$$

The effect of noise on inversion using the singular value analysis may be shown to be less severe than using eigenvalues, as is to be expected since in the first case more information is used in the full image than in the image between -1 and +1 and in the second case the restriction of the object to finite support serves also to increase the available information for the inversion.

The true information-theoretic resolution limits for these two situations have been calculated and significant gains over the conventional Rayleigh limit and the equivalent McWhirter-Pike limit for the Laplace transform may be achieved, the more so, the greater the difference between the image and object domains.

Numerical simulation has verified these conclusions and they have important implications in the fields, for example, of polydispersity analysis by laser scattering from macromolecular suspensions, where in fact, they are already in use, and of diffraction limited imaging systems.

REFERENCES

McWhirter, J.G. and Pike, E.R., 1978 J Phys A, 11, 1729.
Bertero, M., and Pike, E.R., 1982 Optica Acta 29, 727.

A technique for the calculation of the global extremum of a function of several variables.

C.H.Slump and B. J. Hoenders, Department of Applied Physics, State University at Groningen, Nijenborgh 18, 9747 AG Groningen, the Netherlands.

Introduction

The determination of the global extremum of a function is a notorious numerical problem when there are local extrema present. The numerical algorithm which has to determine the global extremum iteratively is in this case very likely to produce a local extremum in the neighborhood of the initial guess of the solution. Moreover, usually one cannot be sure not to have missed an extremum, i.e. various procedures together with various initial values might still overlook the global extremum.

It would therefore be of great value if the total number of stationary points of a function in a certain domain could be calculated easily and exactly: This information would tell us whether or not the numerical algorithm had determined all stationary points.

The stationary points of a function $f(\underline{x})$ are the zeros of the set of eqs. $\nabla f(\underline{x})=0$, and we are therefore looking for a simple procedure determining the number of zeros of a set of eqs. in a certain domain D in the $\underline{x}$-space.

Such a procedure is provided by an integral derived by Picard [1] from previous work by Kronecker [2]. The integration is over the domain of interest, whereas the integrand only contains simple algebraic functions (viz.eqs.(1), (2), (3)). An extensive discussion of this integral is given by Hoenders and Slump [3].

This so-called Picard-Kronecker integral (P.K.integral) is equal to the number of zeros of a set of eqs. in a certain domain provided that all these zeros are simple. The case of multiple zeros will be discussed in a forthcoming paper. (See also Davidoglou [4] and Tzitzéica [5]).

We will illustrate the use of this elegant formalism with the following problem: Estimate the position a and width σ of a Gaussian wave in a Poissonian pulse train using the maximum likelihood method.

The estimation boils down to the determination of the absolute maximum of the likelihood function in a certain domain of the two parameter space $\underline{x} = (a,\sigma)$ (We neglect the possibility of extreme values occurring at the boundary which is only an inessential complication).

The P.K. integral then provides an excellent tool for the calculation of the total number of stationary points of the likelihood function $L(\underline{x})$ in a domain. This number is very useful for the calculation of these points.

We determine to this end with the P.K. integral the total number of roots of the set of eqs. $\nabla L(\underline{x}) = 0$ in the domain of interest and subdivide this domain (nesting) till only one or no root is located in the domain. The position of a stationary point can then eventually be determined with zero locating numerical procedures, e.g. Powell [6].

Theory

The basic idea of the P.K. integral is the generalisation of the concept of the solid angle into higher dimensional space. It is then intuitively clear that the solid angle is a measure for the number of zeros of a set of functions in a certain domain by the following argument: Suppose that in two dimensions the transformation: $x = x(u,v)$ and $y=y(u,v)$ admits n zeros in a certain domain D with boundary S of the uv-plane. One would then expect that the solid angle connected with the surface S is equal to $0, \pm 2\pi, \pm 4\pi, \ldots \pm 2\pi n$ depending on the orientation of the mapping $(x,y)\rightarrow(u,v)$ in the neighborhood of zeros. This idea was put into an exact analytical form by Kronecker [2]. (See for a modern formulation and application of this concept in algebraic topology and functional analysis Schwartz [7]). The Kronecker integral is not conclusive for the calculation of the number of zeros as it is proportional to the number of zeros with positive Jacobian minus the number of zeros with negative Jacobian.

Picard [1] showed how to use this idea to obtain the exact number of zeros of a set of eqs. in a certain domain by a very simple extension of the original set of eqs. He derived that the number n of simple zeros of the function $y = f(x)$ in the interval $a \leq x \leq b$ is equal to:

$$n = -(\pi)^{-1} \varepsilon \int_a^b \frac{f(x)\, f''(x) - f'^2(x)}{f^2(x) + \left(\varepsilon\, f'(x)\right)^2}\, dx + (\pi)^{-1} \operatorname{arctg}\left(\frac{\varepsilon f'(y)}{f(y)}\right)\Bigg|_a^b, \qquad (1)$$

where ε denotes an arbitrary constant.

The number of simple zeros n of the equations: $f(x,y)=0$; $g(x,y)=0$ in the domain D with boundary S and Jacobian J can be shown to be equal to (no zeros are allowed on S):

$$n = (2\pi)^{-1} \int_S (P\,dx + Q\,dy) + \varepsilon (2\pi)^{-1} \iint_D R\left(f^2+g^2+\varepsilon^2 J^2\right)^{-\frac{3}{2}} dx\,dy, \qquad (2)$$

where $P = (f\frac{\partial g}{\partial x} - g\frac{\partial f}{\partial x})(f^2 + g^2)^{-1}(f^2 + g^2 + \varepsilon^2 J^2)^{-\frac{1}{2}}$, (3)

and Q is obtained from P changing $\frac{\partial}{\partial x}$ into $\frac{\partial}{\partial y}$. R is the determinant $\frac{\partial(zf,zg,zJ)}{\partial(z,x,y)}$ evaluated at $z = 1$.

The r.h.s. of eqs. (1) and (2) can be shown to be independent of ε, Picard [1].

Example.

In e.g. optical communication practice and in the processing of seismic signals one encounters the problem how to detect a partially known wave shape in measurements corrupted by noise. We consider the problem of the estimation of the position a and (half)-width σ of a Gaussian wave shape in a Poissonian pulse train, (see fig.1). The measurements are assumed to consist of $\underline{n} = (n_1, n_2, \ldots n_N)$ Poisson - distributed random variables with parameter

$$\lambda_{t_i}(a,\sigma) = \alpha\left[1 + \beta(2\pi)^{-\frac{1}{2}}\sigma^{-1}\exp\left\{-(2\sigma^2)^{-1}(t_i-a)^2\right\}\right] + \lambda_0, \quad (4)$$

$i = 1, \ldots N$, λ_0 denotes the dark current of the detector, α and β are constants (see fig.1), a and σ are estimated by the absolute maximum of the likelihood function:

$$L(\underline{n};a,\sigma) = \prod_{i=1}^{N} \exp(-\lambda_{t_i}(a,\sigma))\lambda_{t_i}(a,\sigma)^{n_i}/n_i!. \quad (5)$$

Applying the method described above to the set of eqs. $\frac{\partial}{\partial a}\ln L = 0$, $\frac{\partial}{\partial \sigma}\ln L = 0$ reveals that there are 4 zeros located in the region $50.0 \leqq a \leqq 80.0$; $2.5 \leqq \sigma \leqq 12.5$. Subdividing the domain and applying [6] leads to the stationary points in table 1.

Acknowledgement.

One of us (C.H.S.) acknowledges the financial support by the Netherlands Organization for the advancement of Pure Research (Z.W.O.).

References:

[1] Picard,E.,Journ.de Math.Pure et Appl.(4e série)8, p.5-24, 1892.
[2] Kronecker,L.,Werke,vol.1,p. 175-212 and p. 213-226,Leipzig:Teubner, 1895.
[3] Hoenders,B.J.,Slump,C.H.,submitted to Computing.
[4] Davidoglou,A.,Compt.Rend.hebd.,133,p. 784-786, p.860-863, 1901.
[5] Tzitzéica,G.,Compt.Rend.hebd.,133,p. 918-920, 1901.
[6] Powell,M.J.D. in:Numerical methods for non-linear algebraic equations,ed. by Rabinowitz,P.,Gordon and Breach, 1970.
[7] Schwartz,L.Cours d'analyse, vol.2. p. 230-269,Paris: Hermann 1967.

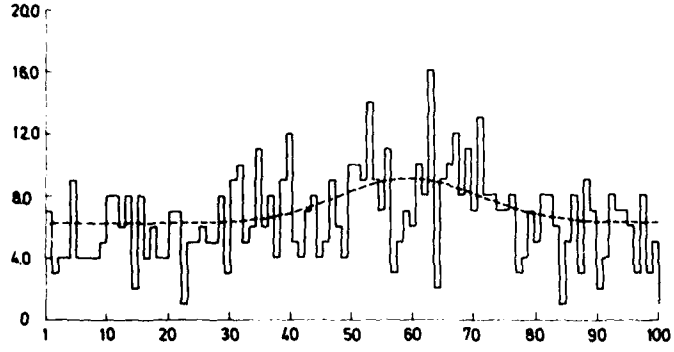


fig.1. The simulated Poissonian pulse train, (drawn line) and λ_t (dotted line), with the values: α =1.75 β=18.0 a =60.0 σ= 8.0 λ_0= 4.0 N=100

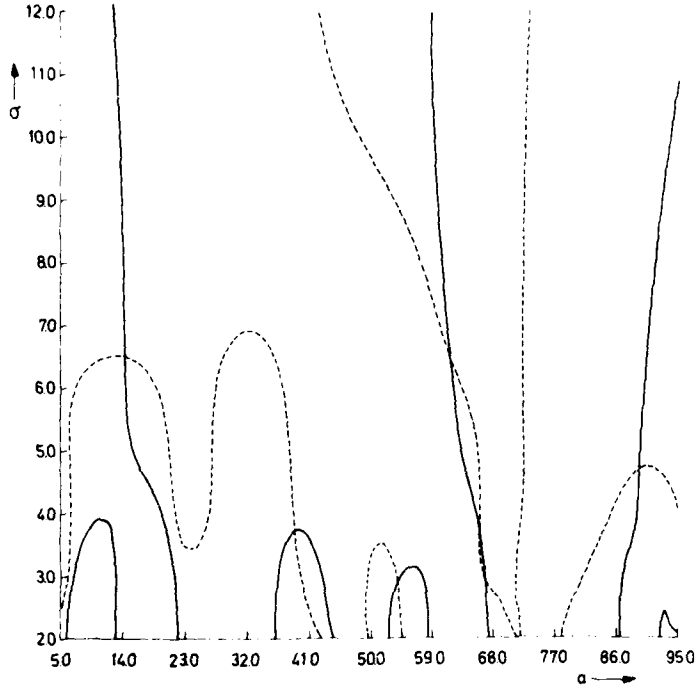


fig.2. $\frac{\partial}{\partial a}$ ln L() = 0 (drawn line)
$\frac{\partial}{\partial \sigma}$ ln L() = 0 (dotted line)

Table 1.

a	σ	Magnitude sequence
53.74	4.12	4
62.88	7.59	2
65.47	5.78	3
66.87	4.26	1

SESSION IX

INTERPOLATION AND TOMOGRAPHY I

N. Hurt, *Presider*

An Algorithm for Incomplete Range of Views Reconstruction

Heang K. Tuy

Medical Imaging Section, Dept. of Radiology
Hospital of the University of Pennsylvania
3400 Spruce Street, Philadelphia, PA 19104

I. Introduction. To reconstruct a cross-section of a 3D object, most algorithms require knowledge of the projection data in a full range of views [1]. In some practical situations [2, 3], reconstruction from an incomplete range of views is inevitable although it is not desirable from a mathematical point of view. Objects to be reconstructed are of compact support. Their Fourier transforms can be extended to an entire function of exponential growth (band-limited function). Consequently, there is a unique solution to the incomplete range of views reconstruction problem. On the other hand, the problem is an ill-posed problem. For example, it has been indicated [4] that the spectrum of the singular values of the Radon transform for limited range of views is split up into two parts. One part consists of singular values near one, and the other part consists of singular values near zero. The recovery of small singular values is necessary for a process to reconstruct objects with good quality. This, however, exacerbates instability in the process, i.e., a small error in the projection data might lead to an undesirable large difference in reconstructed images. Making use of a priori information on image and projection data is being examined to reduce this instability.

In the following discussion, we reformulate the problem of reconstruction from an incomplete range of views using a priori information as an intersection problem. More precisely, the image solution is shown to be a point which belongs to the intersection of a finite number of convex subsets in a Hilbert space. This view allows the derivation of an iterative algorithm for limited range of views reconstruction problems. In the experimental results reported below, we illustrate the importance of the use of qualitative a priori information [5, p. 27] in both parallel and fan-beam geometry.

II. The algorithm. The derivation of the algorithm is based on the following theorem [6]:

Theorem. Let $\{C_j\}_{j=1,\ldots,m}$ be a family of convex subsets in a separable Hilbert space H. Let $f_0 \in H$, and $\{f_n\}$ be a sequence in H defined by

$$f_{n+1} = f_n + r_n(P_{j_{n+1}}(f_n) - f_n)$$

where $0 < e_1 \leq r_n \leq 2 - e_2 < 2$ for some positive real numbers e_1 and e_2, $j_{n+1} = j_n \pmod m + 1$, and $P_j(g)$ is the orthogonal projection of g onto the convex set C_j, for $j = 1,\ldots,m$. Then if the intersection of the C_j's is not empty, the sequence $\{f_n\}$ converges weakly to a point $f \in \bigcap_{j=1}^{m} C_j$.

A geometrical illustration of the theorem corresponding to the case where the relaxation parameters $r_n = 1$ is given in Fig. 1.

Suppose H is $L^2(\mathbf{R}^2)$, the space of square integrable functions defined on $\mathbf{R}^2$, and the attenuation of a cross-section is represented by a function f in H. Let p be the projection data in a given range Γ (incomplete range of views). If S represents the shadowgraph operator (the Radon transform R in the case of parallel beam, and the divergent transform D in the case of fan-beam) then f belongs to a convex subset C_1 of H where $C_1 = \{g \mid Sg = p\}$. The orthogonal projection onto C_1 is given below:

Proposition: Let $q(\rho,\theta) = p(\rho,\theta)$ if θ is in the range Γ, $q(\rho,\theta) = Sg(\rho,\theta)$ otherwise. Denote by $\bar{R}(q)$ the inverse 2D Fourier transform of $\tilde{q}$, which is defined in polar coordinates by $\tilde{q}(s,\theta) = F_1(\, q(\rho,\theta)\,)\,(s)$, the 1D Fourier transform of q with respect to the first variable ρ, evaluated at the point s.

If S is the Radon transform R then $P_1(g) = \bar{R}(q)$.

Proof. Let $g' = \bar{R}(q)$. From the projection theorem [1, p. 149], it follows that g' belongs to C_1, and furthermore, for any point h of C_1, the distance from $\hat{h}$ ($\hat{h}$ is 2D Fourier transform of h) to $\hat{g}$ is at least the distance from g' to $\hat{g}$. Since the Fourier transform is an isometry, g' is indeed the orthogonal projection of g onto C_1.

Note that $\bar{R}$ is a natural extension of the inverse Radon transform to the whole space H.

Quantitative a priori information. Sometimes the following information is available: the attenuation function f to be reconstructed is bounded from below by a function ℓ and bounded from above by a function u. In this case the function f belongs to two other convex subsets C_2 and C_3 of H, where $C_2 = \{g \mid g \geq \ell\}$, and $C_3 = \{g \mid g \leq u\}$. The analytic expressions of the orthogonal projections onto these two convex subsets are given in [7].

Qualitative a priori information. If we assume further that the function f to be reconstructed and its shadowgraph Sf are piecewise smooth, then f is an element of the following two convex subsets: $C_4 = \{g \mid g$ is piecewise smooth$\}$, and $C_5 = \{g \mid Sg$ is piecewise smooth$\}$.

Based on the theorem previously stated, an algorithm to reconstruct f making use of the above a priori information can be stated as follows:

Starting from an arbitrary image f_0, produce a sequence of images $\{f_n\}$ by successive orthogonal projections (with or without relaxation) onto the convex subsets C_1, C_2, C_3, C_4 and C_5.

III. Experimental results. There is some evidence showing the importance of the use of quantitative a priori information; see for example [8, 9, 10]. In this article, the only quantitative a priori information used is the non-negativity of the attenuation of the object to be reconstructed. Our experimental results show the major role played by qualitative a priori information in the reconstruction process. Bearing in mind that knowledge of this type of a priori information is limited, we shall assume only that the object to be reconstructed and its shadowgraph are piecewise smooth. Based on the above algorithm, an analytic expression for the orthogonal projection onto the subset of piecewise smooth functions is required. Such an expression is not known to the author. However, since practical results are of primary interest, a selective smoothing operator [1, p. 193] is used instead of the orthogonal projection operators suggested by the above theoretical work. A schematic representation of the algorithm used in this experiment is given in fig. 2.

The phantom (fig. 3a), which was mathematically simulated, represents a cross-section of a human thorax. Each reconstructed image (and the phantom) was embedded in a square region of 127*127 square pixels each of area $.3*.3cm^2$. The projection data used were mathematically simulated, noiseless, and equally spaced in the given range of views. The dimensions of the projection data will be denoted by (npr, nrays, pinc), where npr is the total number of projections, nrays the number of rays per projection, and pinc the distance between two consecutive rays in a projection.

Parallel beam. The range of the projection angles is from -67° to 67°. Data are missing around the vertical axis. We start off with the projection data set of dimensions (135, 127, .3). We apply the algorithm where the only a priori knowledge used is the non-negativity. The starting point is obtained by applying the convolution algorithm to the projection data filled with zero in the missing range [11] (fig. 3b). As can be seen in fig. 3c, there is some improvement in the quality of reconstructed images up to the 11th iteration. However, the pictures present some undesirable artifacts. The streak artifacts cannot be removed by increasing the number of rays in the projection data set. Fig. 3d shows different streak pattern for projection of dimensions (270, 255, .15). However, there is a big improvement if the reconstructed images are selectively smoothed at each iteration (fig. 4a). A noticeable improvement is also observed in the case where both the reconstructed images and the calculated pseudo-projections are selectively smoothed (fig. 4b).

Fan-beam. Since we do not have an analytic expression for the orthogonal projection on the subset C_1 in this case, an operation similar to the one used in the parallel beam case is employed. In spite of this, our experimental results show that the situation in the fan-beam case is rather more favorable than the situation in the parallel beam case, as illustrated in figs. 4c, 4d. The projection data set was of dimensions (133, 255, .2). The range of views was from 97.5° to 262.5° (Geometry: distance from origin to x-ray source was 78cm, from origin to detectors was 110.7cm).

Fig. 5 shows the graph of the L^2-norm error between the phantom and each reconstructed image. The curve 5a is the error curve corresponding to the algorithm producing the image 3d. Similarly, 5b corresponds to 4a, 5c to 4b, and 5d to 4d. Note that the use of more a priori information not only accelerates the convergence of the iterated sequence but also brings reconstructed images closer to the true image before the error starts to shoot up.

Acknowledgements. This work is supported by NIH grants RR01372, HL4664 and HL28438.

References

1. G.T. Herman, Image Reconstruction from Projections, Academic Press (1980).
2. W.H. Berninger, et al, J. CAT, v. 3, no. 2, pp. 155-163 (1979).
3. D. Townsend, et al, IEEE Trans. Nucl. Sci. NS-27, pp. 463-469 (1980).
4. M.E. Davison, to appear in SIAM J. App. Math.
5. A. Tikhonov, V. Arsenin, Solutions of Ill-posed Problems, Winston Willey (1977).
6. L.G. Gubin, et al, USSR Comp. Math. & Math. Phys. v. 7, no. 6, pp. 1-24 (1967).
7. A. Lent, H. Tuy, J. Math. Anal. & App., v. 83, no. 2, pp. 554-565 (1981).
8. A. Jeavons, et al, Proc. IAEA-SM-247/65 (1980).
9. K.C. Tam, IEEE Trans. Nucl. Sci. NS-29, pp. 512-515 (1982).
10. H. Stark, et al, J. Opt. Soc. Am., v. 72, no. 8, pp. 993-1000 (1982).
11. H. Tuy, J. Math. Anal. & App. v. 80, no. 2, pp. 598-616 (1981).

