Blind Deconvolution of Still Images using Recursive Inverse Filtering

by

Deepa Kundur

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Electrical and Computer Engineering
University of Toronto

ISBN 0-612-07722-5

Name DEEPA KUNDUR

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

SUBJECT TERM

0544 U·M·I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984

Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

THE UNIVERSITY OF TORONTO LIBRARY
MANUSCRIPT THESIS - MASTER'S
AUTHORITY TO DISTRIBUTE

NOTE: The AUTHOR will sign in one of the two places indicated. It is the intention of the University that there be NO RESTRICTION on the distribution of the publication of theses save in exceptional cases.

a) Immediate publication in microform by the National Library is authorized.

Author's signature Deepa Krishnan Date June 7, 1995

- OR -

b) Publication by the National Library is to be postponed until _____ 19 ____
(normal maximum delay is two years).

Author's signature _____ Date _____

This restriction is authorized for reasons which seem to me, as Chair of the

Graduate Department of _____, to be sufficient.

Signature of Graduate Department Chair _____ _____

Date _____

INCLUDE IN UNBOUND COPY

BORROWERS undertake to give proper credit for any use made of the thesis, and to obtain the consent of the author if it is proposed to make extensive quotations, or to reproduce the thesis in whole or in part.

Signature of Borrower	Address	Date

1990

Blind Deconvolution of Still Images using Recursive Inverse Filtering

Master of Applied Science and Engineering 1995
Deepa Kundur
Department of Electrical and Computer Engineering
University of Toronto

Abstract

This thesis presents a novel blind deconvolution technique for the restoration of linearly degraded images without explicit knowledge of either the original image or point spread function. The technique applies to situations in which the scene consists of a finite support object against a uniformly grey background. This occurs in applications such as astronomy, and medical imaging. The only information required are the nonnegativity of the true image and the support size of the original object. A novel support-finding algorithm is proposed for situations in which the exact object support is unknown.

The restoration procedure involves equalization of the blurred image using a convex cost function. The performance of the technique for truncated equalizer parameters, and in the presence of noise are examined analytically. The new approach is experimentally shown to be more reliable and to have faster convergence than the existing nonparametric finite support blind deconvolution methods.

Acknowledgements

This thesis would not have been possible without the guidance of my supervisor Professor Dimitrios Hatzinakos; his insight, support and enthusiasm is very much appreciated.

I wish to also thank Takis Zourntos for his encouragement and help, as well as, Firas Jaton, and Jacek Ilow for their constructive comments during the development of this work. Lastly, I would like to acknowledge the Natural Sciences and Engineering Research Council of Canada for providing funding for this endeavour.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Image restoration is an important and difficult task in image processing. The goal of restoration is to reconstruct or recover an image that has been degraded by using some a priori knowledge of the degradation phenomenon and true image. Most restoration techniques model the degradation process and attempt to apply an inverse procedure to obtain an approximation of the original image. The various approaches that have appeared in literature depend on the particular degradation and image models [1].

The difficulty of image restoration arises largely because of the following:

1. The degradation process is often unknown and must be estimated from the blurred image characteristics.

2. Even when the degradation is known, the inversion process required may not be unique or may be ill-conditioned.

3. The actual noise value at each pixel is unknown and must be estimated using the statistics of the noise or degraded image.

1.1 The Linear Image Restoration Problem

A general degraded image-formation model is represented by the following equation:

$$g(u,v) = s[\int_{-\infty}^{\infty} h(u,v,u_1,v_1,f(u_1,v_1))du_1dv_1] \oplus n(u,v) \qquad (1.1)$$

where $g(u,v)$ is the degraded image, $f(\cdot)$ is the true image, $h(\cdot)$ is the point spread function[1] (PSF), and $n(u,v)$ is the noise process that affects the image formation at the time of recording. The symbol $\diamond$ corresponds to a point-by-point operation. Most recording systems, such as those which use photographic film, introduce a point nonlinearity represented by $s[\cdot]$ in equation 1.1.

Some assumptions may be made about the degradation process to simplify equation 1.1: the point-nonlinearity is negligible, the noise introduced to the system is purely additive, and the PSF acts as a linear shift invariant operator on the true image. These assumptions are commonly used in practice for a variety of applications [2]. The model simplifies to

$$g(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u-u_1, v-v_1) f(u_1,v_1)\,du_1 dv_1 + n(u,v) \tag{1.2}$$

With uniform sampling, the degradation model can be reduced to a discrete approximation of the form [1]

$$g(x,y) = h(x,y) * f(x,y) + n(x,y) \tag{1.3}$$

where (x,y) represents the discrete pixel coordinates, $h(x,y)$ is the impulse response of the sampled system (referred to as the PSF or blur), $g(x,y)$ and $f(x,y)$ are the average values of $g(u,v)$ and $f(u,v)$, respectively, over a pixel area in the sampling grid, and $*$ represents the two-dimensional discrete linear convolution operator.