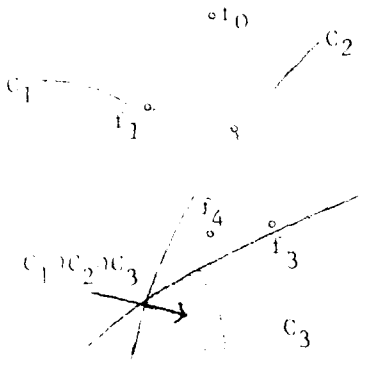


Fig. 1

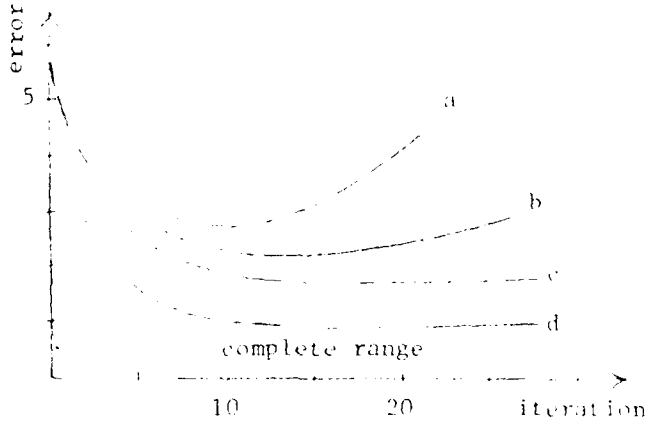


Fig. 5

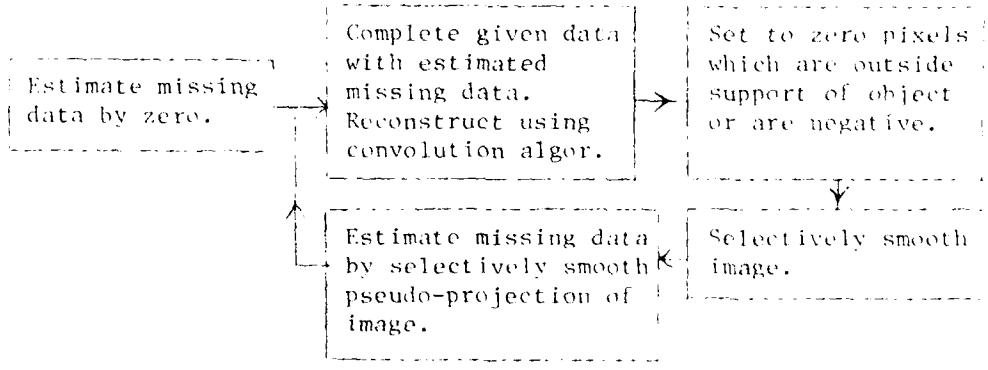
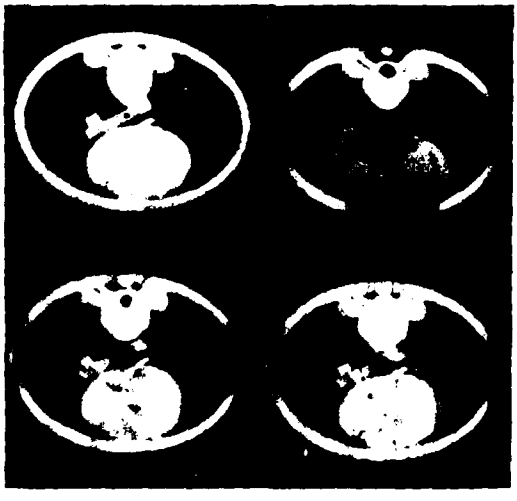


Fig. 2

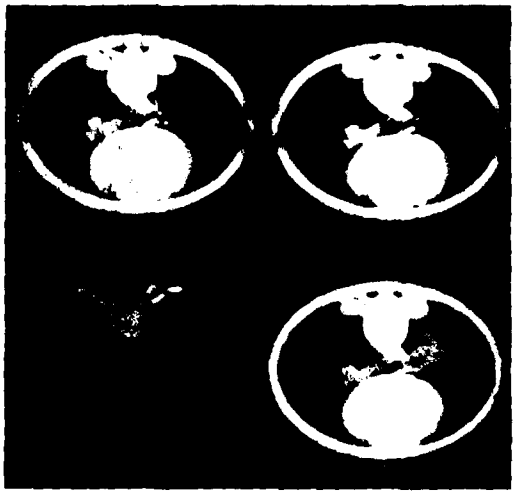
3a 3b



3c 3d

Fig. 3

4a 4b



4c 4d

Fig. 4

FA2-1

GEOMETRIC DECONVOLUTION OF ARTIFACTS IN LIMITED VIEW COMPUTED TOMOGRAPHY

Dr. Rangaraj M. Rangayyan
Electrical Engineering Department
University of Manitoba, Winnipeg, Canada R3T 2N2.

and

Dr. Richard Gordon
Departments of Pathology and Radiology
University of Manitoba, Winnipeg, Canada R3E 0W3.

Images reconstructed using a limited number of projections measured over a narrow angle range are characterized by elliptic distortions along the directions of the rays used, and poor contrast at angles not used (anisotropic resolution). Thus, for example, a reconstruction computed using a few views measured about the vertical would have a vertical ellipsoidal distortion and poor resolution along the horizontal. We are interested in this particular case in connection with our experiments to achieve computed tomography from a few radiographs acquired at different angles using an ordinary overhead or mammographic x-ray unit. The aim is to transmit these images over telephone lines to provide inexpensive computed tomography to people living in remote areas. The angular coverage is restricted to the range 55-125 degrees. We are also attempting to compute high resolution tomographic images of the breast from a few film mammograms. Similar cases arise in industrial non-destructive testing and electron microscopy of biological macromolecules.

The ordinary reconstruction process introduces a systematic geometric distortion of the image. We have the following problem: Given a distorted reconstruction, compute a distortion-free reconstruction of the image. In other words, we wish to deconvolve the geometric distortion. If we restrict ourselves to linear reconstruction algorithms (unconstrained multiplicative ART is an example), the problem comes down to one of estimating the point spread function of the system for a given set of parameters and estimating its inverse. A further simplification can be achieved by applying the projection theorem: we can work on the one dimensional projections instead of the two dimensional image.

The approach we have taken is as follows: Since the reconstructed image has satisfied the projection measurements at all angles for which the projection data have been given (with ART-type algorithms), the distortions lie at the other angles. We estimate the inverse of the one dimensional spread function by deconvolving the projections of a reconstructed test pattern and the corresponding projections of the original test pattern, at angles not used in the reconstruction procedure. In a real case then, after reconstruction using the available projections, we measure the projections of the reconstructed image at other angles. These projections are convolved with the inverse of the spread functions at the corresponding angles. The corrected projections are then used along with the original projections to reconstruct a better estimate of the unknown image. Fast Fourier transform techniques are used to perform the deconvolutions and convolutions.

The problems associated with this procedure are those with division of Fourier spectra: noise amplification and indeterminacy at the zeros of the two spectra. By proper construction of the point spread function, and use of spectral constraints based on a priori knowledge about the projection data, it is possible to get around these difficulties. Once the inverses are computed for a particular linear reconstruction algorithm and given sets of geometric parameters, they should be applicable to any unknown image.

As we carry out our deconvolution on the projections and then use them for the second reconstruction, the second algorithm itself need not be linear. This allows us to incorporate nonlinear and object dependent constraints and a priori knowledge into the second reconstruction procedure. We demonstrate this by the use of our streak preventive algorithm SPARTAF for the second reconstruction. We present initial results obtained in our experiments along these lines.

Localization From Projections Based on Detection and Estimation of Objects*

David J. Rossi
Schlumberger-Doll Research
P.O. Box 307
Ridgefield, CT 06877

Alan S. Willsky
Massachusetts Institute of Technology
Room 35-233
Cambridge, MA 02139

The problem of reconstructing a two-dimensional (2D) function from its 1D projections arises, typically in the context of cross-sectional imaging, in a diversity of disciplines [1-3]. In this problem, a 2D function f(x) is estimated from samples of its Radon transform (line integral measurements)

$$g(t,\theta) = \int_{\underline{x}'\underline{\theta}=t} f(x)ds \qquad (1)$$

where $\underline{\theta}=(\cos\theta \ \sin\theta)'$. The major emphasis of research and applications in this area has been on producing accurate, high-resolution cross-sectional images (requiring a large number of high signal-to-noise ratio (SNR) measurements taken over a wide viewing angle [4,5]) which in practice are post-processed, perhaps by humans, to remove artifacts and extract the information of interest about the cross-section. For example, in nondestructive testing applications [3], reconstructed images are post-processed to determine whether flaws or defects are present within a homogeneous medium; in oceanographic applications, reconstructed images are post-processed to determine where within the cross-section an oceanographic cold-core ring is located [2]. Such post-processing is effectively the utilization of a priori information about the medium being measured to enhance and extract specific pieces of information.

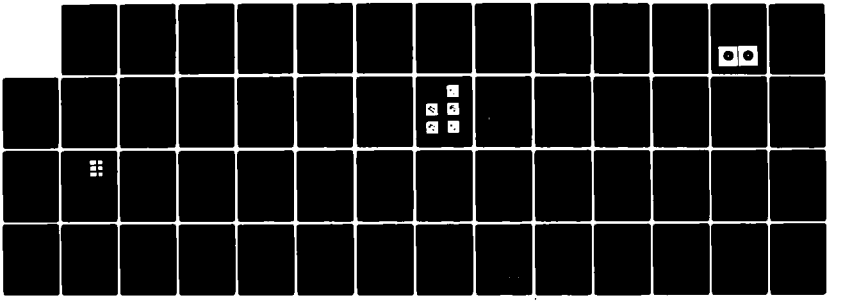
In this paper, we develop a framework for utilizing a priori information directly in the solution to the inverse problem. In particular, we consider using the projection measurements directly (rather than post-processing the reconstruction) in order to detect, locate and characterize objects contained within the cross-section. This framework offers the potential for significant improvements in applications where (1) attempts to perform direct inversion with insufficient measurement data result in severely degraded reconstruction and (2) the ultimate goal of the reconstruction process is to obtain a few specific pieces of information about the cross-section.

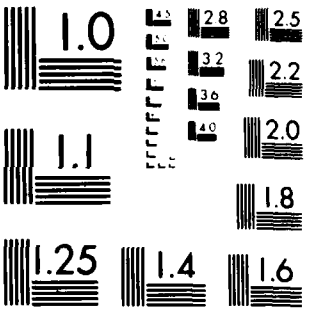
To illustrate this perspective, we consider the case of a cross-section containing a single object; let the cross-sectional density profile f(x) consist of the sum of a deterministic background field $f_1(x)$ and $f_o(x-c_o)$, a single object which is located at some unknown point $c_o \in R^2$ within the cross-section, but whose size, shape, orientation and brightness or contrast relative to the background are known (i.e. $f_o(x)$ is known). We consider the problem of obtaining a maximum likelihood (ML) estimate for the location parameter c_o from full-view projection measurements (i.e. projections available at all t and θ values) corrupted by additive white measurement noise. The case where projection measurements are available over only a limited viewing angle or at a discrete set of projection angles may be considered in this framework in a similar way [6], as can the problem of estimating parameters that characterize other aspects of the object, such as size, shape and orientation. It should be noted that the parameters characterizing an

*The research described in this paper was supported by the National Science Foundation under Grant ECS-8012668.

AD-A135 629 TOPICAL MEETING ON SIGNAL RECOVERY AND SYNTHESIS WITH INCOMPLETE INFORMAT..(U) OPTICAL SOCIETY OF AMERICA WASHINGTON D C J W QUINN 31 AUG 83 AFOSR-TR-83-1094 3/3

UNCLASSIFIED AFOSR-83-0026 F/G 20/6 NL





object's location, size, shape and orientation enter the problem nonlinearly, and consequently lead to a nonlinear estimation problem of relatively small dimensionality. This is in contrast to full image reconstruction, in which a linear estimation problem of high dimensionality is solved.

ML Location Estimation

When the cross-sectional density function $f(x)$ consists of the sum of a background field $f_b(x)$ and the object $f_o(x-c_o)$ located at some unknown point c_o in the cross-section, the Radon transform in (1) is

$$g(t,\theta) = \int_{\underline{x}'\underline{\theta}=t} f_b(x)ds + \int_{\underline{x}'\underline{\theta}=t} f_o(x-c_o)ds \triangleq g_b(t,\theta) + g_o(t-\underline{c}_o'\underline{\theta},\theta) \qquad (2)$$

A Shift in the location of the object c_o is seen to result in a shift in the t variable of the Radon transform, the amount of the shift depending cosinusoidally on θ. At any angle θ, $g(t,\theta)$ is a 1D projection of $f(x)$; let the projection measurement at angle θ be given by a noisy observation of the projection $g(t,\theta)$ convolved with a 1D measurement aperture function $h(t)$ (see for example [1])

$$y(t,\theta) = g(t,\theta)*h(t) + w(t,\theta) = g_b(t,\theta)*h(t) + g_o(t-\underline{c}_o'\underline{\theta},\theta)*h(t) + w(t,\theta)$$
$$\triangleq y_b(t,\theta) + g_h(t-c_o'\theta,\theta) + w(t,\theta) \qquad (3)$$

where * denotes 1D convolution in the t variable, and $w(t,\theta)$ is zero-mean white Gaussian noise with spectral level $N_o/2$.

The location of the object $f_o(x)$ is estimated from the noisy projection measurements in (3) by locating the peak of the log likelihood function [7]

$$\ell(c) = \frac{2}{N_o}\int_o^{\pi}\int_{-\infty}^{\infty} y(t,\theta)g_h(t-\underline{c}_o'\underline{\theta},\theta)dtd\theta - \frac{1}{N_o}\int_o^{\pi}\int_{-\infty}^{\infty} g_h^2(t-\underline{c}_o'\underline{\theta},\theta)dtd\theta \qquad (4)$$

The first term is the result of a 2D matched filtering operation in the measurement space (Radon space). Notice that this 2D function is obtained by a convolution back-projection operation [4] on the measurements, with a convolving kernel that is specified in the optimal solution to the object localization problem, and is generally not θ-independent. The second term in (4) is the Radon-space energy in the matched filtering template; this term is c-independent and can be ignored, as can the $2/N_o$ scaling factor. Substituting (3), where without loss of generality we subtract the known quantity $y_b(t,\theta)$ from the measurements

$$L(c) = \int_o^{\pi}\int_{-\infty}^{\infty} y(t,\theta)g_h(t-\underline{c}'\underline{\theta},\theta)dtd\theta = \int_o^{\pi}\int_{-\infty}^{\infty} g_h(t-\underline{c}_o'\theta,\theta)g_h(t-\underline{c}'\underline{\theta},\theta)dtd\theta$$
$$+ \int_o^{\pi}\int_{-\infty}^{\infty} w(t,\theta)g_h(t-\underline{c}'\underline{\theta},\theta)dtd\theta = a(c-c_o) + n(c) \qquad (5)$$

The first term in (5) is a 2D deterministic generalized ambiguity function, which has its peak located at the actual object location $c=c_o$. The second term in (5) is a zero-mean random field with correlation $E[n(c)n(\tilde{c})] = (N_o/2)\cdot a(c-\tilde{c})$.

EXAMPLE

Consider the case where the object is a weighted indicator function on

a disk of radius R

$$f_o(x) = d\chi_{D(R)}(x) \tag{6}$$

with contrast d and Radon-space energy $\varepsilon=(16/3)\pi d^2R^3$. Let the aperture function correspond to spatial bandlimiting of W cycles per unit distance in the t direction, i.e. $h(t)=(\sin 2\pi Wt)/2\pi Wt$. For this circularly symmetric object, the ambiguity function in (5) is $a(c)=\varepsilon a^*(\|c\|/R,RW)$ which is circularly-symmetric and depends on the spatial bandwidth W only through the radius-bandwidth product RW (RW is the number of wavelengths at frequency W in the distance R). Figure 1 illustrates a central section of $a^*(\|c\|/R,RW)$ for several values of RW.

Performance Analysis

Suppose that the object is known to be located inside D(T), a disk of radius T centered at the origin. Let $\underline{e} \triangleq \hat{c}_{ML}-c_o$, where $\hat{c}_{ML}$ is the location of the peak of the log likelihood function L(c) in (5). The ML estimate error covariance $\Lambda_e \triangleq E[\underline{e}\underline{e}']$ may be approximately characterized by dividing D(T) into M square cells, the jth denoted C_j with centroid c_j (the actual object is located at the centroid of C_0 for simplicity).

$$\Lambda e = \sum_{j=0}^{M-1} \left[\int\int_{C_j} \underline{e}\underline{e}'\, p_{e|E_j}(e|E_j)de\right] \text{Prob}[E_j] \triangleq \sum_{j=0}^{M-1} \Lambda_j p_j \tag{7}$$

where E_j is the event that $\hat{c}_{ML} \in C_j$. By approximating p_j as $\text{Prob}[L(c_j) \geq L(c_i), \forall i]$, approximating Λ_j as $(\underline{c}_j-\underline{c}_o)(\underline{c}_j-\underline{c}_o)'$ for $j\neq 0$, and using the Cramer- Rao bound to characterize Λ_0 (the Cramer-Rao bound may be written in terms of the second partial derivative of a(c) evaluated at c=0, which may be calculated using frequency-domain techniques), one obtains an approximate expression for the estimate error covariance. For the disk object in the example, the approximate error covariance is $\sigma^2 I$; Figure 2 is a plot of T^2/σ^2 versus R for several values of the ratio of contrast squared to noise level.

Conclusions

As indicated by Figure 2, the problem of locating an object in a cross-section from noisy projection measurements is characterized by a definite threshold behavior -- for given values of contrast and noise level, there exists a smallest object size that can be located reliably. This type of error covariance analysis may be performed for other object profiles $f_o(x)$, as well as for limited-view and discrete-view projection measurements [6].

REFERENCES

[1] Bracewell, R. 1956. Two-dimensional aerial smoothing in radio astronomy. Austr. J. Phys. 9:297-314.

[2] Cornuelle, B. 1982. Acoustic Tomography. IEEE Trans. Geo. and Rem. Sensing GE-20:326,332.

[3] Klyuev, V., E. Vainberg, I. Kazek. V. Kurozaev. 1980. Computational tomography - a new radiation method of nondestructive testing, I, II. Soviet J. Nondestr. Test. 16:180-185, 186-193.

[4] Mersereau, R., A. Oppenheim. 1974. Digital reconstruction of multi-dimensional signals from their projections. Proc. IEEE 62:1319-1338.

[5] Vest, C., P. Radulovic. 1977. Measurement of three-dimensional temperature fields by holographic interferometry. In Proc. of the Intl. Conf. on Appl. of Holography and Optical Processing. Ed. E. Marom, A. Friesem, E. Wiener-Avnear. Pergamon Press, New York.

[6] Rossi, D. 1982. Reconstruction from projections based on detection

and estimation of objects. Ph.D thesis, MIT, Dept. of EECS, 341 p.

[7] Van Trees, H. 1968. Detection, estimation, and modulation theory, Part I. John Wiley and Sons, New York.

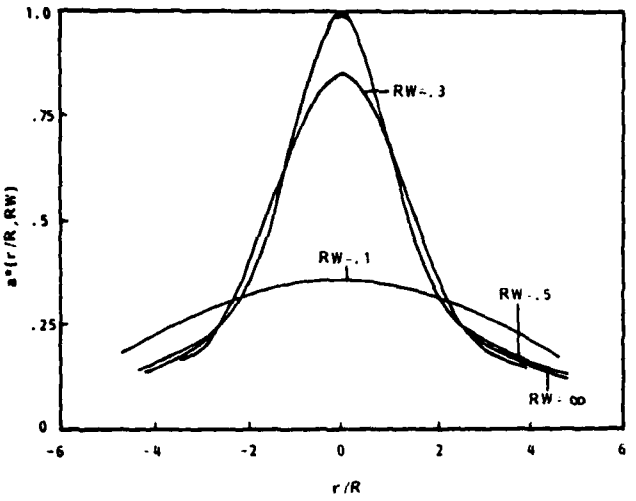


Figure 1

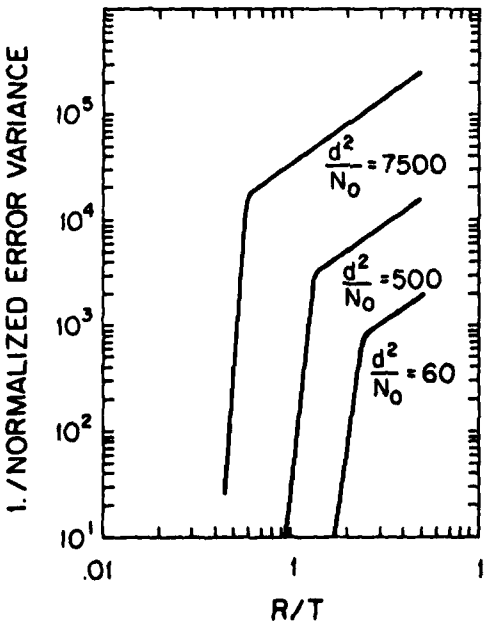


Figure 2

TOMOGRAPHIC IMAGING WITH LIMITED VIEW ANGLE USING AN EXPANSION ON A SET OF EIGENFUNCTIONS ADAPTED TO SPACE-LIMITED OBJECTS

Line GARNERO *, Jean BRUNOL **.

* Institut d'Optique, Université de Paris XI, 91406 ORSAY cedex, France.

** C.G.R., Médecine Nucléaire, rue de la Minière 78530, BUC, France

I - INTRODUCTION

In many radiation imaging applications, such as X-ray computerized tomography or nuclear medicine, tomographic systems have been recently developped, which image an object from an angulary restricted number of projections. It is well known that in such a case, some Fourier components are missing {1}, and thus analytical reconstructions, such as Fourier synthesis can no longer be used. But, the algebraic methods allow to take advantage of "a priori" informations on the object, in order to recover the missing data. Many authors have proposed different reconstruction techniques, that combine limited projection data with a priori object information through iterative revisions in both image and transform spaces {2-3}. These methods in general consume much computer time.

We propose here a new approach of the problem, based on the fact that generally the imaged organs or objects do not fill the whole detector field. The knowledge of their boundaries allows to find, for each class of objects of the same extent, a new set of expansion functions, which are more appropriate than a Fourier expansion. We shall see how the expansion of the object distribution on this new set of functions allows the object reconstruction from projection data by a mere operation of matrix multiplication. As a matter of fact, the terms of the resulting linear system of equations are the same for each object of same extent, and so the corresponding matrix inversion has to be performed only once.

In this summary, we present first the mathematical principles, and we describe the numerical realization of the method, and a reconstruction simulation of an object viewed through a limited angle.

II - MATHEMATICAL PRINCIPLE

2.1. The basis function

Let us take the problem in one dimension first, we will generalize later for the 2.D case.

We consider a detector field $x \in \{-b,+b\}$, and a number of samples equal to N, corresponding to the finite detector

.../...

resolution $(\frac{2b}{N})$. Thus a general object distribution f(x) will be reduced to the N terms series $f(x_p)$, where :

$$x_p = \frac{2pb}{N} \quad , \quad p \in \left[-\frac{N}{2}+1, \frac{N}{2}\right]$$

In the following, all the mathematical development will concern digitized rather than continuous functions.
The N terms Fourier series of $f(x_p)$ can be expressed as :

$$f(x_p) = \sum_{k=-N/2+1}^{N/2} \tilde{f}_k \cdot \exp(-2i\pi \frac{kx_p}{2b}) = \sum_{k=-N/2+1}^{N/2} \tilde{f}_k \cdot \exp(-2i\pi \frac{kp}{N}) \qquad (1)$$

Now, we consider the class $\mathcal{L}_a$ of 1-D digitized functions $f(x_p)$ equal to zero outside the interval {-a,+a} , with a < b, i.e^p:

$$f(x_p) = \text{rect}(\frac{x_p}{2a}) \cdot f(x_p) \qquad (2)$$

where rect $(\frac{x_p}{2a}) = \begin{cases} 1 & |x_p| < a < b \\ 0 & \text{otherwise} \end{cases}$

The Fourier series of a series $f(x_p)$ $\mathcal{L}_a$ can be expressed as follows {4},|cf(1) and (2)|:

$$f(x_p) = \sum_{\ell=-N/2+1}^{N/2} \tilde{f}_\ell \cdot \chi_\ell(x_p) \qquad (3)$$

with $\chi_\ell(x_p) = \text{rect}(\frac{x_p}{2a}) \cdot \exp(-\frac{2i\pi x_p \ell}{2b})$.

From (2) it can be easily seen that the expansion coefficients of equations (1) and (3) are the same.
We can expand these functions $\chi_p(x_p)$ into the Fourier terms such as :

$$\chi_\ell(x_p) = \sum_k \tilde{\chi}_{\ell k} \exp(-2i\pi \frac{kx_p}{2b}) \qquad (4)$$

and equation (3) now becomes :

$$f(x_p) = \sum_\ell \sum_k \tilde{f}_\ell \tilde{\chi}_{\ell k} \exp(-2i\pi \frac{kx_p}{2b}) \qquad (5)$$

Comparing (1) and (5) leads to the relations :

.../...

$$\tilde{f}_k = \sum_{\ell=-N/2+1}^{N/2} \tilde{f}_\ell \tilde{\chi}_{\ell k} \quad , \; k \in [-\tfrac{N}{2}+1, \tfrac{N}{2}],$$

that can be expressed in a matrix form :

$$[\tilde{f}] = [\tilde{\chi}][\tilde{f}], \tag{6}$$

where $[\tilde{f}]$ is a N x 1 column vector of the Fourier components of $f(x_p)$ and $[\tilde{\chi}]$ is a N x N matrix, corresponding to the $[-a,+a]$ truncation operator. Equation (6) shows that, if $f(x_p) \in \mathcal{L}_a$, then the N Fourier terms vector $[\tilde{f}]$ must be an eigenvector of matrix $[\tilde{\chi}]$ associated with the eigenvalue $\lambda = 1$. The diagonalization of $[\tilde{\chi}]$ leads to only two degenerated eigenvalues : 0 or 1, and the number of eigenvectors associated to $\lambda = 1$ is N_a, with $N_a = N \times a/b < N$.

To conclude, the functions $f(x_p) \in \mathcal{L}_a$ can be expanded into a basis of N_a eigenfunctions (we shall note them $f_K(x_p)$, $K = 1, N_a$) which corresponds to the eigenvectors of $[\tilde{\chi}]$. associated to the eigenvalue $\lambda = 1$.

Now, we generalize to the 2-D case : for the reconstruction, it is useful to deal with the class $\mathcal{L}_a{}^2$ of digitized functions $f(x_p, y_p)$ equal to zero outside the disk of radius a. Of course, the basis eigenfunctions could be calculated in the same way, but the new matrix $[\tilde{\chi}]$ will be $N^2 \times N^2$, and much computation time and storage required. So we prefer to deduce the 2-D eigenfunctions from the 1-D ones $\{f_K(x_p)\}$.
Now, we can easily see that the set of the $N_a \times N_a$ functions $\{f_K(x_p), f_{K'}(y_p)\}$ where $\{K, K' \in [1, N_a]\}$ is a complete set of basis functions, for any digitized function $f(x_p, y_p)$ which is nonzero only inside the square region $\{-a < x_p \leq +a, -a \leq y_p \leq +a\}$, and a fortiori for the functions belonging to $\mathcal{L}_a{}^2$.
So, if we consider a 2-D object distribution $O(x,y)$ whose extent is included in the circle of radius a (with $a < b$, b radius of the detector field), its corresponding discrete form $O(x_p, y_p)$ can be expanded into this set of $N_a \times N_a$ 2-D eigenfunctions :

$$O(x_p, y_p) = \sum_{K=1}^{N_a} \sum_{K'=1}^{N_a} \tilde{O}_{KK'} f_K(x_p) f_{K'}(y_p) \tag{7}$$

Let us now consider the problem of the reconstruction of this distribution from an incomplete set of projections.

2.2. The reconstruction problem

The available data are the projections $P_\theta(x_{rp})$ in various directions of the object. Since $O(x_p, y_p)$ belongs to $\mathcal{L}_a{}^2$, its projections $P_\theta(x_{rp})$ belongs to $\mathcal{L}_a$, and thus can be expanded to the set of 1-D eigenfunctions, i.e.

$$P_\theta(x_{rp}) = \sum_{L=1}^{N_a} P_{\theta L} f_L(x_{rp}) \tag{8}$$

Now, "projecting" the two terms of (7) we get :

$$P_\theta(x_{rp}) = \sum_{K=1}^{N_a} \sum_{K'=1}^{N_a} \tilde{O}_{KK'} P_{\theta KK'}(x_{rp}) \tag{9}$$

where $p_{\theta KK'}(x_{rp})$ can be considered as the θ-projections of the eigenfunction $f_K(x_p)f_{K'}(y_p)$, these functions $p_{\theta KK'}(x_{rp})$ do not belong to $\mathcal{L}_a$, but since $P_\theta(x_{rp}) \in \mathcal{L}_a$, the two terms of (9) can be truncated, thus :

$$P_\theta(x_{rp}) = \text{rect}\left(\frac{x_{rp}}{2a}\right) . \; P_\theta(x_{rp}) = \sum_{K=1}^{N_a} \sum_{K'=1}^{N_a} \tilde{O}_{KK'} \; p^{+}_{\theta KK'}(x_{rp}) \qquad (10)$$

where $p^{+}_{\theta KK'}(x_{rp})$ is the truncated projection of the eigenfunctions, and thus can be expanded as any function belonging to $\mathcal{L}_a$. We get then :

$$p^{+}_{\theta KK'}(x_{rp}) = \sum_{L} \tilde{p}_{\theta KK'L} \; f_L(x_{rp}) \qquad (11)$$

Combining (10) and (11) and identifying the component of $f_L(x_r)$ in (8) and (10), we get :

$$\tilde{P}_{\theta L} = \sum_{K} \sum_{K'} \tilde{O}_{KK'} \; \tilde{p}_{\theta KK'L} \qquad (12)$$

We obtain a linear system, whose unknowns are the $N_a \times N_a$ components of the object (cf.(7)). The given data are the $M \times N_a$ values of $P_{\theta L}$, with M the number of observed directions θ . To resolve (12), M must be equal to N_a. The components $\tilde{O}_{KK'}$ are obtained by the inversion of the matrix $[\tilde{p}_{\theta KK'L}]$, which is the same for all objects belonging to a same class $\mathcal{L}_a{}^2$, and imaged through the same projection directions.

As an illustration of the above considerations, we shall present numerical reconstructions of simulated objects whose support is limited to a quarter of the detector area and viewed through a total angle of 56°, resulting 32 x 32 images will be shown.

III - CONCLUSION

As a matter of fact, this method uses the relations existing between the spatial frequencies of every object, which has a given limited support ; thanks to these relations, all the long and big calculations need to be done only once, and the reconstruction itself of an image from the projection data is rapid.

At this time, we have only done simulations with ideal data. We have now to study the influence of noisy data on the reconstruction accuracy, so as to test the validity of the following intuitive argument : a limited number $M \times N_a$ of data is used to reconstruct an object of size $N_a \times N_a$; since M has to be at least equal to N_a, any extrapolation is avoided, and the sensitivity of the method to noise should be limited.

IV - REFERENCES

{1} J. BRUNOL, L. GARNERO, J.C. SAGET, Opt. Comm., 35(1980)311.
{2} K.C. TAM, V. PEREZ MENDES, J. Opt. Soc. Am., 71(1981)582.
{3} T. SATO, S.J. NORTON, M. LINZER, O. IKEDA, M. HIRAMA, Appl. Opt. 20(1981)395.
{4} A. JERRY, Proc. of IEEE, 65(1977)1565.

FA5-1

Linear Estimation with a Size Constraint

M. J. Lahart
Naval Research Laboratory
Washington, D. C. 20375

Least squares estimation has been used in image processing since Helstrom showed several years ago that Wiener filters could be used to deblur images[1]. In its most usual application, the technique uses the image data and the object and noise autocorrelation functions to compute the object that corresponds to the minimum of the sum of the squares of the noise values. We show here how least squares techniques can also be used to estimate spectral components when the size and shape of an object are known. Examples of missing components are high frequency Fourier components of an object that has been subjected to low pass filtering (blurring) and transforms of missing projections of an object that is to be restored through computed tomography.

The least squares of estimate of an unknown quantity y is the linear sum $\hat{y}$ of measured quantities $X = (x_1, \ldots x_i, \ldots)$

$$\hat{y} = M_{UK}(M_{KK})^{-1} X \tag{1}$$

that minimizes the expectation of the square of the actual and estimated values.[2] The matrix M_{KK} is an array of correlation coefficients $\langle x_i x_j \rangle$ between the measured quantities x_i and x_j, and M_{UK} is an array of correlation coefficients $\langle x_i x_j \rangle$ between measured quantities x_i and unknown quantities y_j. Functional relationships among known and unknown quantities, as well as the stationarity characteristics of the object, are used to compute these correlation coefficients. For example, the imaging equation, which describes the image as the convolution of the object and a point spread function, is the basis of the Wiener filter. In that application the known quantities are the measured values of the image and the unknown quantities are the values of the object. Stationarity characteristics relate the product of two quantities to a Wiener spectrum or a correlation function. In the Fourier domain, stationarity implies that spectral components are delta-correlated[3], i.e.,

$$\langle o_o^*(\omega_i)\, o_o(\omega_j) \rangle = 2\pi O_o(\omega_i)\, \delta(\omega_i - \omega_j) \tag{2}$$

and similarly for noise. Here, $O_o(\omega)$ is the Wiener spectrum of the object, proportional to the Fourier transform of the autocorrelation function, and $\delta(\omega_i - \omega_j)$ is a Dirac delta function.

SPATIAL LIMITATION

An object whose spatial extent is known may be represented as the multiplication of a stationary process by a truncating function that is zero outside the known boundaries of the object. In the Fourier domain this is a convolution:

$$o(\vec{\omega}_i) = \frac{1}{2\pi} \int t(\vec{\omega}) o_o(\vec{\omega}_i - \vec{\omega}) d\vec{\omega} + n(\vec{\omega}). \quad (3)$$

The correlation coefficients between measured spectral components $o(\vec{\omega}_i)$ and $o(\vec{\omega}_j)$ are found by multiplying variations of Eq. 3, taking expectations, and using Eq. 2 to express spectral components $o(\omega_i-\omega)$ and $o(\omega_j-\omega)$ in terms of a Wiener spectrum.[4] This leads to

$$\langle o^*(\vec{\omega}_i) o(\vec{\omega}_j) \rangle = \frac{1}{2\pi} \int t^*(\vec{\omega})\, t\,(\vec{\omega}_i + \vec{\omega}_j - \vec{\omega}) O_o\,(\vec{\omega}_i - \vec{\omega})\, d\vec{\omega} + N(\vec{\omega}), \quad (4)$$

where $O_o(\vec{\omega})$ is the Wiener spectrum of the untruncated object. The expression for the correlation coefficient between measured and unknown components are the same, except that the noise term $N(\omega)$ is not present. In the one-dimensional restoration problem $t(x)$ is often a rectangle function and its Fourier transform $t(\omega)$ is a sinc function.

It is convenient to express Eq. 3 in circular coordinates when estimating spectral components in computed tomography. Here, projections are made through the object at several angles and Fourier transforms $o(\omega,\phi)$ of each of these projections are calculated. We let $o_o(\nu,\alpha)$ represent the Fourier transform of the untruncated object at frequency $(-\nu \sin\alpha,\ \nu \cos\alpha)$, (See Fig. 1 for illustration of the coordinate system). If the object is limited in extent by a circle of radius R, the function $o(\omega,\phi)$ is[3]

$$o(\omega,\phi) = \frac{R^2}{2\pi} \int_0^{\infty} \int_0^{2\pi} o_o(\nu,\alpha) \Lambda\{R[\omega^2 - 2\omega\,\nu \cos(\alpha-\nu) + \nu^2]^{1/2}\}\, \nu\, d\nu\, d\alpha, \quad (5)$$

where $\Lambda(\ )$ is a besinc function. This is the form of Eq. 3 that is used to compute correlation coefficients between measured spectral components in the projections. Products of $o(\omega,\phi)$ at different spatial frequencies and angles, which comprise the matrix M_{KK}, can be expressed in terms of the object Wiener spectrum with the aid of Eq. 2, and similar coefficients are calculated for M_{KU}. With the measured $o(\omega,\phi)$ as the data, the $o(\omega,\phi)$ may be estimated for missing views and used to provide better reconstructions than could be made by reconstructing from the data alone.

We should note a similarity between Wiener spectrum estimation and estimation using knowledge of spatial extent. It exists in the sense that Eq. 3 - the equation on which the computation of correlation coefficients is based - is a convolution, as is the imaging equation, which is the basis of the Wiener filter. The two problems are not duals, however, because the stationarity condition of Eq. 2 is applied in the Fourier domain in both cases.

Unlike the Wiener spectrum problem, the matrix M_{KK} that is computed in the spatial limitation problem is not easily diagonalized, and must be inverted laborously. This is practical only if it has a small number of elements. For most problems the convolution kernal $t(\omega)$ is large when ω is small, and is relatively small when ω is larger, so that only a few correlation coefficients are significantly different from zero. We have

estimated each spectral component separately, using only that data which strongly influence it, and have constructed M_{KK} only from coefficients that relate this data. These separate computations for each estimation have allowed the size of M_{KK} to be manageable.

COMPUTATIONS

Figure 2 is an example of a one-dimensional restoration based on bandwidth extrapolation by least squares methods. Part (B) is a low pass filtered version of part (A) and part (C) is a reconstruction which uses estimated spectral components.

Figure 3 is a tomographic reconstruction of an object from 15 equally spaced parallel beam projections, using the filtered back projection method. The original object consists of a circle of radius 32 pixels inside of which were smaller holes of radii 6,4,3,2 and 1. The object is binary, that is, the large circle is white and the holes and surrounding areas is black. The entire frame is 128 x 128 pixels.

Figure 4 is a similar reconstruction from 60 equally spaced projections that have been estimated from 15 data projections. The object has significantly higher contrast and fewer artifacts.

In order to use Eq. 2, we must have an estimate of $O_o(\omega)$, the Wiener spectrum of the untruncated object. Because this quantity appears as a part of an integrand, it need not be exact, but we have found empirically that some improbable forms of $O_o(\omega)$ produce poor restorations. For the tomographic reconstruction we took $O_o(\omega)$ as the average of the sum of the real and imaginary parts of the spectral components of all measured views,

$$O_o(\omega) = \frac{1}{n} \sum_{\text{all } \phi} [o_{r2}(\nu,\phi) + o_{i2}(\nu,\phi)] \quad . \tag{6}$$

The reconstructions presented here use noiseless data. Restorations have also been made with noise whose standard derivation is a few percent of that of the signal. There is some loss of image quality in the reconstructions, but the imagery is still significantly better than the unrestored imagery. Examples of one- and two-dimensional restorations with noise are in the references 3,4.

REFERENCES

1. C. W. Helstrom, "Image restoration by the method of least squares", J. Opt. Soc. Am. 57, 297-303 (1967).

2. M. J. Lahart, "Local image restoration by a least-squares method", J. Opt. Soc. Am. 69, 1333-1339 (1979).

3. M. J. Lahart, "Estimation of reconstructions in computed tomography", J. Opt. Soc. Am. 71, 1155-1161 (1981).

4. M. J. Lahart, "Maximum-likelihood restoration of non-stationary imagery", J. Opt. Soc. Am., 64, 17-22 (1974).

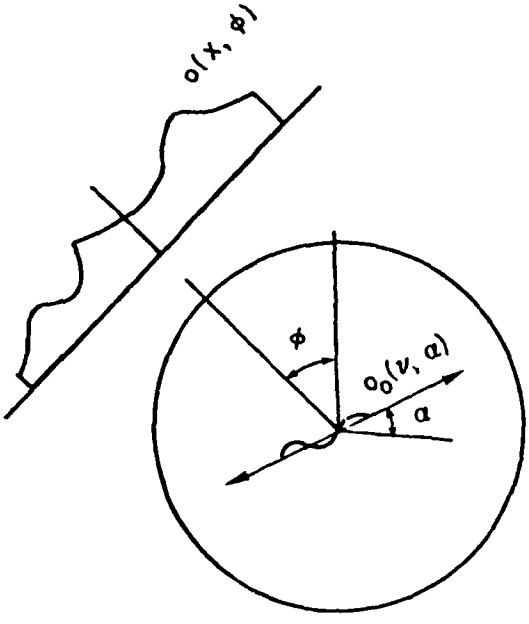


Figure 1

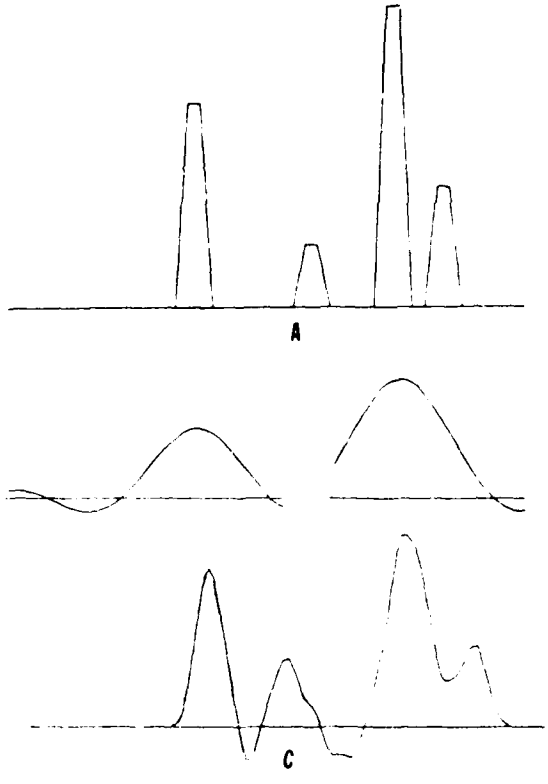


Figure 2

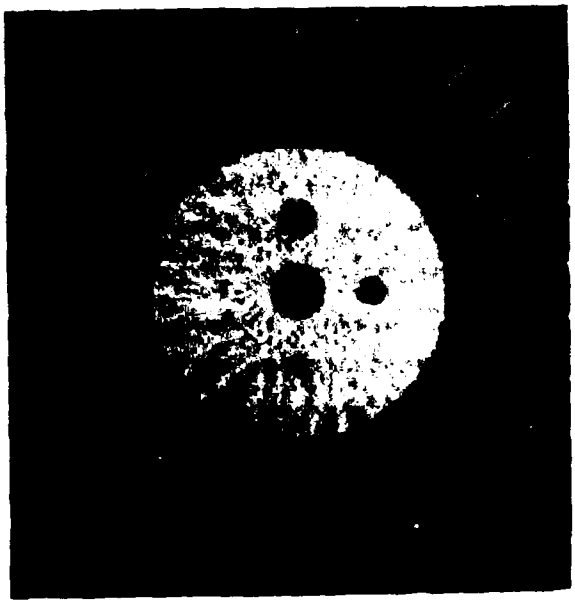


Figure 3

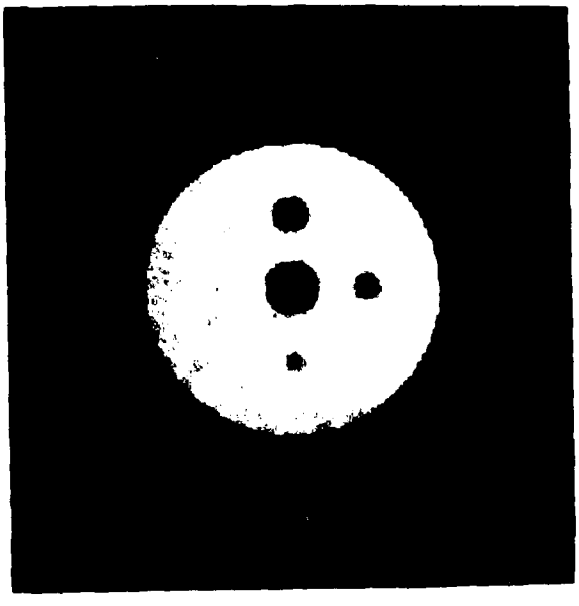


Figure 4

Bayesian Approach to Limited-Angle CT Reconstruction*

by

Kenneth M. Hanson
University of California
Los Alamos National Laboratory
Los Alamos, NM 87545

George W. Wecksung
University of California
Los Alamos National Laboratory
Los Alamos, NM 87545

Consider the function f(x,y) to belong to the set of all integrable functions with compact support. The projections of f(x,y) may generally be written as

$$g_i = \int\int h_i(x,y) f(x,y) dxdy \qquad ,i=1,\ldots,N \tag{1}$$

where the h_i are strip-like response functions corresponding to each of the N available projection measurements. The objective of computed tomography (CT) is to reconstruct the source function f(x,y) from these N measurements. Clearly a limited number of such measurements cannot completely specify an arbitrary f(x,y). Since Eq. 1 may be viewed as an inner product between h_i and f in the Hilbert space of all acceptable functions, each measurement consists of a projection of the unknown vector f onto the basis vector h_i. The available measurements can only provide information about those components of f that lie in the subspace spanned by the response functions called the measurement subspace. The components of f that lie in the orthogonal (null) subspace do not contribute to the measurements and, hence, cannot be

*This work was supported by the U. S. Department of Energy under Contract #W-7405-ENG-36.

determined from the measurements alone. Without prior information about $f(x,y)$ it is at least necessary to restrict the solution to the measurement space in order to make it unique, i.e., have minimum norm. The null-space components of such a solution are obviously zero. It is known that this leads to identifiable, objectionable artifacts when the projections span a limited range of angles.[1,2] In its generality, Eq. 1 is representative of any discretely sampled, linear-imaging process. Thus, the above statements and the approach that follows are applicable to many other problems such as restoration of blurred images and coded-aperture imaging.

The Bayesian approach to CT reconstruction[3] is based on the assumption that the image to be reconstructed belongs to an identifiable ensemble of similar images. The best estimate for the reconstruction is taken to be that particular image f that maximizes the a posteriori conditional probability density of f given the measurements g. This probability is given by Bayes' law

$$P(f|g) = \frac{P(g|f)P(f)}{P(g)} \tag{2}$$

in terms of the conditional probability of g given f and the a priori probability distributions of f and g separately. We assume that the measurement noise is additive with a probability distribution that has zero mean and is gaussian distributed. P(f) is also assumed to be gaussian distributed about a mean value $\bar{f}$. The covariance matrices of the noise and ensemble image-vectors are R_n and R_f, respectively. Under these assumptions, the maximum a posteriori (MAP) solution is easily shown to satisfy

$$R_f^{-1}(\bar{f}-f) + H^T R_n^{-1}(g-Hf) = 0 \tag{3}$$

where H is the linear operator (matrix) corresponding to the projection process described by the integral in Eq. 1. The transpose of H is the familiar backprojection operation. It can be seen from Eq. 1 that the desired solution strikes a balance between its difference with the ensemble mean $\bar{f}$ and the solution to the measurement equation ($g = Hf$). This balance is determined by the covariance matrices R_f and R_n that specify the confidence with

which each difference is weighted as well as possible correlations between the differences.

We have adopted an iterative approach to the solution of Eq. 3 based on the scheme proposed by Herman and Lent.[3] The nth estimate f^n is given by the iteration scheme:

$$f^0 = \bar{f}, \tag{4a}$$

$$f^{n+1} = f^n + c^n r^n, \tag{4b}$$

$$r^n = \bar{f} - f + R_f H^T R_n^{-1}(g - Hf^n), \tag{4c}$$

$$c^n = \frac{r^{nT} s^n}{s^{nT} s^n}, \tag{4d}$$

$$s^n = (I + R_f H^T R_n^{-1} H) r^n, \tag{4e}$$

where vector r^n is the residual of Eq. 3 (multiplied by R_f) and the scaler c^n is chosen to minimize the norm of r^{n+1}. This iterative scheme is very similar to the one proposed by Hunt[4] for nonlinear MAP-image restoration. We have found that this technique works well although convergence typically requires 10 to 20 iterations. It is easy to see from the form of this iterative procedure that significant null-space contributions to f can arise from the a priori information. First, the zero-order estimate is $\bar{f}$, which can contain null-space contributions. Second, in Eq. 4c, R_f can generate null-space contributions when it operates on the result of the backprojection (H^T) process, which lies wholly in the measurement space.

The above MAP-reconstruction technique has been applied to an example consisting of a fuzzy annulus with variable amplitude (see Ref. 2), which roughly emulates the nuclear-isotope distribution in the cross section of a heart. The available projection data are eleven views covering 90° in projection angle. Each projection contained 128 samples. The measurement-space reconstruction obtained using ART[5] shows severe artifacts that tend to obscure much of the source distribution. In the MAP approach it was assumed that $\bar{f}$ was an annulus with constant amplitude and that the covariance matrix R_f was large (1.0) at the peak of the annulus and small (0.2) inside and

outside. The measurement noise was assumed to be uncorrelated, constant, and low in value. The resulting MAP reconstruction is vastly superior to the ART result showing much more of the detailed variation in amplitude present in the source.

In past comparisons of MAP results to more standard techniques in the areas of CT[3] and image restoration, [5,6-8] the MAP approaches yielded little or no benefits. The reasons for the success of the MAP approach in the above limited-angle CT problem are 1) the solution is severely underdetermined because of the limited data set, and 2) the a priori assumptions about $\bar{f}$ and r_f can be made quite restrictive. It is expected that the MAP analysis will be most useful in situations where these two conditions hold.

REFERENCES

1. K. M. Hanson, "CT Reconstruction from Limited Projection Angles," to be published in Proc. SPIE (Vol. 347) Photo-Optical Instrumentation in Medicine X, New Orleans, 1982.

2. K. M. Hanson, "Limited-Angle CT Reconstruction Using A Priori Information," to be published in Proc. First IEEE Computer Society International Symposium on Medical Imaging and Image Interpretation, Berlin, 1982.

3. G. T. Herman and A. Lent, "A Computer Implementation of a Bayesian Analysis of Image Reconstruction," Info. Control 31, 364-384 (1976).

4. B. R. Hunt, "Bayesian Methods in Nonlinear Digital Image Restoration," IEEE Trans. Comp. C-26, 219-229 (1977).

5. R. Gordon, R. Bender, G. T. Herman, "Algebraic Reconstruction Techniques for Three-Dimensional Electron Microcopy and X-Ray Photography," J. Theor. Biol. 29, 471-481 (1970)

6. H. J. Trussell and B. R. Hunt, "Improved Methods of Maximum A Posteriori Restoration," IEEE Trans. Comp. C-27, 57-62 (1979).

7. H. J. Trussell, "Notes on Linear Image Restoration by Maximizing the A Posteriori Probability," IEEE Trans. Comp. C-27, 57-62 (1978).

8. T. M. Cannon, H. J. Trussell and B. R. Hunt, "Comparison of Image Restoration Methods," Appl. Opt. 17, 3384-3390 (1978).

SESSION X

INTERPOLATION AND TOMOGRAPHY II

J. C. Dainty, *Presider*

The Use of *A Priori* Information In Image Reconstruction From Limited Data

B.P. Medoff, W.R. Brody and A. Macovski

Stanford University
Stanford, CA 94305

Introduction

Image reconstruction algorithms implemented on existing CT scanners require the collection of line integrals that are evenly spaced over 360 degrees.[1] In many practical situations, requirements for high temporal resolution or the presence of an x-ray opaque structure prevent the measurement of all the line integrals. Attempts to use existing algorithms in this "limited data" situation result in images with severe streak artifacts.[2]

Recently, there has been interest in an iterative approach to image reconstruction from limited data.[3,4] The algorithm employs repeated transformation between the space and frequency domains, and is in essence a two dimensional extension of a frequency domain extrapolation algorithm introduced by Gerchberg[5] and Papoulis.[6] Because of its reliance on a frequency domain interpretation, this algorithm is limited to a particular type of CT data collection geometry (parallel beam) and a particular pattern of missing data (missing complete projections). While the importance of using *a priori* information was recognized, results have only been shown using the *a priori* knowledge of finite extent and positivity.

In this paper, an operator framework for limited data image reconstruction that applies to arbitrary scanning geometries, and to arbitrary patterns of missing line integral data, is introduced. This framework is not based on a frequency domain interpretation, and hence can be used when the pattern of missing data does not correspond to a missing sector in the frequency domain, as happens, for example, in the *Bagel Problem.*[7] The operator framework allows the reconstruction algorithm to incorporate a wide class of *a priori* information about both the line integral data and the underlying density.

This paper shows how a certain kind of *a priori* information - knowledge of the density in part of the object - can be exploited to improve the quality of the reconstructed image. The technique is demonstrated with images reconstructed from real x-ray data obtained from a clinical CT scanner.

This work was supported by the National Heart, Lung and Blood Institute (contract 1R01-HL25905) and by a grant in aid from the American Heart Association. Dr Brody is an Established Investigator of the American Heart Association.

Limited Data Reconstruction

Let the full set of line integral data z be divided into two subsets: the measured line integrals y and the missing line integrals x. Let f be an image representing the attenuation coefficients of the object in the plane through which the measurements are made. The set of line integrals z will be referred to as the shadow of f. Our goal is to produce an estimate $\hat{f}$ that approximates f. The algorithms to be developed are based on two fundamental observations.

(1) *Any technique that generates an image must implicitly assign values to the missing line integrals.* Instead of attempting to compute the estimate $\hat{f}$ directly from the limited set of data, we can attempt to estimate the missing line integral values that would be predicted by this image estimate. Then the estimated line integrals can be combined with the measured data, and the resulting full set of data can be reconstructed using convolution back projection.

(2) *Missing measurements correspond to missing information.* Two objects that are distinguishable given a full set of data might be indistinguishable given a limited set of data. A unique image is only determined if the information not contained in the measurements is available in the form of *a priori* knowledge about the object or line integral data. The *a priori* information can be represented as constraints that must be satisfied by the reconstructed image and line integral data. Examples of constraints are: The image and line integrals are nonnegative; the image is confined within a known boundary; part of the image has a known value; all of the values of the image lie in some known range.

Define z to be a ***consistent full set of projection data*** if and only if: (1) z satisfies the line integral constraints; (2) z can be reconstructed to produce an image that satisfies the image constraints; and (3) the line integrals calculated from this image equal the values in z. Define $\hat{x}$ to be a ***consistent estimate of the missing data*** if and only if it can be combined with the measured data y to form a consistent full set of data.

The limited data image reconstruction problem can be viewed as the problem of finding a consistent estimate of the missing data.

The Consistency Principle

Define a constraint operator $\mathbf{C}^z$ which enforces constraints on a full set of line integral data. For example, an operator that enforces a nonnegativity constraint might set negative line integral values to zero. Define the following: A constraint operator $\mathbf{C}^I$ which enforces constraints on an image; a linear reconstruction operator $\mathbf{R}$ which produces an image from a full set of projection data; and a line integral projection operator $\mathbf{S}$ which produces a shadow from an image. $\mathbf{R}$ can be approximated by the usual convolution back projection algorithm. $\mathbf{S}$ can be approximated by numerical integration of a discrete representation of an image. The full set of projection data, z, can be thought of as a vector partitioned into two component vectors, x and y. This will be represented as $z = \begin{bmatrix} x \\ y \end{bmatrix}$.

With these definitions, we introduce three *Consistency Conditions*. If $\hat{x}$ is a consistent estimate of the missing data, it can be combined with the measured line integrals to form a full set of line integral data which satisfies the line integral constraints,

$$\begin{bmatrix} \hat{x} \\ y \end{bmatrix} = \mathbf{C}^z \begin{bmatrix} \hat{x} \\ y \end{bmatrix} \quad . \tag{1}$$

A consistent estimate can be combined with the measured data and reconstructed to produce an image which satisfies the image constraints,

$$\mathbf{R}\begin{bmatrix} \hat{x} \\ y \end{bmatrix} = \mathbf{C}^I\mathbf{R}\begin{bmatrix} \hat{x} \\ y \end{bmatrix} \quad . \tag{2}$$

The line integrals of the reconstructed image must agree with the measured and estimated line integrals,

$$\mathbf{SR}\begin{bmatrix} \hat{x} \\ y \end{bmatrix} = \begin{bmatrix} \hat{x} \\ y \end{bmatrix} \quad . \tag{3}$$

Based on these relationships we introduce the *Consistency Principle*: There exists a unique reconstruction from a limited set of line integral data if and only if there is a unique estimate of the missing data that satisfies (1)-(3).

Ghosts

Suppose that there are two different estimates of the missing data, $\hat{x}_1$ and $\hat{x}_2$, which both satisfy (1)-(3). Then, from (3) and the fact that $\mathbf{S}$ and $\mathbf{R}$ are both linear operators,

$$\mathbf{SR}\begin{bmatrix} \hat{x}_1 \\ y \end{bmatrix} - \mathbf{SR}\begin{bmatrix} \hat{x}_2 \\ y \end{bmatrix} = \begin{bmatrix} \hat{x}_1 \\ y \end{bmatrix} - \begin{bmatrix} \hat{x}_2 \\ y \end{bmatrix} \tag{4}$$

$$\mathbf{SR}\begin{bmatrix} \hat{x}_1 - \hat{x}_2 \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{x}_1 - \hat{x}_2 \\ 0 \end{bmatrix} \quad . \tag{5}$$

Define $g = \hat{x}_1 - \hat{x}_2$, and $G = \mathbf{R}\begin{bmatrix} g \\ 0 \end{bmatrix}$. Let $\mathbf{S}^y$ represent an operator that computes the shadow only in the measured data region. Then

$$\mathbf{S}G = \begin{bmatrix} g \\ 0 \end{bmatrix} \tag{6}$$

$$\mathbf{S}^y G = 0 \quad . \tag{7}$$

Define a *Ghost Image* to be a nonzero image which has zero line integrals in the regions where data has been measured. Equivalently, a *Ghost Image* is a nonzero image which is in the null space of the operator $\mathbf{S}^y$. When the geometry of the missing data is such that there may be images which are invisible to the measured data, the constraints must provide enough information to determine a unique consistent estimate.

Iterative Algorithms

The consistency conditions (1)-(3) can be combined in a *single* necessary functional relationship that must be satisfied by a consistent estimate of the missing data. One of the many possible combinations is,

$$\begin{bmatrix} \hat{x} \\ y \end{bmatrix} = \mathbf{SR}\begin{bmatrix} \hat{x} \\ y \end{bmatrix} = \mathbf{SC}^I\mathbf{R}\begin{bmatrix} \hat{x} \\ y \end{bmatrix} = \mathbf{SC}^I\mathbf{RC}^z\begin{bmatrix} \hat{x} \\ y \end{bmatrix} \quad . \tag{8}$$

In general, there is no computationally efficient way to solve directly for a consistent estimate $\hat{x}$. A consistent estimate can be determined using iterative techniques, as shown in Figure 1. An estimate of the missing data is combined with the measured data, and the resulting full set of data is forced to satisfy the line integral constraints. The full set of data is reconstructed, and the reconstructed image is forced to satisfy the image constraints. Line integrals of this image are calculated and used to update the estimated line integrals. We have previously presented update techniques based on the methods of successive substitution, steepest descent, and conjugate gradients.[7] Let $\mathbf{S}^z$ represent an operator that computes the shadow only in the missing data region. Then the recursion for the method of successive substitution is

$$\hat{x}_{i+1} = \mathbf{S}^z\mathbf{C}^I\mathbf{RC}^z\begin{bmatrix} \hat{x}_i \\ y \end{bmatrix} \quad . \tag{9}$$

A detailed proof of convergence and analysis of this algorithm is presented elsewhere.[8]

When C^s is the identity, C^t is a known extent constraint, parallel beam geometry is used, and complete projections are missing over a continuous range of angles, (9) is equivalent to the Gerchberg-Papoulis algorithm. Our framework, and the resulting algorithms, are not restricted to this special case.

The Use of *A Priori* Information

The algorithm based on the method of successive substitutions has been implemented on a third generation (fan beam) Varian CT body scanner. This section presents experimental results demonstrating the use of *a priori* information about the density in a region of the underlying object. In the experiment, parts of some of the fan beam projections are missing. The missing data does not correspond to a missing sector in the frequency domain; therefore, a Gerchberg-Papoulis frequency domain extrapolation algorithm can not be directly applied.

Figure 2 is a full data reconstruction of a slice through a phantom that contains within it three smaller rods. Assume that in 10% of the fan beam projections in a continuous angular range, the central 20% of the nonzero line integrals are missing. Figure 3 shows the streak artifact generated when convolution back projection is used to reconstruct the limited set of data.

First, a reconstruction is obtained using the iterative algorithm with positivity and known outer boundary constraints. The resulting steady state image is shown in Figure 4. Although this image is an improvement over the image obtained using convolution back projection directly on the limited set of data, there is still significant streak artifact. The residual error can be determined by taking the difference between the steady state image and the actual density values of the object. The nonzero line integrals of this residual error image are concentrated in the missing data region. This error represents a ghost which can not be eliminated using only the positivity and known extent constraints.

Next, assume that a region in the center of the object has a known uniform value. This information can be represented by a constraint operator which resets the central part of the image to this known value, as shown in Figure 5. The image produced by the iterative algorithm using positivity, known extent, and this additional known background information, is shown in Figure 6. Compared with the steady state result shown in Figure 4, there is substantial reduction in streak artifact. Ghost components that are not consistent with the additional constraint have been eliminated.

This example illustrates three important points. First, it shows the potential of using deterministic *a priori* information to reduce ghosts and improve the quality of the reconstructed image. Second, it shows how the operator framework incorporates this information into the reconstruction process. Finally, it illustrates the robustness of the technique. Note that although the background constraint removes the small structure in the center of the image, this structure is resolved in detail in the steady state reconstruction. Although information about this structure is not retained in the estimated line integrals, it is retained in the measured data and reincorporated at each step of the iteration. The steady state reconstruction is the best compromise between the measured data and all of the available *a priori* information.

Conclusion

A new operator framework for image reconstruction from limited data was introduced. This framework makes no special assumptions about the pattern of missing line integrals or the scanning geometry. Improved reconstruction was obtained by using this operator framework to capture deterministic *a priori* information. Experimental results were presented which demonstrate the performance of the technique.

References

1. G. T. Herman and A. Naparstek, SIAM J. Appl. Math. **33**, 3 (1977) pp. 511-533.
2. G. S. Harell, D. F. Guthaner, R. S. Breiman, C. C. Morehouse, E. J. Seppi, W. H. Marshall, and L. Wexler, Radiology **123** (1977) pp. 515-517.
3. K. C. Tam and V. Perez-Mendez, J. Opt. Soc. Am. **71**, 5 (1981) pp. 582-592.
4. T. Sato, S. J. Norton, M. Linzer, O. Ikeda, and M. Hirama, Applied Optics **20**, 3 (1981) pp. 395-399.
5. R. W. Gerchberg, Optica Acta **21**, 9 (1974) pp. 709-720.
6. A. Papoulis, IEEE Trans. Circuits Syst. **CAS-22**, 9 (1975) pp. 735-742.
7. B. P. Medoff, W. R. Brody, and A. Macovski, in *Proc. International Workshop on Physics and Engineering in Medical Imaging*, (Pacific Grove, CA, 1982).
8. B. P. Medoff, Ph.D. dissertation. Dept. of Electrical Engineering, Stanford Univ., Stanford, CA, 1982.

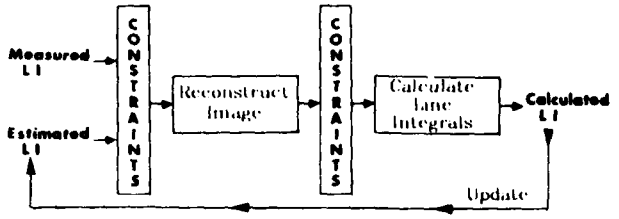


Figure 1: The iterative limited data image reconstruction procedure.

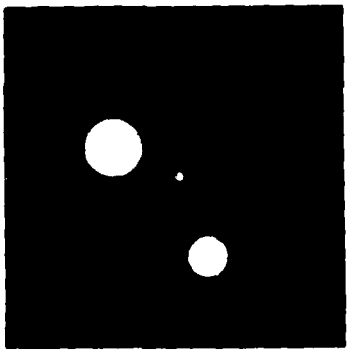


Figure 2: Full data reconstruction of a slice through a phantom that contains within it three smaller rods.

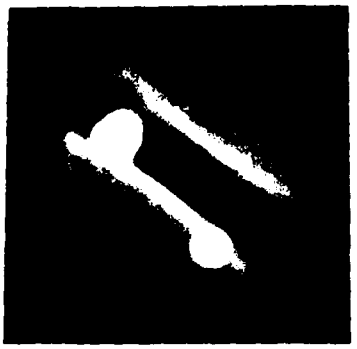


Figure 3: Limited data reconstruction with severe streak artifact produced by convolution back projection. In 10% of the projections, the central 20% of the nonzero data is missing.

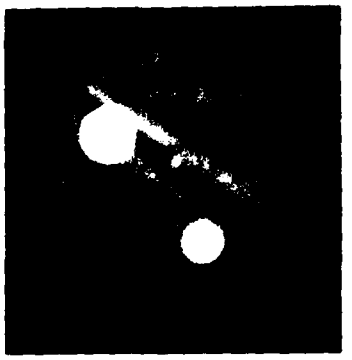


Figure 4: Steady state image produced by the iterative algorithm using known extent and positivity constraints. There is significant residual streak artifact.

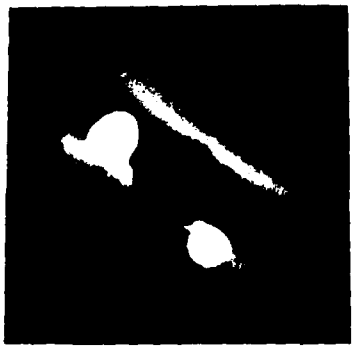


Figure 5: *A priori* knowledge of the background density in a region of the object can be represented as a constraint operator. This operator resets the central region of the image to the known value at each step of the iteration.

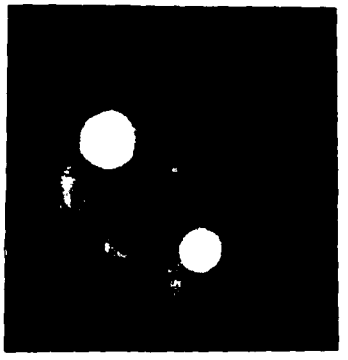


Figure 6: Steady state image produced by the iterative algorithm using the known background constraint in addition to known extent and positivity. Compared with Figure 4, there is significant reduction in the residual streak artifact.

Incorporation of Prior Constraints in Tomographic Reconstructions from Coded Images

Richard G. Paxman, Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Gene R. Gindi, Department of Diagnostic Radiology, School of Medicine, Yale University, 333 Cedar Street, New Haven, CT 06510

Harrison H. Barrett, Optical Sciences Center, University of Arizona, Tucson, Arizona 85721, and Department of Radiology, Arizona Health Sciences Center, University of Arizona, Tucson, Arizona 85724

The study of coded-aperture imaging has usually centered on the restoration of planar objects from planar coded images. The concept has been generalized to the problem of restoring a 3-dimensional object from a planar coded image. These two tasks are illustrated in Figure 1.

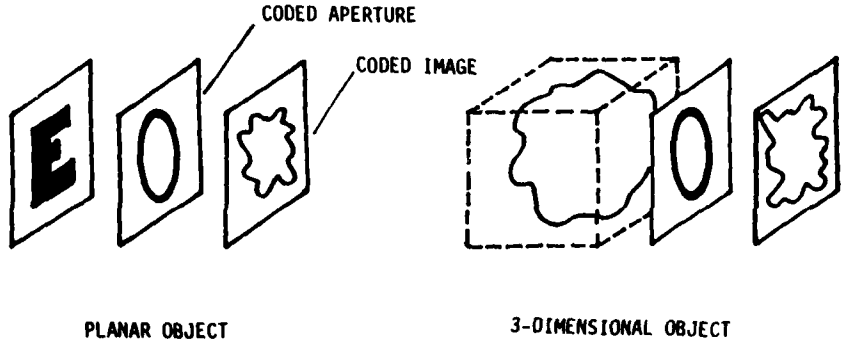


Figure 1

Clearly the second problem is more ambitious since we seek to restore a 3-dimensional function from a 2-dimensional data set. Said in another way, we are trying to solve N^3 unknowns given N^2 equations. The two-dimensional data set has been shown to be deficient in two ways: (1) Because this is a

restricted-view-angle system, "missing cones" of information exist in the 3-dimensional Fourier transform of the object; (2) Because the coded image can be considered to be a series of overlapping pinhole projections of the object, there is a mixing or "multiplexing" of data in the 3-dimensional Fourier transform of the object.

As a first estimate of the object one can perform mismatched scale correlation on the coded image. The output of this process is equivalent to a blurred version of the object in which the blurring psf is shift-variant and anisotropic. This estimate can be processed using the Jacobi method - an iterative technique which in the limit of many iterations becomes inverse filtering. The iterative nature of the restoration provides a convenient framework for the injection of a priori information. Constraints such as positivity and spatial extent of the object can be enforced at each iteration in the hopes of overcoming the ambiguity arising from the deficiencies in the data set.

As a first try at implementing these ideas, we performed a computer simulation in which the dimensionality of the problem was reduced so that the object was 2-dimensional while the coded aperture and coded image were each 1-dimensional. The object was further simplified to consist of 4 discrete layers. A 25 element pinhole code approximating a non-redundent array (NRA) was used. The geometry of this simulation is shown in Figure 2.

The constraints used in the constrained reconstruction include positivity, lateral extent of object, and power per layer, that is knowledge of the integrated source strength in each layer presumably derived from an independent measurement. The unconstrained and constrained reconstructions after 175 iterations are shown in Figures 3 and 4, respectively. The dotted lines

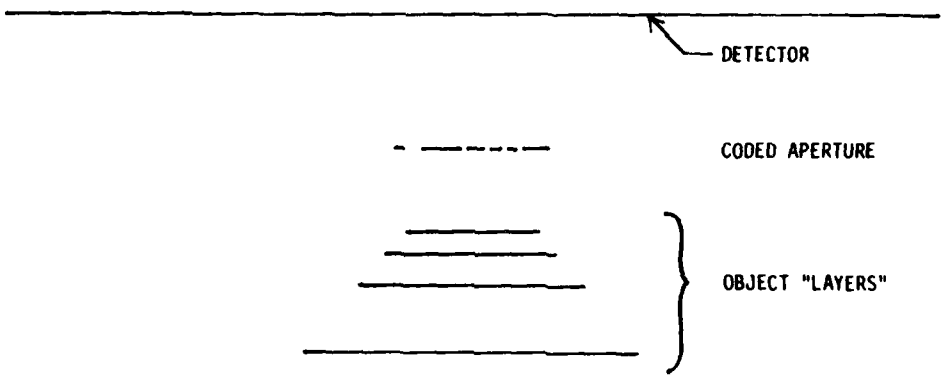


Figure 2

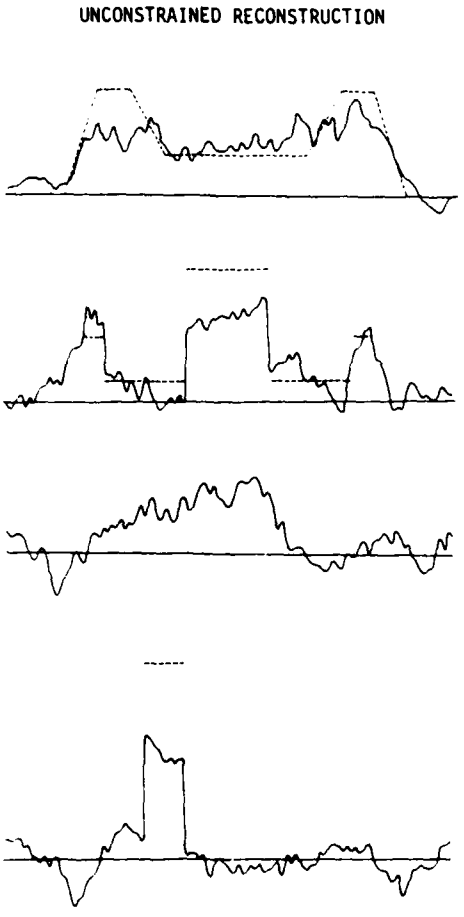


Figure 3

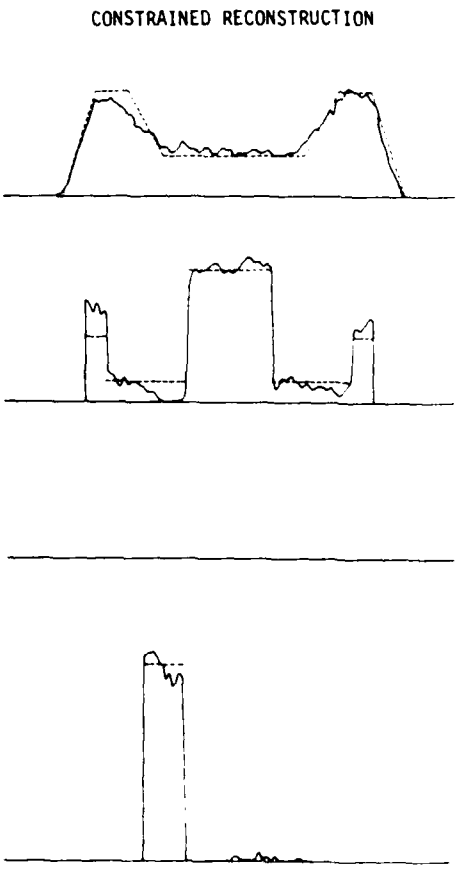


Figure 4

represent the true object. When photon noise was added to the simulation the constrained reconstruction was found to be relatively stable.

Admittedly, the object used in this simulation is far from being realistic. Furthermore, a large amount of prior knowledge about the object is implied by reconstructing on only the 4 given planes. Nevertheless the results are promising because of the dramatic improvement in the reconstruction due to the constraints.

To perform more realistic simulations a continuous object made of superimposed circles of various weights was used so that coded images could be analytically derived. Reconstructions were performed on a finely sampled grid. As might be expected, the results in the continuous case are not as dramatic as those in the discrete layer case. Even though the constraints do in fact improve the mean square error of the reconstruction, thus far the improvement has been rather mild.

Continuing work includes attempts to improve the design of coded apertures for out-of-focus behavior, reconstructions from orthogonal-view coded images, and full 3-dimensional reconstructions from 2-dimensional coded images.

Inverse Scattering Reconstructions From Incomplete Fourier Space Data

N.H. Farhat
University of Pennsylvania
Electro-Optics and Microwave-Optics Laboratory
The Moore School of Electrical Engineering
200 S. 33rd St., Philadelphia, Pa. 19104

ABSTRACT

We show that 3-D tomographic inverse scattering reconstruction of a scattering object is obtainable from data lying on a curved surface, rather than within a volume, of its accessed Fourier space as would ordinarily be required.

I. Introduction

It is known from inverse scattering theory [1]-[4], that multiaspect *monostatic* or *bistatic* coherent measurement of the far field scattered by a plane-wave illuminated conducting or non-dispersive object under conditions that satisfy the *physical optics* and Born approximations can be used to access the 3-D Fourier space $\Gamma(\bar{p})$ of the object *scattering function* $\gamma(\bar{r})$. Here $\bar{r}$ is a 3-D position vector in object space measured relative to a common origin in the object or its vicinity and $\bar{p} = k\,(\bar{1}_R - \bar{1}_i)$ is a 3-D position vector in Fourier space or $\bar{p}$-space with $\bar{1}_R$ and $\bar{1}_i$ being unit vectors in the directions of observation and incident illumination respectively and k is the wavenumber of illumination. The scattering function $\gamma(\bar{r})$ represents the 3-D geometrical distribution and strengths of those visible *scattering centers* or *differential scattering cross-sections* of the body that give rise to the measured field. Correction of the field measured in practice in this fashion for range-phase, clutter, and system response [5],[6] leads to accessing $\Gamma(\bar{p})$ over those values of $\bar{p}$ employed in the measurement. We will designate the measured data manifold by $\Gamma_m(\bar{p})$ and assume without further elaboration here that the values of $\bar{p}$ utilized always sample the $\bar{p}$-space in a manner satisfying the Nyquist criterion to avoid aliasing in the reconstruction. The size and shape of the accessed Fourier region depends on geometry and on the extent of the spectral and angular apertures utilized i.e., on the values k and $(\bar{1}_R, \bar{1}_i)$. It is possible then as shown by computer simulation in [3] and [4] to retrieve tomographically a *diffraction and noise limited* version $\gamma_d(\bar{r})$ of the object scattering function through application of the *Projection-Slice Theorem* derivable from the multi-dimensional Fourier transform [7]-[10].

The aim of this paper is to show that high resolution reconstruction of $\gamma_d(\bar{r})$ is possible by measuring $\Gamma_m(\bar{p})$ over a curved surface in $\bar{p}$-space rather than within a volume as would ordinarily be required for the retrieval of 3-D detail of $\gamma_d(\bar{r})$. Because, for a given fixed spectral range, the number of angular observation points needed to adequately sample the $\bar{p}$-space over a volume is considerably higher than the number needed to access the outer surface of the volume or a portion of it, a sizable reduction in the number of coherent sensors or receivers is achieved. In practice this translates into a proportionate reduction in the projected cost of high resolution wavelength diversity imaging apertures and would for example, open the way for cost-effective implementation of envisioned [11] giant *imaging radar networks*.

II. Theoretical Considerations

Let $H(\bar{p})$ be a Fourier space sampling function describing the values assumed by the vector $\bar{p} = k(\bar{l}_R - \bar{l}_i)$ during data acquisition. We can express then the Fourier space data manifold accessed by measurement as,

$$\Gamma_m(\bar{p}) = \Gamma(\bar{p})\ H(\bar{p}) \tag{1}$$

which can be regarded as a 3-D *Fourier transform hologram* of the scattering object. A diffraction and noise limited version $\gamma_d(\bar{r})$ of the object function $\gamma(\bar{r})$ can be obtained by Fourier inversion of eq. (1). That is (within a constant $1/(2\pi)^3$),

$$\gamma_d(\bar{r}) = \int \Gamma(\bar{p})H(\bar{p})\ e^{-j\bar{p}\cdot\bar{r}}\ d\bar{p} = \gamma(\bar{r}) *** (\bar{r}) \tag{2}$$

where

$$h(\bar{r}) = \int H(\bar{p})\ e^{-j\bar{p}\cdot\bar{r}}\ d\bar{p} \tag{3}$$

is the 3-D *impulse response* or *point spread function* of the system and the triple asterisks denote a 3-D convolution. Clearly, because $H(\bar{p})$ (which can be now identified also as a 3-D *transfer function* of the system) is dependent on recording geometry through $\bar{p} = k(\bar{l}_R - \bar{l}_i)$, the impulse response here, in contrast to conventional monochromatic imaging systems, is *spatially variant*. A valid question then is the identification of favorable recording geometries for which a narrow $h(\bar{r})$ is realized for a wide range of object bearings utilizing a minimum number of observation and/or illuminating points i.e., $(\bar{l}_R, \bar{l}_i)$ values in order to keep to a minimum the number of broadband coherent receivers and transmitters employed in the recording geometry. To provide an answer to this question we apply the projection-slice theorem to eq. (3). There are two forms of this theorem. One states that a projection (central slice) in Fourier space and a central slice (projection) in object space are a Fourier transform pair. The second form establishes a similar relationship between parallel *weighted projections* and parallel slices in the two domains [3],[9]. In the context of this analysis, the first version means that the 2-D Fourier transform of the projection of $H(\bar{p})$ on an arbitrarily oriented plane in p-space is a central slice through $h(\bar{r})$ oriented parallel to the projection plane. This immediately suggests that desirable recording geometries are those for which the projection of their $H(\bar{p})$ in any direction cover always extended areas whose 2-D Fourier transform will necessarily be concentrated in a narrow region indicating an $h(\bar{r})$ with central slices exhibiting peak amplitudes of narrow extent. If all central slices of $h(\bar{r})$ possess peaks of narrow extent the 3-D impulse response $h(\bar{r})$ will consequently be narrow. With this condition established, we consider next the two bistatic recording geometries shown in Fig. 1. In one (a) a randomly or regularly sampled circular recording aperture of diameter D is used to access a truncated conical volume in p-space with the truncation being set by the initial and final values of the range of wavenumbers k utilized in the measurement. In the second geometry (b), a number of random or equally spaced sampling points or coherent receivers distributed in a circle of diameter D are employed to access an identically shaped hollow truncated cone. Both geometries assume a centrally located coherent transmitter or illuminator T. The sampling functions $H(\bar{p})$ realized in both cases will coincide over the sidewalls of the truncated cones. A brief study of the two cases reveals that the shape and extent of the areas covered by nearly all projections of the solid truncated cone and the hollow truncated

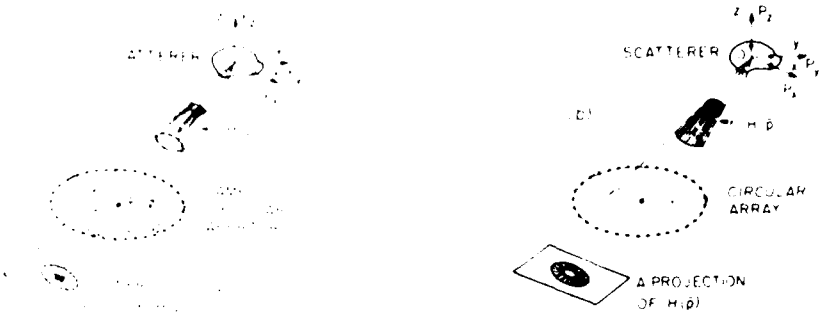


Fig. 1. Geometries for accessing the Fourier space of a 3-D scatterer. (a) with a sampled 2-D circular aperture, (b) with a circular array of sampling points.

cone are the same except for a few projections in those directions forming a small solid angle surrounding the OT Line where the hollow nature of $H(\overline{p})$ in (b) will be evidenced by a missing central region. This difference is illustrated by the two projection examples included in Fig. 1. However, because the outer boundaries of comparable projections of $H(\overline{p})$ in (a) and (b) are identical, the associated Fourier transforms representing the slices of $h(\overline{r})$ for both geometries are expected to possess central peaks of the same extent. The amplitude and shape of side-lobe structure in the outlying regions surrounding the central peaks will however differ somewhat because of the different number and distribution of data points in the comparable projections. It can be concluded therefore that the 3-D resolutions obtained with data accessed over a volume of $\overline{p}$-space and with data accessed over the outer surface of the volume are nearly the same. Verification of this conclusion is found in the results of a numerical simulation of microwave wavelength diversity imaging [3] which are presented in Fig. 2. A semi-circular array of sampling points consisting of 50 equally spaced receivers distributed over an arc of 130° as depicted in Fig. 2(a) is assumed. A 3-D test object consisting of two adjoining 1m diameter conducting spheres arranged as shown and centered at a distance R directly above the transmitter T is chosen. A spectral range of (2-4) GHz and a ratio of R/D =1 are assumed in computing $\Gamma_m(\overline{p})$. The far field scattered by the two spheres was computed using the Mie series formulation [12]. Weighted parallel projections in the direction p_z of the accessed $\overline{p}$-space data manifold lying on a truncated semi-conical surface represented by H(p) in Fig. 2(a) were obtained by multiplying $\Gamma_m(\overline{p})$ by the complex factor $\exp j(\alpha p_z)$, $\alpha = 0$, -30cm, -40cm, before computing the projections. This yields three weighted projection holograms shown to the left in Fig. 2(b) that correspond from top to bottom to parallel horizontal slices through the object in Fig. 2(a) at $z = 0$, -30cm and -40cm. The optically retrieved images from these projection holograms are shown to the right in Fig. 2(b). These demonstrate clearly the 3-D tomographic imaging capability from the limited p-space data accessed by the geometry of Fig. 2(a).

III. Acknowledgement

This research was supported by the Army Research Office under contract DAAG-29-80-K-0024P02 and by the Air Force Office of Scientific Research, Air

Force Systems Command under grant AFOSR-81-0240B.

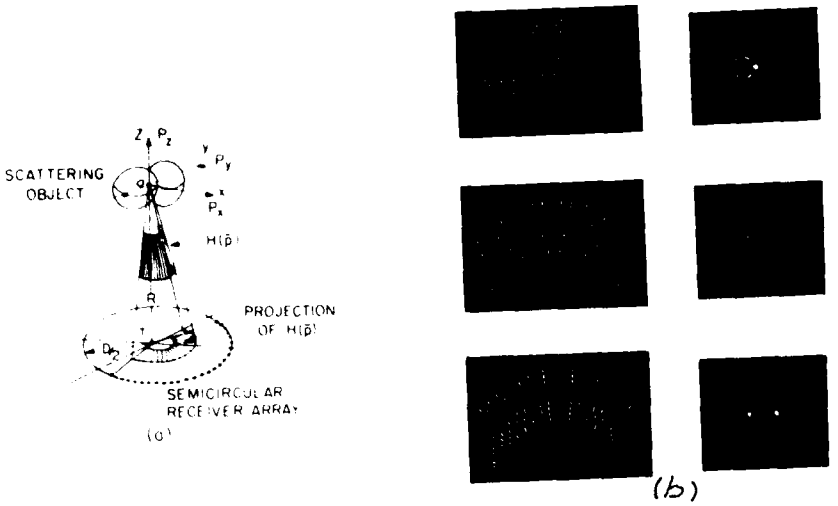


Fig. 2. Details of numerical simulation of 3-D tomographic image reconstruction from limited Fourier space data. (a) Recording geometry (b) Weighted projection holograms (left) and retrieved images of three slices through the illuminated portion of the two spheres.

References

1. N. Bojarski, Naval Air Command Final Report, N00019-73-C-0312F, Feb. 1974.
2. R.M. Lewis, IEEE Trans. on Ant. and Prop., Vol. AP-17, pp. 308-314, May 1969.
3. C.K. Chan and N. Farhat, IEEE Trans. on Ant. and Prop., Vol. AP-29, pp. 312-319, March 1981.
4. N.H. Farhat and C.K. Chan, "Three-Dimensional Imaging by Wave-Vector Diversity", Presented at the 8th Int. Symp. on Acoust. Imaging, Key Biscayne, Fla. 1978, and published in *Acoustical Imaging*, Vol. 8, A. Metherell (Ed.), pp. 499-516, Plenum Press, New York 1980.
5. N.H. Farhat and C.L. Werner, "An Automated Microwave Measurement Facility For Three-Dimensional Tomographic Imaging by Wavelength Diversity", Presented at the 1981 Intern. IEEE AP-S Symp./Nat. Radio Science Meeting, L.A., June 1981.
6. N.H. Farhat, C.L. Werner and T.H. Chu, "Microwave Projection Imaging With Near Optical Resolution", Submitted for publication, IEEE Trans. on Ant. and Propagation.
7. J. Radon, Ber. Saechs. Akad. Wiss. (Leipzig), Vol. 19, pp. 262-278, 1917.
8. R.N. Bracewell, Australian J. of Phys., Vol. 9, pp. 198-217, 1951.
9. G. Stroke and M. Halioua, Trans. Amer. Crystal. Assoc., Vol. 12, pp. 27-41, 1976.
10. H.H. Barrett and W. Swindle, *"Radiological Imaging:Theory and Image Formation, Detection and Processing*, Acad. Press, New York, 1982.
11. N.H. Farhat, W.I. Landauer and W.E. Wallace, "Computer Assisted Naval Applications of Holography", Computer Command and Control Co., Report No. 132-4, prepared for the Naval Analysis Program, Office of Naval Research, under contract N0014-69-C-0167, Feb. 1973.
12. R.F. Harrington, *"Time Harmonic Electromagnetic Fields*, McGraw-Hill, 1961.

Deblurring and Three-Dimensional Reconstruction from Multiple Linear-Tomograms

Satoshi Kawata
Department of Applied Physics
Osaka University
Suita, Osaka 565 Japan

Jack Sklansky
School of Engineering
University of California, Irvine
Irvine, California 92717

ABSTRACT

The image of the tomogram obtained by a conventional x-ray tomographic machine is degraded by the superposition of motion-blurred images of nonpivotal planes. We introduce a method to eliminate these blurred images from a tomogram. In this method a set of tomograms, each focused on one of a set of parallel planes, are combined to form a three-dimensional reconstruction of blur-free tomograms. This approach is equivalent to the inversion of a linear system. By a mathematical analysis of linear-motion tomography, we found that linear-motion tomography is restricted to angularly-limited frequency information. An iterative matrix inversion algorithm with the constraints of nonnegativity and finite-extent is applied to the reconstruction of the plane of interest from a set of tomograms.

SUMMARY

1. INTRODUCTION

Even after the introduction of the computed tomographic (CT) scanner, conventional x-ray tomographic machines are still widely used in hospitals. This is because conventional tomography has the following benefits in comparison to CT:

1) simple and inexpensive equipment,
2) small radiation dose, and
3) no restriction with slice-angle.

However, compared to CT, conventional tomography has poor image quality. The images of the planes adjacent to the pivot plane are blurred and superimposed on the tomographic image of the pivot plane. To suppress this blur, high-pass spatial filters have been applied to tomograms [1].

In this paper we propose another approach to suppress superimposed non-pivotal planes in a tomogram. We combine the image of multiple tomograms, each focused on a distinct member of a set of parallel planes, so that nonpivotal planes are eliminated from the image of the tomogram focused on the plane of interest. In current practice, one often takes a set of tomograms rather than a single tomogram at one session. For this purpose, a multiple-film cassette, which provides multiple tomograms by a single motion-scan, is commercially available.

2. BASIC FORMULATION OF CONVENTIONAL X-RAY TOMOGRAPHY

In conventional x-ray motion tomography, an x-ray source moves synchronously with the film so that the projection of a particular plane (the pivot plane) in the object remains stationary relative to the film during the exposure. The projections of all other planes move relative to the film and are blurred on the film. Figure 1 illustrates the geometry of linear-motion tomography.

Let the three-dimensional (3-D) distribution of the attenuation coefficient of the object and the one-dimensional distribution along the path of the x-ray source during the exposure be denoted by $a(x,y,z)$ and $h(x)$, respectively, where x and y form the x-ray source plane, and z is the distance from the x-ray source plane to the film plane. The image of the tomogram in the pivot plane $z = p$ is given by

$$t(x,y,p) = \iint a\left(\frac{\xi}{(kp-\zeta)/\zeta}, \frac{y}{(kp-\zeta)/\zeta}, \zeta\right) \cdot \frac{\zeta}{k|p-\zeta|} \; h\left(\frac{x-\xi}{\frac{k}{\zeta}|p-\zeta|}\right) d\xi \, d\zeta \qquad (1)$$

where the film is located at $z=kp$ (k times farther than the distance of the pivot plane from the source plane), and where the path of the moving x-ray source is assumed to be on the x-axis.

Because $|p-\zeta|$ appears in the denominator of the integrand in Eq. (1), most of the integration is produced for values of ζ close to p. Thus, in the above integral,

$$\frac{kp-\zeta}{\zeta} \cong \frac{kp-p}{p} = k-1$$

and

$$\frac{k}{\zeta} \cong \frac{k}{p} .$$

Hence

$$t(x,y,p) \cong \iint a\left(\frac{\xi}{k-1}, \frac{y}{k-1}, \xi\right) \cdot \frac{p}{k|p-\zeta|} \; h\left(\frac{x-\xi}{\frac{k}{p}|p-\zeta|}\right) d\xi \, d\zeta$$

Since we are interested in obtaining a three-dimensional tomographic reconstruction, we replace p by z. This yields

$$t(x,y,z) \cong \iint a\left(\frac{\xi}{k-1}, \frac{y}{k-1}, \zeta\right) \cdot \frac{z}{k|z-\xi|} \; h\left(\frac{x-\xi}{\frac{k}{z}|z-\zeta|}\right) d\xi \, d\zeta$$

The above equation may be expressed as the following two-dimensional convolution:

$$t(x,y,z) \cong a\left(\frac{x}{k-1}, \frac{y}{k-1}, z\right) ** h'(x,z) \qquad (2)$$

where ** denotes two-dimensional convolution, and $h'(x,z)$ is the following two-dimensional (2-D) point spread function (PSF) expressed in terms of the one-dimensional PSF $h(x)$:

$$h'(x,z) = \frac{\beta}{|z|} h\left(\frac{\beta x}{|z|}\right). \qquad (3)$$

where β is a constant in the convolution, and subsequent to the convolution is replaced by z/k. Thus the PSF $h'(x,z)$ varies spatially with respect to the pivot-to-source distance z (or, equivalently, p).

Figure 2 illustrates this 2-D PSF and its 2-D Fourier transform (the transfer function) of linear tomography, in the case of the constant-speed motion of the x-ray source within the finite path length. The transfer function, illustrated in Figure 2(b), has angularly bounded ranges where all values are zero. This means that linear tomography system misses the information of the spatial frequency components of these ranges in the original signal. In other words, we cannot obtain a unique solution of the double integral equation of Eq. (1). Similar problems can be seen in CT reconstruction with limited view angle [2,3] and in coded-aperture imaging [4].

3. CONSTRAINED ITERATIVE RECONSTRUCTION

If some constraints given by equalities and inequalities are added to Eq. (1), which agree with the physics of the object and the system, the solution of Eq. (1) with these constraints may be unique, and the missing information can be recovered. There have been many works of such constrained signal recovery algorithms proposed mainly in the field of diffraction-limited optics [5].

Our approach for 3-D reconstruction from multiple linear tomograms is an iterative matrix inversion algorithm, which is well known in numerical analysis. This algorithm has been successfully used for digital image restoration [6,7]. It also appears in the well-known CT reconstruction algorithm, ART (Algebric Reconstruction Technique).

As a digital version of Eq. (1), we can express linear tomography by a linear matrix equation:

$$\underline{t} = [H]\,\underline{a}, \qquad (4)$$

where vectors $\underline{t}$ and $\underline{a}$ are 3-D distributions of the tomograms and the object, respectively. The elements of the individual vectors represent the sampled voxels (3-D elements) of the tomograms and the object, respectively. $[H]$ is the matrix of the PSF, or the sampled version of Eq. (3).

The estimate of $\underline{a}$ in n iterations, $\hat{\underline{a}}^{(n)}$, by the successive displacement method with constraints is given by

$$\underline{\hat{a}}^{(n)} = \underline{\hat{a}}^{(n-1)} + [R^{(n)}][D]^{-1}\{\underline{t} - ([D] + [f])\hat{a}^{(n-1)} - [E]\underline{\hat{a}}^{(n)}\}, \qquad (5)$$

where [D] is the diagonal matrix whose diagonals are the same as those of [H], [E] and [F] are the lower and the upper triangle matrices of [H] with zeros on the diagonals, respectively, and $[R^{(n)}]$ is the diagonal matrix to constrain the output within a bounded region. The a priori information, such as the maximum limitation of the object size in 3-D space and the possible attenuation coefficient values, that is between 0 and 1, can be used for this constraint matrix.

If the number of tomograms available is insufficient to form the 3-D space of the object, we have to interpolate or extrapolate the tomograms along the z axis. The method applicable for this purpose is an iterative algorithm [8] that minimizes the sum of the squared error between the object (i.e. the true attention function and the computed 3-D reconstruction. This algorithm is also based on an iterative matrix inversion algorithm [9].

4. EXPERIMENTAL RESULTS AND CONCLUSIONS

We carried out computer simulations of a 3-D reconstruction of a computer generated phantom with 11 planes by the constrained iterative matrix inversion algorithm. The results encourage us to believe that this algorithm will be useful for 3-D reconstruction of blur-free tomograms from a set of conventional x-ray linear tomograms.

The iterative matrix inversion algorithm is very efficient in computation time and in memory capacity when the matrix is very sparce and the diagonal elements are relatively larger than others. Linear tomography satisfies these conditions, as well as the condition for convergence. The condition for convergence is guaranteed by the fact that any eigenvalue λ_j of $[D]^{-1}[H]$ lies between 0 and 1. In other words, the optical transfer function (OTF) of linear tomography has a nonnegative value at every frequency, as illustrated in Figure 2(b) [6].

ACKNOWLEDGMENT

This research was supported by the National Cancer Institute of U.S. Public Health Service under Grant No. CA-32847.

REFERENCES

1. P. Edholm and L. Quiding, "Elimination of blur in linear tomography," Acta Radiologica (Diagn), Vol. 10, 441-447 (1970).
2. T. Inouye, "Imag reconstruction with limited angle projection data," IEEE Trans. Nucl. Sci., Vol. NS-26, 2666-2669 (1979).
3. K. C. Tam and V. Perez-Mendez, "Tomographical imaging with limited-angle input," J. Opt. Soc. Amer., Vol. 21, 582-592 (1981).
4. M. Y. Chiu, H. H. Barrett, R. G. Simpson, C. Chou, J. W. Arendt and G. R. Gindi, "Three-dimensional radiographic imaging with a restricted view angle," J. Opt. Soc. Amer., Vol. 69, 1323-1333 (1979).
5. A. Papoulis, Signal Analysis, (McGraw-Hill, New York, 1977) Chap.7-2.
6. S. Kawata and Y. Ichioka, "Iterative image restoration for linearly degraded images. I. Basis," J. Opt. Soc. Amer., Vol. 70, 762-768 (1980).
7. S. Kawata and Y. Ichioka, "Iterative image restoration for linearly degraded images. II. Reblurring procedure," J. Opt. Soc. Amer., Vol. 70, 768-772 (1980).

8. R. W. Gerchberg, "Super-resolution through error energy reduction," Opt. Acta, Vol. 21, 709-720 (1974).
9. B. P. Medoff, W. R. Brody and A. Macovski, "Image reconstruction from limited data," Proc. Int. Workshop on Phys. Eng. in Med. Imag., Pacific Grove, California, 1982) available from IEEE Computer Society.

FIGURES

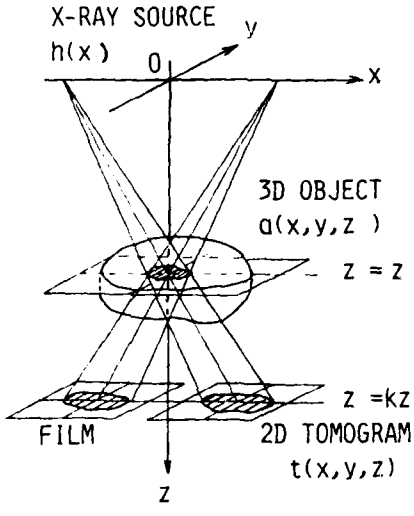


Figure 1 Geometry of linear-tomography.

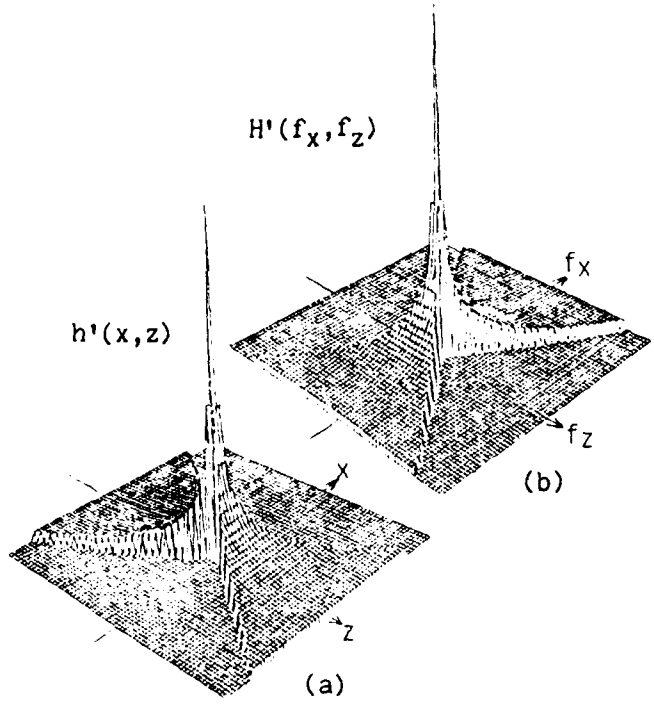


Figure 2 (a) PSF; and (b) transfer function of linear-tomography.

Two-dimensional Reconstructions from One-dimensional Data by Maximum Entropy

Sibusiso Sibisi

Department of Applied Mathematics and Theoretical Physics,
Silver Street, Cambridge CB3 9EW, England

The Maximum Entropy Method (MEM) is a powerful information-theoretic approach to the inversion of many types of data in science and engineering. It has been used for reconstructing positive images in such areas as radio astronomy [1,2], medical tomography [3], plasma diagnostics [4] and crystallography [5]. Practical data are always corrupted by noise and are usually incomplete: for example, there may be missing projections or missing Fourier components. Also the instrumental response function may be incompletely known [6]. These factors can lead to severely ill-posed inversion problems.

We apply the method here to Fourier Transform Nuclear Magnetic Resonance (FTnmr). The experiment consists of perturbing the magnetic moments of a set of chemical species with a transient magnetic field and observing the profile of oscillating decays as the moments relax to their equilibrium configuration in an underlying uniform field. The data are a time-series representing the overall magnetic moment, from which we wish to recover the decays and frequencies of the individual species. The early data points in the time-series frequently need to be rejected because of distortions due to finite pulse width. The data are therefore incomplete. There is also usually an imperfectly known time delay before detection and an imperfectly known misalignment angle between the detectors and the net magnetic moment at time t = 0. MEM can recover these parameters as well as finding both the frequencies and decays, even though the recovery of decays is a notoriously ill-posed problem.

Our problem is of the novel type

$$(1) \qquad F(t) = \int_0^\infty dk \int_{-\infty}^{\infty} d\omega f(k,\omega) \exp(-kt - i\omega t)$$

in which a 2-dimensional function $f(k,\omega)$ representing a set of oscillators of strength $fd\omega dk$ in the frequency/decay cell $d\omega dk$ produces a 1-dimensional time-series $F(t)$. With practical FTnmr data there is also an unknown origin shift in t and an unknown overall phase shift $\exp(i\phi)$: determination of these is discussed elsewhere [7]. The data D_t are noisy samples

$$(2) \qquad D_t = \mathrm{Re}(F(t)) + r_t \sigma_t$$

where σ_t is a standard deviation and r_t is random with unit variance. From D_t one wishes to reconstruct $f(k,\omega)$.

The conventional technique is based on Fourier transforming D_t. An individual species μ of frequency ω_μ and decay k_μ would give a Lorentzian frequency spectrum $f_\mu k_\mu/(k_\mu^2 + (\omega-\omega_\mu)^2)$ peaked at ω_μ with width k_μ, and one attempts to describe the data transform $D(\omega)$ as a sum of such Lorentzians. However, close rapidly decaying components overlap and become confused. Apodisation techniques [9,10] are used to improve resolution. For instance, if the decays k_μ are all the same, equal to k_o say, then multiplying $F(t)$ by $\exp(k_o t)$ and then Fourier transforming yields, ideally

$$(3) \qquad D(\omega) = \sum_\mu f_\mu \delta(\omega-\omega_\mu)$$

But this is unsatisfactory when the decay rates differ and a simple exponential filter is liable to give large noise amplification. Noise may be suppressed by modifying the apodising filter. For example, $\exp(k_o t - Ct^2)$ where $C>0$ is often used but suppressing noise leads to poorer resolution. In general there is no all-purpose apodisation filter.

To do MEM we first construct the chisquared statistic

$$(4) \qquad \chi^2(f) = \sum_t r_t^2 = \sum_t \left[(\mathrm{Re}(F(t)) - D_t)/\sigma_t\right]^2$$

and define the set $E = \{f:\chi^2(f)\leq\chi_o^2\}$ as the 'feasible' set of reconstructions consistent with the data, where χ_o^2 is some chosen tolerance level (say, the value corresponding to 99% confidence). Any f for which $\chi^2(f)>\chi_o^2$ is rejected: its residuals are too large for it to be consistent with the data. However, we still need to select just one reconstruction from the large set E. To this end we introduce entropy.

The configurational entropy of a reconstruction is [8]

$$(5) \qquad S(f) = -\int\int dk d\omega\, p(k,\omega)\log(p(k,\omega)/m(k,\omega));$$

$$p(k,\omega) = f(k,\omega)/\int\int dk d\omega\, f(k,\omega)$$

Here, $m(k,\omega)$ is a measure representing the relative importance of different regions of the (ω,k) plane. The global (unconstrained) maximum of S is attained when $f(k,\omega)$ is directly proportional to the measure. Maximising S subject to the feasibility constraint yields that unique reconstruction which has least overall structure relative to m. Spurrious structures due to noise and instrumental artefacts are automatically suppressed [1].

To determine m, we observe that the natural measure in the frequency/decay-time phase-plane (ω,T) is uniform i.e. there is an uncertainty relation $\delta\omega\delta T = \text{const.}$ between ω and $T=k^{-1}$. This transforms to a measure $m(\omega,k) = \partial(\omega,T)/\partial(\omega,k) = k^{-2}$ in the (ω,k) plane. We can now use MEM to find the optimal two-dimensional reconstruction $f(k,\omega)$ which fits the data.

We are currently using the technique on real FTnmr data [7], but here we report simulated results from a 5-component example. Fig.1 shows this example as a 2-dimensional plot against decay k and frequency ω. Fig.2 is the corresponding time-series data D_t sampled at 128 points with gaussian noise of unit standard deviation (relative to $\int\int f(k,\omega)dkd\omega = 200$). Fig.3 is the conventional spectrum of Lorentzians $D(\omega)$. Peak (b) appears to be of lower amplitude than (c) because of its faster decay. But, because of the strong overlap, it is not possible to tell which is the broader, let alone deduce the width. Fig.4 is the spectrum obtained after applying the apodisation filter $\exp(k_o t - Ct^2)$ to the data. Here, k_o is the decay of the peaks (d) and (e) and $C=k_o/T_o$ where T_o is the length of the time-series. Noise has been amplified and the amplitude ordering is clearly wrong.

The MEM 2-dimensional reconstruction is in Fig.5. Peaks are well resolved in ω and the decay values may be read off the k axis although peaks are broad along this axis. This is because of the inherent difficulty of recovering decays. The width in k gives the substantial uncertainty in the decay values whilst the small width in ω shows that each frequency is well-determined.

Fig.6 is the projection of Fig.5 onto the frequency axis. By comparison with Figs.3 and 4 the peaks are much better resolved and the ordering of the amplitudes has been corrected. The wavy background in Figs.3 and 4 due to noise is absent from Fig.5. Thus, even when decay information is removed by projecting onto the frequency axis, MEM is still inherently superior to conventional apodisation because it allows each component to be fully resolved in ω, regardless of its decay rate.

Finally, we point out that this analysis is possible because of the relative sparseness of the components; it is not realistic to hope to recover more than about one component in the decay axis for a given frequency. The method, however, remains relevant for FTnmr and for problems in engineering which have relatively few modes of oscillation.

References

1. Gull, S.F., Daniell, G.J.: 1978, Nature 272, 686.
2. Bryan, R.K., Skilling, J.: 1980, Mon.Not.R.astr.Soc.191,69.
3. Kemp,M.C.: 1980, Intl.Symposium on Radionuclide Imaging, IAEA-SM-247, 128, Heidelberg.
4. Skilling, J.: 1981, Workshop on Maximum Entropy and Data Analysis, Univ. of Wyoming, Laramie, Wyoming.
5. Collins, D.G.: 1982, Nature 298, 49.
6. Scott, P.F.: 1981, Mon.Not.R.astr.Soc. 194, 25P.
7. Sibisi, S.P.: 1982, Ph.D. Thesis, in preparation.
8. Jaynes, E.T.: 1968, IEEE Trans. SCC-4, 227.
9. Ernst, R.R.: 1966, Advances in Magnetic Resonance,ed. J.S. Waugh.
10. Campbell, I.D., Dobson, C.M., Williams, R.J.P., Xavier,A.V.: 1973, J.Mag.Res. 11, 172.

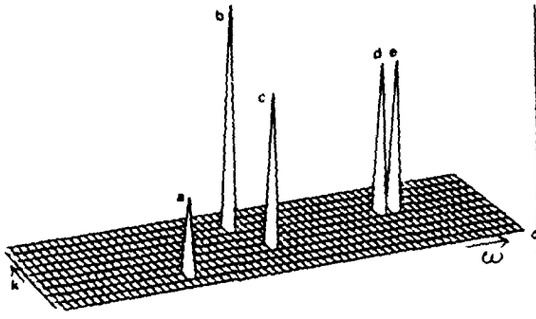


Fig.1 Simulation

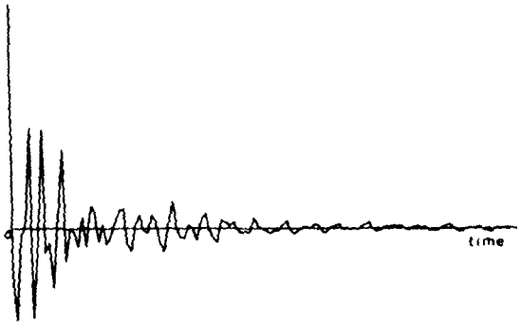


Fig.2 Corresponding time-series data

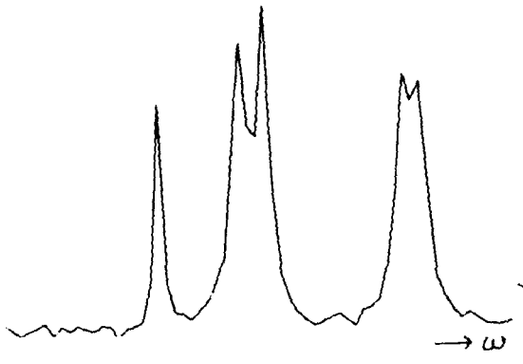


Fig.3 Conventional reconstruction. Fourier transform of time-series.

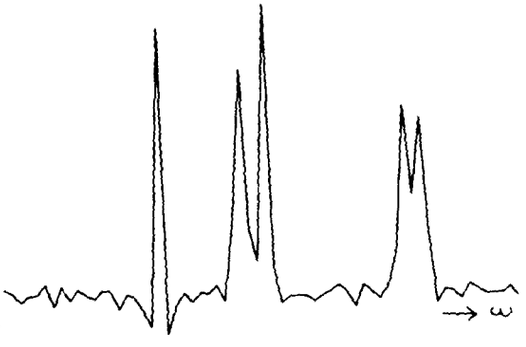


Fig.4 Fourier transform of apodised time-series

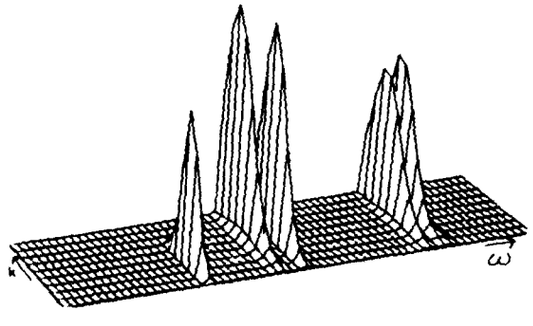


Fig.5 MEM 2-dimensional reconstruction

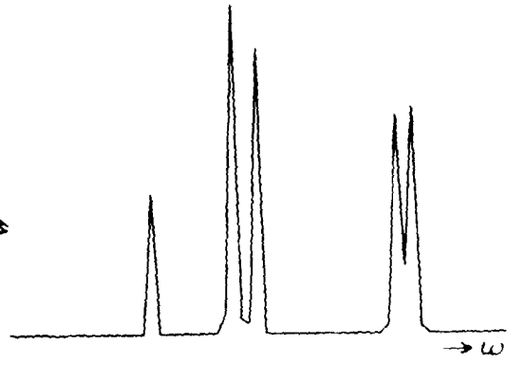


Fig.6 Projection of Fig.5 to frequency axis

SESSION XI

PROBLEMS AND METHODS II

J. A. Neff, *Presider*

The Phase Problem of X-Ray Crystallography from the Viewpoint of Signal Recovery

GERARD BRICOGNE
Department of Biochemistry
College of Physicians & Surgeons
Columbia University
630 West 168th Street
New York, New York 10032

The diffraction of X-rays by crystals was discovered in 1912 by Laue who proposed a theory of the phenomenon based on its analogy with optical diffraction by gratings. It is most easily understood from the standpoint of Fourier analysis, in which language it was reformulated shortly afterwards. The electron density in a crystal is periodic, and may therefore be conceived as a superposition of plane waves whose wave vectors belong to a "reciprocal lattice" dual to the crystal lattice. The contribution of each wave, or system of fringes, is described by a complex Fourier coefficient, whose amplitude gives the strength of this system of fringes, and whose phase determines the position of the fringes relative to some fixed origin. If all the amplitudes and phases of these waves are known, it is possible to obtain a picture of the electron distribution in the crystal by superposing them; that is, by a simple Fourier synthesis.

In an X-ray diffraction experiment, however, only the intensities of the diffracted beams can be measured, and these are essentially the squares of the amplitudes of the Fourier coefficients; the corresponding phases are lost in the measurement process, and must be restored by some means before the desired picture of the crystal structure can be obtained. This constitutes the "Phase Problem" of X-ray crystallography.

The early crystal structure determinations were performed by trial-and-error methods in which the atomic positions were first guessed, then adjusted so as to produce the best fit between calculated and observed intensities.

The first conceptual advance took place in 1934, when Patterson had some seminal discussions at M.I.T. with Norbert Wiener who was then developing the Fourier transform into a tool for statistical analysis. As a result, he realized that a Fourier series having as its coefficients the directly observable intensities, and thus containing no phase information, would yield, in Wiener's language, the autocorrelation function of the electron density in the crystal. The positions of the peaks of this function (the Patterson function) would thus give the interatomic vectors of the structure. In simple cases, the atomic positions can be inferred by trial and error from this set of interatomic vectors. This method of "Patterson deconvolution" became the basis of most crystal structure determinations. Patterson's work was important in another respect: he

showed that the phase problem may not always have a unique answer, by giving an example of two crystal structures (which he called "homometric") which were distinct and yet had the same autocorrelation function.

The first break-away from trial-and-error methods was the use of heavy atom substitutions to determine the phases of the Fourier coefficients. A classical example of the application of this technique is Robertson's solution of the phthalocyanin structure in 1936: by comparing the intensities diffracted by the organic molecule to those obtained after insertion of a nickel or platinum atom at the centre of the organic group, most phases can be determined and the Fourier synthesis can proceed without any previous model.

Both Patterson and heavy-atom substitution methods have since grown enormously in sophistication and power in the hands of protein crystallographers, but their conceptual foundations still go back to the mid-thirties.

Only relatively late was it realized that some special features of the electron density in a crystal, namely positivity and atomicity, lead to mathematical relationships between the phases and the amplitudes of the Fourier coefficients, so that a purely mathematical solution to the Phase problem could be contemplated.

Harker & Kasper showed in 1948 that positivity implies a set of inequalities involving the Fourier coefficients, which were later (1950) shown by Karle & Hauptman to be the simplest of a large family of determinantal inequalities. The general theory of such inequalities had been known to mathematicians since the work of Toeplitz in 1911.

Atomicity was exploited in two different ways, one of them analytical and the other statistical.

The analytical approach is that of Sayre, who in 1952 derived a set of nonlinear convolution equations satisfied by the Fourier coefficients of any structure consisting of equal, resolved, spherical atoms. Sayre's equations had an immediate impact, and their implementation led to the first "direct method" of X-ray structure determination.

In the statistical formulation, proposed by Hauptman & Karle in 1953, atomicity is interpreted as a statement that crystal structures may be thought of as being generated by placing identical atoms randomly, and independently of each other, throughout the unit cell of the crystal. The Fourier coefficients of such structures are therefore sums of a large number of independent identically distributed random variables (the contributions from individual atoms); the Central Limit Theorem of probability theory may then be invoked to obtain asymptotic estimates of their joint distribution. Such joint distributions of Fourier coefficients can yield phase information if knowledge of the amplitudes is assumed. These ideas have had a profound, if somewhat delayed, influence on the development of direct methods. Thirty years later, the initial formulation of Hauptman & Karle has been much improved upon, and its connections with classical methods of probability theory have been greatly clarified. However, their original intuition of the nature and power of the

probabilistic approach to the phase problem has remained the basis of all later developments.

Recent work by the present author has been aimed at relating the mathematical apparatus of probabilistic direct methods to that used in various branches of signal processing, and in particular to Jaynes's maximum-entropy formalism.

The representation of a crystal structure as an "atomic chaos", used by Hauptman & Karle, is a device for ranking sets of Fourier coefficients by the value of their joint probability. An alternative procedure would consist in ranking the corresponding electron density maps according to their relative frequency of occurrence in the statistical ensemble of structures generated by the same random process. The relative frequency of a map can be evaluated by simple combinatorial arguments, and its logarithm is found to be proportional to the Shannon entropy of the map. Probabilistic direct methods are thus closely related to Jaynes's maximum-entropy principle, which has been applied successfully to many problems of spectral estimation and of image reconstruction from incomplete or noisy data. This connection can be made more precise: it will be shown that the maximum-entropy principle can be rigorously derived from the hypotheses on which probabilistic direct methods are based, and that it embodies the optimal exploitation of these hypotheses.

In spite of this unification of hitherto distinct formalisms, the X-ray phase problem retains a unique level of difficulty among the signal recovery problems so far tackled by the maximum-entropy method. Because of the non-convexity of the observational constraints represented by the knowledge of the intensities, the maximum-entropy condition does not define a unique set of phases, and any solution algorithm must accommodate a high degree of branching. A strategy for dealing with this problem by sequential decision methods will be presented.

General References

BLUNDELL, T.L. & JOHNSON, L.N. (1976). "Protein Crystallography," Academic Press, London.

GIACOVAZZO, C. (1980). "Direct Methods in Crystallography," Academic Press, London.

HAUPTMAN, H. & KARLE, J. (1953). "The Solution of the Phase Problem I. The Centrosymmetric Crystal." ACA Monograph no. 3, Polycrystal Book Service, New York.

LEVINE, R.D. & TRIBUS, M. (1979). "The Maximum Entropy Formalism." M.I.T. Press, Cambridge.

IMAGE RECONSTRUCTION AND PARTIAL DECONVOLUTION WITH SUPPORT CONSTRAINT SEPARATION ANGLE AND LEAST-SQUARES INTERPOLATION PROCEDURES

A. Lannes

Laboratoire d'Optique Electronique
29, rue Jeanne Marvig, 31055 Toulouse Cedex, France

In image reconstruction from projections [1], and more generally, in any partial deconvolution with support constraint, *the central problem is to specify the conditions under which it is possible to interpolate, in a bounded region W, the Fourier transform of an object function with support in a bounded region V* [2,3,4]. In particular, it is then essential to understand, in an analytical way, the parts played by the size of V and W, and the geometry of the whole V,W-configuration (cf. for example Fig. 1). The aim of this communication is to summarize the corresponding analysis, and to visualize the main results with the aid of appropriate representations.

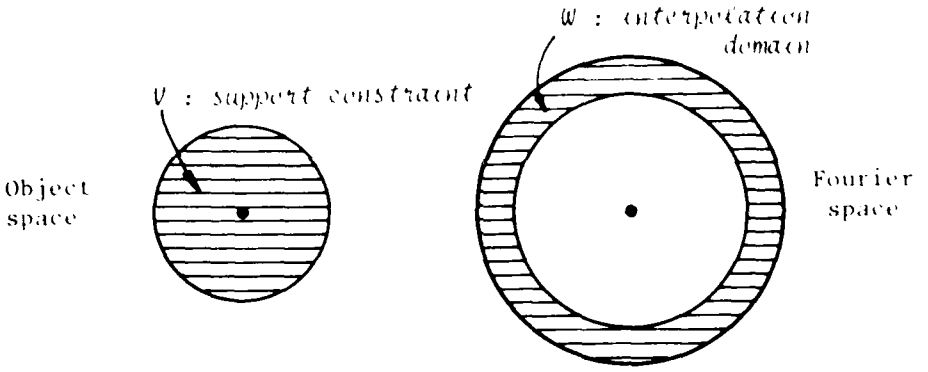


Fig. 1 : - Fourier transform interpolation of functions with bounded supports. An example of V,W-configuration.

1. PRINCIPLE OF THE INTERPOLATION METHOD

In the geometrical illustration of the problem, presented on Figure 2, the infinite dimensional Hilbert space of square-integrable functions is schematically identified with the three-dimensional Euclidean space. By hypothesis, the object function Φ has its support contained in V ($\Phi = 0$ outside V). Denoting by F the Fourier transform operator, and by w the characteristic function of W ($w = 1$ in W, $w = 0$ outside W), the binary filter eliminating the spectral regions in W is the operator of orthogonal projection onto K' : $Q' = F^{-1}(1-w)F$. Let $\Psi = Q'\Phi$ be the projection of Φ onto K', $\tilde{\Psi}$ K' be an approximation to Ψ, and ε'^2 be the corresponding quadratic error :

$$\varepsilon'^2 = \|\tilde{\Psi} - \Psi\|^2 = \int|\tilde{\Psi} - \Psi|^2 = \int|\hat{\tilde{\Psi}} - \hat{\Psi}|^2 \qquad (\hat{\Psi} = F\Psi).$$

The problem is to obtain an approximation to Φ, Φ_s, using as data $\tilde{\Psi}$, and in addition, to understand how the error

$$\varepsilon_s' \quad \|\Phi_s - \Phi\|^2$$

depends on V and W, in particular. Clearly, when W is bounded, the problem consists essentially in interpolating $\hat{\Phi}$ in W from its values outside W. Strictly speaking, this is only true if $\varepsilon' = 0$.

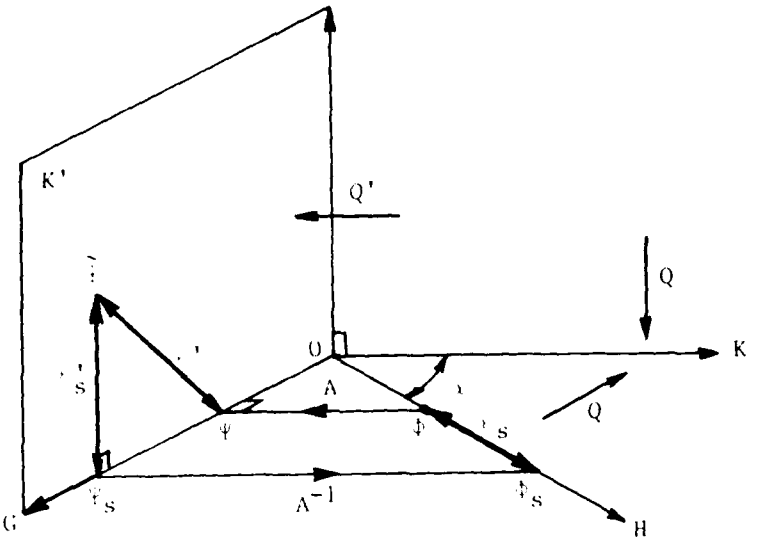


Fig. 2 : - geometrical illustration of the interpolation principle.

- H : *subspace of functions* Φ *with support in* V *(object space)*
- K : *subspace of functions the support of the Fourier transform of which is in* W
- K' : *subspace of functions the support of the Fourier transform of which is outside* W *(image space)*
- Q : *operator of orthogonal projection onto* K *(binary filter eliminating the spectral regions outside* W*)*
- Q' : *operator of orthogonal projection onto* K' *(binary filter eliminating the spectral regions in* W*)*
- G : *subspace of functions* Φ *filtered by* Q' *(*G = Q'H = AH *is the range of* A*)*
- α : *separation angle between* H *and* K.

Let A be the mapping of H onto G induced by Q' (cf. Fig. 2). The least-squares solution of the equation $A\Phi - \tilde{\Psi} = 0$ is defined by the relation

$$\Phi_s = A^{-1}\Psi_s$$

where Ψ_s is the orthogonal projection of $\tilde{\Psi}$ onto G. In the Euclidean case presented on Figure 2, A is invertible *iff* (if and only if) the angle α between H and K is different from 0. Then, $\varepsilon_s^2 = (\varepsilon'^2 - \varepsilon'^2)/\sin^2\alpha$.

The concept that generalizes the traditional notion of angle with respect to the problem under consideration is that of *separation angle*. It is easy to show from Proposition 1 in [2] that

$$\cos^2\alpha = \sup_{\|\Phi\|=1} \|w\hat{\Phi}\|^2 \ . \qquad (1)$$

Thus, $\cos^2\alpha$ is the largest fraction of energy that the Fourier transform of a function with support in V can have in W. According to the Parseval identity, $\sin^2\alpha$ is the smallest fraction of energy that $\hat{\Phi}$ can have outside W. It then

follows that the quadratic error ε_s^2 is governed by the relation :

$$\varepsilon_s^2 \leq \frac{\varepsilon'^2 - \varepsilon_s'^2}{\sin^2\alpha} \quad . \tag{2}$$

This inequality shows in an explicit way that *the closer to $\pi/2$ is the separation angle, the more reliable is the interpolation process*. Conversely, when α is small, the problem is practically ill-conditioned.

2. INTERPOLATION PROCEDURE

Denoting by P the operator of orthogonal projection onto H, the least-squares solution is explicitly given by the relation

$$\Phi_s = A^{-1}\Psi_s \quad \text{with} \quad \Psi_s = A(PA)^{-1}P\tilde{\Psi}; \tag{3}$$

thus, $\Phi_s = (PA)^{-1}P\tilde{\Psi}$ where $P\tilde{\Psi} = \nu\tilde{\Psi}$ ($\nu = 1$ in V, $\nu = 0$ outside V). Since, for any $\Psi \in K'$ and $\Phi \in H$, $\langle\Psi \mid \Phi\rangle = \langle\Psi \mid Q'\Phi\rangle = \langle P\Psi \mid \Phi\rangle$, the operator of H into K' induced by Q', and that of K' into H induced by P are mutually adjoint. Note that the smallest spectral value of PA is equal to $\sin^2\alpha$ (cf. [2] and [3]).

Always subject to the condition $\alpha > 0$, and provided that ω lies in the open interval $]0, 2[$, the Jacobi sequence

$$\Phi_0 \ , \ \Phi_{n+1} = \Phi_n + \omega P(\tilde{\Psi} - A\Phi_n) \tag{4}$$

converges towards Φ_s for any starting function $\Phi_0 \cdot H$ (cf. Proposition 1 in [3] with $\alpha' = 0$). The optimal relaxation parameter ω_o and the corresponding asymptotic rate of convergence are given by the relations :

$$\omega_o = 2/(2 - \cos^2\alpha) \quad \text{and} \quad \nu(\omega_o) = \cos^2\alpha/(2 - \cos^2\alpha). \tag{5}$$

The fact that $\nu(\omega_o)$ is a decreasing function of α shows that *the closer α is to $\pi/2$, the easier is the interpolation process.*

According to Equation (5), ω_o is greater than 1 and less than 2. As indicated in the next section, it is generally possible to specify its value , *a priori*, from the knowledge of V and W alone.

When Φ_0 is taken equal to 0, and ω is set equal to 1, the Jacobi sequence (4), $\Phi_{n+1} = P\tilde{\Psi} + PQ\Phi_n$, corresponds to explicit calculation of the series

$$\sum_{n=0}^{\infty} (PQ)^n \, P\tilde{\Psi} \ , \tag{6}$$

which yields Φ_s from the alternating Projection Theorem (cf. Eq. (13) in [4]; by hypothesis, the projection of Φ onto the orthogonal complement of H is equal to 0). It should be noted that this choice of ω is only suited to cases where α is close to $\pi/2$.

3. BEHAVIOUR IN V AND W OF THE SEPARATION ANGLE

For clarity, we assume that V and W are two-dimensional bounded regions, and give the corresponding results without proof. The reader interested in the underlying mathematical material may consult References [2] and [5].

Let η be the geometric mean of the measures (or areas) of V and W :

$$\eta = [\mu(V)\ \mu(W)]^{\frac{1}{2}}. \tag{7}$$

By definition, η is a size parameter without physical dimension. As expected, when η tends to infinity, α tends to 0 more quickly than any integral power of of $1/\eta$. It follows that the interpolation process can only be performed if η is

relatively small. The behaviour of α for small η is essentially governed by the relation

$$\sup\{0, \eta(1 - J\eta^2)\} < \cos\alpha < \inf\{1, \eta\} \text{ where } J = 2\pi^2 \int_V \int_W (\vec{x}.\vec{u})^2 d\vec{u}\, d\vec{x}. \quad (8)$$

Here, V and W, the centres of gravity of which coincide with the origin, denote the sets of measure 1 similar to V and W respectively. More precisely, provided that η is sufficiently small, $\cos\alpha$ can be obtained from a series of the form

$$\cos\alpha = \eta \sum_{m=0}^{\infty} (-1)^m k_m \eta^{2m} \quad (9)$$

where the coefficients k_m (for $m > 0$) depend on the geometry of the V,W-configuration ($k_0 = 1$, $k_1 = J$, etc...). For example, when V is a disk and W has rotational symmetry (cf. Fig. 1), we have

$$k_1 = \frac{1}{2} D_1 \ , \ k_2 = \frac{5}{6} D_2 - \frac{1}{12} D_1^2 \ , \ k_3 = \frac{7}{4} D_3 - \frac{1}{6} D_1 D_2, \text{ etc...}$$

with

$$D_m = \frac{2\pi}{m!^2} \int r(\pi r^2)^m w(r)\, dr \quad (w = 1 \text{ in W}, w = 0 \text{ outside W}).$$

The coefficients k_m for m greater than 2 or 3 can be easily obtained by performing appropriate algebraic calculations with the aid of the computer[6]. In the special case where V and W are disks, the previous expressions yield :

$$k_1 = \frac{1}{4} \ , \ k_2 = \frac{7}{144} \ , \ k_3 = \frac{1}{192} \ , \ \text{etc...}$$

We then retrieve in a quite independent manner, the coefficients that can be directly obtained from an analysis based on the prolate-spheroidal wave functions [7]. *The question raised by Landau and Pollak (in Reference [8], Para. 5.2) about how to deal with problems involving regions (essentially) other than disks has therefore been solved.*

We may now advance the following conjecture : for a given value of η, the minimal value of α is attained when V and W are disks. Intuitively, with regard to the interpolation process, this is indeed the worst situation. According to our analysis, this conjecture is at least true for η asymptotically small : the minimal value of $k_1 = J$ is attained when V and W are disks (cf. Eq. (8)). For large η the problem remains open.

References

[1] G.T. Herman, Image Reconstruction from Projections, 1979, Springer-Verlag, Berlin.
[2] A. Lannes, J. Math. Anal. Appl., 1980, 74, n° 2, 530.
[3] A. Lannes, J. Optics (Paris), 1982, 13, n° 1, 27.
[4] W.D. Montgomery, Optics Letters, 1982, 7, n° 1.
[5] A. Lannes, J. Math. Anal. Appl. (to be published).
[6] E.W. Ng, Symbolic and Algebraic Computation, 1979, Springer-Verlag, New-York.
[7] D. Slepian, B.S.T.J., 1964, 43, 3009.
[8] H.J. Landau, H.O. Pollak, B.S.T.J., 1961, 40, 65.

LBI Image Recovery Using Sharpness Maximization

D.G. Steer, M.R. Ito, P.E. Dewdney
Dept. of Electrical Engineering, U.B.C.
Dominion Radio Astrophysical Observatory, H.I.A.
P.O. Box 248, Penticton, B.C. V2A 6K3, CANADA

In radio Long Baseline Interferometry (LBI), the image is recovered from measurements made in the aperture plane. However, two major difficulties hinder the imaging process.

The atmosphere contributes errors in the phase of the sampled complex signal, and the small number of points sampled in the aperture makes the point response function or "beam" of the instrument have significant positive and negative sidelobes over a large area of the image. Imaging under these conditions of unknown phase and poor sampling is an active area of interest for radio astronomers.

One of the significant effects of th phase errors is to "spread out" the image. While the true image is usually empty sky with a few isolated patches of brightness, the distorted image is often a wide distribution of low level brightness. One possible way for recovering the true image is to adjust the phase at each antenna in the array until the image with the minimum spreading is obtained. Muller [1], Brown [2], and McCall [3] report on adaptive optic systems and simulations to recover images based on this principle. These systems make use of a parameter measured in the image plane to control the aperture plane phase adjustments. This parameter is often referred to as the "sharpness", and is considered for this discussion to be the summation of all pixel amplitudes cubed. The operative assumption is that the sharpness parameter will be a maximum for the undistorted (true) image.

A series of simulation experiments were conducted to see if the sharpness system could be used for LBI image recovery. These experiments were undertaken as part of a wider survey of a number of potential and practical LBI image recovery techniques. There are two practical reasons for considering the sharpness approach. The first is that the imaging can proceed without the need to remove the "beam" from the image during the phase recovery. Most other LBI imaging processes (eg. Readhead [4] and Cornwell [5]) are iterative, and the image must be deconvolved from the beam as an integral part of each iteration. As the deconvolution is a tedious calculation, there is an advantage in separating the two operations. The sharpening method allows the "dirty image" (including the beam) to be recovered, and then the beam deconvolution to be done only once in a separate, independent operation. The second advantage is that the phases may be corrected individually for each antenna and it is thus possible to obtain the sharpened image which is in agreement with the "phase closure constraint". This constraint specifies that phase errors cancel in certain sums of the measured phases, and thus recovered images should have the same closure sums (eg. Jennison

[6]). The sharpened dirty image is thus simple to calculate and remains true to the measured data.

The simulation experiments were conducted by choosing some simple test images and convolving them with an LBI beam. Typically, the image consisted of 3 or 4 Gaussian sources and the LBI telescope array included 6 antennas. The simulated LBI data was then spoiled by artificial phase errors which were added to simulate errors at individual antennas and could be adjusted so that the maximum error was a fixed fraction of 360°. No amplitude errors were included. It was found that the sharpness was always larger for the true image than for the distorted image for random phase errors. There was a factor of approximately 10 between the sharpness with phase errors of 360°. Unfortunately, it was also found that the autocorrelation (AC) image had an even larger sharpness of about twice the true image value. The AC image may be obtained by setting all the aperture phases to a constant, and the result is an image that is symmetric about the origin. The large sharpness is a consequence of the AC image having a large brightness at its center. This result suggests that the sharpness recovery process will be hindered by confusion between the true and the AC image.

Tests were also made to examine the variation in the image sharpness as the antenna phase was varied from 0 to 360°. It was found that the variation was not the expected simple sine curve. The sharpness was more sensitive to phase changes at antennas involved in short baselines, and several peaks in the sharpness function were observed for these antennas. For antennas involved in longer baselines, the effect on the sharpness was less pronounced, and a single shallow peaked curve usually resulted (Figure 1). The difference in sensitivity can be explained by the differences between the aperture amplitudes sampled by the short and the long baselines. Since the amplitudes are largest near the origin, where they will be sampled by the shortest baselines, the sharpness is more sensitive to phase changes in these signals.

Finally tests were conducted to see how the process worked at restoring images. In this case the antenna phases were adjusted sequentially by a successive approximation method to obtain the maximum sharpness for the image. Usually several passes were made with the data until the sharpness was no longer increasing. It was found that, for images with small phase errors (±36°), a reasonably correct image could be recovered if the individual phase adjustments were kept small (± 23°). For larger errors and larger adjustments to the antenna phase, an image resembling the AC image would usually result. This image sharpness was, as expected, larger than the true image value. With larger phase errors and smaller phase adjustments an incorrect image was produced. This image would usually contain a number of compact sources; however, their distribution and sizes would not correspond to the true image. The final sharpened image was influenced by the initial image pattern with the sharpened peaks corresponding to the largest speckles in the distorted image. An

example showing the initial and the sharpened images is given in figure 2.

In conclusion it has been shown that image recovery by maximizing a global image parameter such as sharpness is not practical for LBI imaging. With large phase errors, the sharpness function has several maximums, each of which corresponds to a different, compact brightness pattern. The sharpening operation is unable to distinguish one peak from another. Similar problems are to be expected with comparable optical systems with large phase errors.

References

[1] R.A. Muller & A. Buffington
"Real-Time Correction of Atmospherically Degraded Telescope Images Through Image Sharpening"
J.Opt.Soc.Am. Vol 64 #9 Sept 1974 pp1200-1210

[2] T.M. Brown
"Reconstruction of Turbulence-Degraded Images Using Nonredundant Aperture Arrays"
J.Opt.Soc.Am. Vol 68 #7 July 1978 pp883-892

[3] S.L. McCall, T.R. Brown & A. Passner
"Improved Optical Stellar Image Using Real-Time Phase-Correction System: Initial Results"
Astrophys.J. Vol 211 Jan 1977 pp463-468

[4] A.C.S. Readhead & P.N. Wilkinson
"The Mapping of Compact Radio Sources from VLBI Data"
Astrophys.J. Vol 223 July 1978 pp25-36

[5] T.J. Cornwell & P.N. Wilkinson
"A New Method for Making Maps with Unstable Radio Interferometers"
Mon.Not.R.Astr.Soc. Vol 196 1981 pp1067-1086

[6] R.C. Jennison
"A Phase Sensitive Interferometer Technique for the Measurement of the Fourier Transform of Spatial Brightness Distributions of small Angular Extent"
Mon.Not.R.Astr.Soc. Vol 118 #3 1958 pp276-284

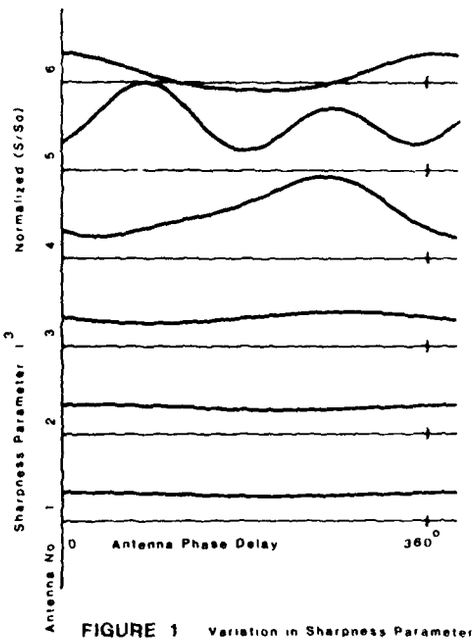


FIGURE 1 Variation in Sharpness Parameter with Antenna Phase

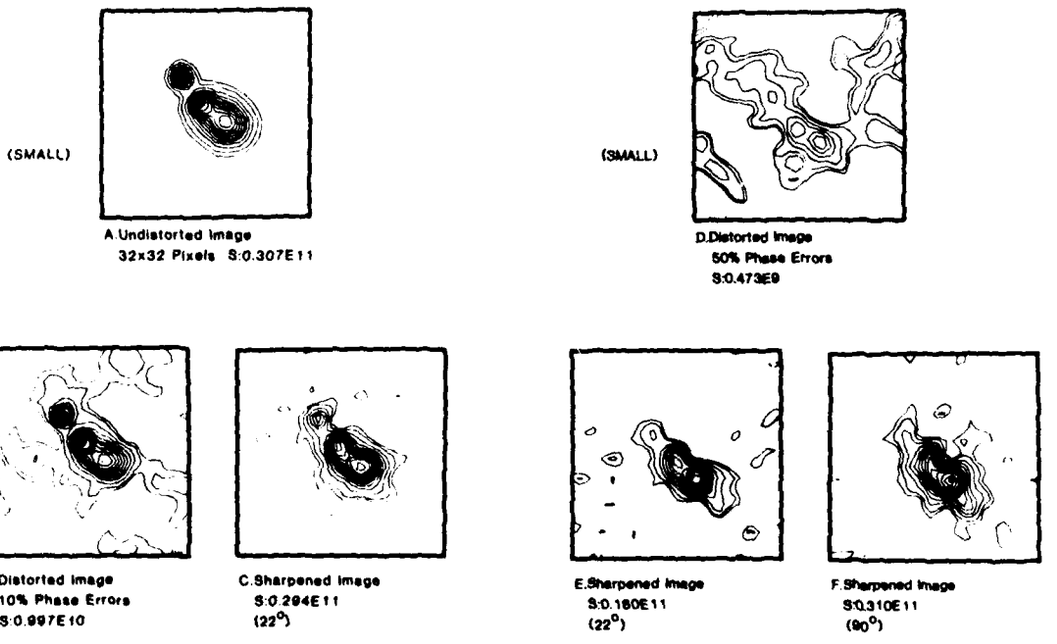


FIGURE 2 Sharpened Images

PHASE SYNCHRONIZATION OF DISTORTED IMAGING ANTENNA ARRAYS

Bernard D. Steinberg
Valley Forge Research Center, University of Pennsylvania
200 S. 33rd Street, Philadelphia, Pennsylvania

I. Diffraction-limited performance of an imaging system is often unattainable without some feedback-controlled compensation built into the image-forming process. Dielectric-constant perturbations due to atmospheric turbulence distort the phasefront of the optical radiation field. Muller and Buffington have discovered a class of integrals of the image intensity which, when maximized by adjustments of a compensating lens or mirror, reduce the error in the image to zero, except for an unknown shift in the optical axis [1]. This is a remarkable theorem. Its success depends upon the spatial incoherence of optical sources. Another approach, due to Gerchberg and Saxton [2], utilizes known properties of the class of expected signals, their autocorrelations or their Fourier transforms. It introduces considerable heuristics into the iterative Fourier transformation process. After each successive transformation, portions of the image or its Fourier transform are retained and portions deleted. Retention or deletion is based on a priori scene information and the physics of the process. Ref. [3] and [4] apply [2] to the optical problem.

Low frequency acoustical wave propagation in the ocean also suffers ray path distortion due to variations in the refractive index of the medium. Unlike the optical atmospheric problem, the distortion-inducing phenomenon is not confined to a thin layer near the imaging instrument. One adaptive-beamforming approach in a medium modeled by a zero-mean, random, additive component to the propagation velocity having prescribed correlation properties has shown that the aperture size can be extended considerably and that diffraction-limited performance can be preserved, although in a severely restricted field of view [5]. A similar problem exists in ultrasonic imaging within the human body [6], although no solutions have been forthcoming.

The microwave imaging problem has its own peculiarities. Because of the long wavelength (1-30 cm), the aperture size required for picture taking with microwaves is huge. To achieve resolving power of common optical instruments such as the camera or small telescope, a microwave antenna need be 100 m to 30 km in size. The large size demands a distributed array such as the radio astronomical Very Large Array in New Mexico [7]. However, for terrestrial imaging purposes, the site cannot be a fully dedicated site, implying that the aperture of the system must be a highly thinned antenna array in which the individual parts are separated from each other and distributed over the requisite area, and the design must circumvent the conventional requirement that the array surface be planar or any other a priori shape. Practicality demands that the design procedure be sufficiently robust to permit the element coordinates (x,y, and z) to be arbitrarily chosen and, furthermore, to possibly vary with time. This latter property imposes still another design constraint, which leads to the most difficult and most interesting condition; it precludes the a priori information needed to design the beamforming and scanning phase-shift program. Instead, some low order intelligence must be built into the system so that beamforming and scanning can proceed in the absence of accurate knowledge of antenna element position. The instrument which accomplishes this task is called the radio camera.

II. The microwave imaging problem differs from the optical problem in several basic ways. The sources in optical astronomy are spatially incoherent and are largely temporally incoherent (due to the very large spectral bandwidth). The converse is true in microwaves. The bandwidth is very narrow and, because the microwave signals are reflections from a transmitter and are not derived from active sources, the sources are spatially coherent. Thus the information necessary to cohere a distorted array must be derived from a coherent radiation field of known or assumed properties across the microwave aperture.

The sources of distortion are different, too. A microwave array may be distorted because it is too large to be surveyed properly. It may suffer from wind-loading. Its installation may be faulty. Or its distortion may be electrical rather than geometric: Medium turbulence can cause the integral of the dielectric constant over a path from a source or a target to the array to vary randomly with position in the array. This is similar to the optical problem. Multipath and scattering of the echoes can alter the amplitudes and phases of the signals received across the array. Electromagnetic coupling from the antenna elements to the local environment may vary with position in the array, causing random errors in the driving point impedances of the antenna elements.

This paper will demonstrate an experimental procedure based on a relatively simple algorithm which provides diffraction-limited performance from large microwave apertures that are nonrigid and poorly surveyed. The experiments were conducted with 3 cm wavelength and a 40 m antenna. Fig. 1 shows the experimental set-up. A single time-shared receiver was moved along a cable 10 m above the ground at the Valley Forge Research Center of the University of Pennsylvania. The cable was 39 m long and was free to sway with the wind, which it did. The radar transmitter was pointed toward a town at a distance of 6-7 km. The reference radiation field was obtained either from a corner reflector placed in the town or from stable target echoes having required reradiation properties [8]. The transmitter was pulsed 1,000 times per sec. The receiver was turned on 200 to 300 times at random intervals as it was moved over the 39 m course. The movement of the system took approximately 2 minutes. The estimated uncertainty in antenna location at each of the sample points was about 1 wavelength. Thus, the received echoes bore an uncertain phase so large as to make conventional imaging impossible.

The echo traces accepted by the receiver were coherently quadrature demodulated, samples every 25 ns, A/D converted and stored in the microprocessor. The beamforming and imaging algorithm consists of three parts. First, the variance of the amplitudes of the group of echoes from each range bin, measured at each element in the array, is calculated. The range bin having the lowest normalized echo variance is selected as the reference range for the system. Next, the processor either multiplied the complex sample at each array element from each range trace by the complex conjugate of the echo at the reference range, or more simply, merely phase rotated the received echoes by the phases of the complex conjugates of the signals at the reference range. This was the adaptive part of the process. Lastly, the processor applies linear phase weighting across the array to electronically scan the adaptively formed beam in angle.

The object of the first step is to find a target or a clutter patch whose reradiation most closely approximates that of the point source. The object of the second step is to self-cohere the array upon that target. The object of the third step is to scan the beam in angle to the left and right of that target.

This algorithm is very simple to implement. It requires no special filtering nor any complicated signal processing. The algorithm was tested extensively with ground based equipment [9] and with airborne radar data [10].

III. Each of the distortion-inducing sources mentioned above set limits to the compensatory capability of the algorithm. Since the algorithm is based upon the use of only a single reference source no possibility exists for surveying the element locations; at least three and preferably four reference targets or beacons are needed for self-survey [11]. As a consequence, compensation for element position errors is effective only in a relatively narrow sector in the direction of the reference source. The theory indicates that the loss in image contrast from the direction of the self-cohering reference source is, in decibels, $\Delta G_P = 4.3k^2\sigma_x^2\theta^2$ [12]. P pertains to position errors, $k = 2\pi/\lambda$ is the wave number, σ_x is the standard deviation of element position error normal to the direction to the reference source, and θ (rad) is the scan angle from that direction. Defining the field of view (FOV) as that angular sector (rad) in which the contrast has diminished by no more than 1 dB at the edges, the FOV of a radio camera is FOV = $\lambda/2\pi\sigma_x$. A value of σ_x equal to 10 wavelengths, which is 100 times the conventional image-forming tolerance, permits diffraction-limited operation with very little contrast penalty over a field of view of approximately one degree.

Another source of phase error is multipath and scattering. Energy from the target may be scattered by reflectors located outside the direct path to the array and some of the scattered energy may arrive at the array. Energy also may be refracted to the array by refractive index variations in the propagation medium. HF propagation in the ionosphere and acoustic propagation in the ocean offer examples of multipath propagation. The direct effect of the scattered energy is distortion of the phasefront of the radiation across the array. In turn, this distortion leads to phase errors in adaptive beamforming.

A theory useful for design purposes has been developed [13]. It shows that the loss in gain due to scattering, in decibels, is $\Delta G_S = 4.3S[1-\rho(\theta)]$. S is the power ratio of the scattered field at the array to the direct signal and $\rho(\theta)$ is the autocorrelation function of the phase error due to the scattered signal as a function of the scan angle from the direction of cohering of the array. The dB loss in gain is zero at the origin and grows asymptotically toward 4.3S dB. Thus the intensity of the scattered field, at the array, can be as large as 1/4 the direct field intensity without causing more than a 1 dB loss in image contrast.

The simplest source of phase error is measurement error associated with receiver noise. Assuming that the phase error is unbiased and uncorrelated from array element to element, the dB loss in gain is $\Delta G_N = 4.3/(SNR)$ where SNR is the power ratio of signal to noise. Since the minimum useful SNR in radar is approximately 20 the maximum expected loss is about 0.2 dB, which is very small. In general, receiver noise can be ignored as a source of loss in array gain when scanning the beam of the radio camera.

The combined effects of the phase errors from these sources of phasefront distortion upon the main-lobe gain of the antenna array and, therefore, upon the contrast in the image is, in decibels, $\Sigma\Delta G$.

IV. The raw image produced by the radio camera suffers two primary deficiencies. The first is the effect of incoherent residual errors. The second is the high artifactual content due to the inherent high sidelobes of sparsely filled apertures. The current state of the research on the following three techniques will be described. The first is the use of diversity techniques, e.g., element position, frequency. The second is image feedback control. Although the Muller-Buffington technique is limited in its direct applicability, there is optimism that a modified form of the theorem will prove useful for the relatively narrow band waveforms employed in microwave imaging [14]. Third, there is some evidence that the effect of the gaps in the aperture of a thinned array can be mitigated somewhat. Instead of forming the image as the square magnitude of the Fourier

transform of the samples of the electric field measured by the thinned array, all cross-products $e_i e_j^*$ are formed and a statistical interpolator such as maximum entropy or maximum likelihood is applied to bridge the gaps between them. The result is a gapless estimate of the autocorrelation function of the radiation field across the aperture. The Fourier transform of the correlation estimate becomes the image [15].

[1] R.A. Muller & A. Buffington, "Real-Time Correction of Atmospherically Degraded Telescope Images...," J.Opt.Soc.Am., 64, 9, Sept. 1974, pp. 1200-1210.

[2] R.Gerchberg & W. Saxton, "A Practical Algorithm for Determination of Phase...," Optik, 35, pp. 237-246.

[3] J. Fienup, "Space Object Imaging...," Optical Engineering, Sept.-Oct. 1979.

[4] J. Fienup, "Iterative Method Applied to Image Reconstruction...," Optical Engineering, May-June 1980, pp. 297-305.

[5] B. Steinberg & A. Luthra, "Simple Theory of the Effects of Medium Turbulence...," J.Acous.Soc.Am., 71 (3), March 1982.

[6] B. Hillenbrand & B. Brendon, Introduction to Acoustical Holography, Plenum, 1972.

[7] R. Hjellming & R. Bignell, "Radio Astronomy with the Very Large Array," Science, Vol. 216, 18 June 1982, pp. 1279-85.

[8] B. Steinberg, "Properties of Phase Synchronizing Sources for a Radio Camera," IEEE Trans. A & P, Nov. 1982.

[9] B.Steinberg, "Radar Imaging from a Distorted Array: The Radio Camera Algorithm and Experiments," IEEE Trans. A & P , Sept. 1981.

[10] B. Steinberg & E. Yadin, "Radio Camera Experiment...," Proc. IEEE, Jan. 1982.

[11] C.N.Dorny, "A Self-Survey Technique for Self-Cohering of Antenna Systems," IEEE Trans. A & P, Nov. 1978, pp. 877-881.

[12] B. Steinberg, "Design Approach for a High Resolution Microwave Imaging Radio Camera," J. Franklin Institute, December 1973.

[13] B. Steinberg & E. Yadin, "Effect of Multipath and Scattering on Array Gain of a Radio Camera," to be submitted to IEEE Trans. A & P.

[14] E. Attia, "The Applicability of the Muller-Buffington Technique to Microwave Images," Valley Forge Research Center Quarterly Progress Report No. 41, Sept. '82.

[15] H. Subbaram, "Maximum Entropy Interpolation in Thinned Arrays," Valley Forge Research Center Quarterly Progress Report No. 41, Sept. 1982.

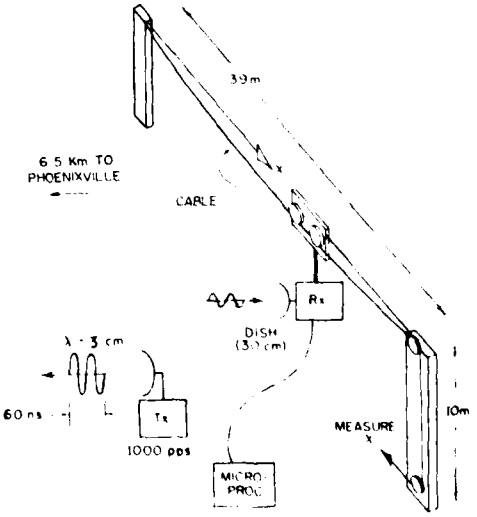


FIGURE 5.15 CABLE-SUPPORTED ARRAY FOR MAPPING PHOENIXVILLE.

KEY TO AUTHORS AND PAPERS