In this simplified model, the observed image $g(x,y)$, true image $f(x,y)$ and noise $n(x,y)$ are coupled linearly, so the problem of recovering $f(x,y)$ from $g(x,y)$ is referred to as the *linear image restoration problem* [3]. Figure 1.1 gives an overview of the discrete linear degradation model.

Linearity of a restoration process refers to the chosen image degradation model, not the selected restoration scheme. Existing linear image restoration algorithms

[1] If we consider the true image to be a single bright point on a black background, the blurring of this point source is called the point spread function. In practice, the term point spread function is used to denote nonnegative blurring functions.

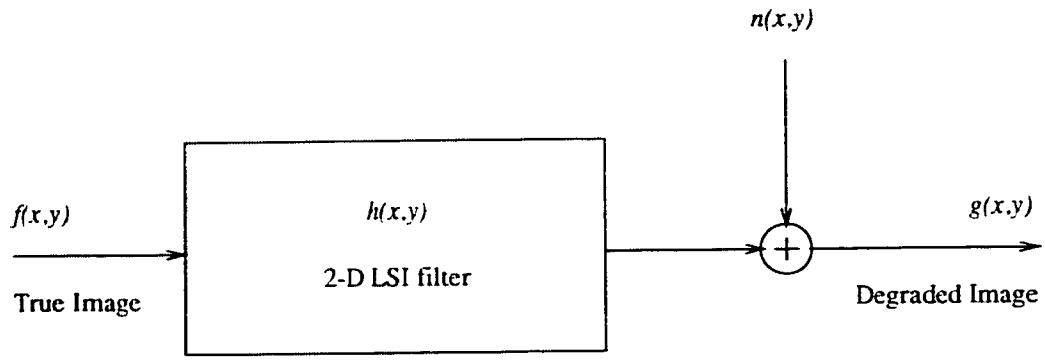

Figure 1.1: Linear Degradation Model

assume that the PSF is known a priori and attempt to invert it and reduce noise by using varying amounts of information about the PSF, true image, and noise statistics. In this thesis the linear degradation model of equation 1.3 is assumed.

1.2 Introduction to the Blind Deconvolution Problem for Images

In many situations, the PSF is unknown and little can be assumed about the original image. Therefore, the majority of existing image restoration techniques are not applicable for solving this type of problem. The process of simultaneously estimating the PSF (or its inverse) and restoring an unknown image is known as *blind image restoration*. For the linear degradation model of equation 1.3, it is specifically referred to as blind deconvolution. The blind deconvolution problem for images may be expressed as follows:

> Given a blurred image, find an estimate of the true image using partial or no information about the PSF and true image.

In practice, some a priori information is required to successfully restore the image. The partial information available is specific to each imaging application, so many diverse techniques for blind deconvolution of images have been proposed. The challenge is to design a method that exhibits the most appropriate compromise among computation complexity, reliability, and robustness to noise for a given application.

1.3 Motivation behind using Blind Deconvolution for Image Restoration

There exist several motivating factors behind the use of blind deconvolution for image processing applications. The most significant are discussed in this section.

In many practical situations, it is costly, dangerous or physically impossible to obtain a priori information about the scene to be imaged. For example, in applications like remote sensing and astronomy it is difficult to statistically model the original image or even know specific information about scenes never imaged before [4], [5]. In addition, the degradation cannot be accurately specified. In aerial imaging and astronomy, the blurring cannot be accurately modelled as a random process, since fluctuations in the PSF are difficult to characterize [6]. In real-time image processing, such as medical video-conferencing, the parameters of the PSF are not known to instantaneously deblur images [7]. Moreover, on-line identification techniques used to estimate the degradation may result in significant error, which can create artifacts in the restored image [8].

The hardware associated with image sensors is complicated and difficult to calibrate. Although calibration methods work well in measuring the impulse response of the imaging system, they are complex, which limits their wide usage [6], [9]. One might also want to restore images for which an impulse response calibration has not been performed or is imprecise. In addition, the high cost of some adaptive imaging systems makes imaging infeasible for some observations facilities. Even with adaptive-optics systems, the potential for phase error exists, so post-processing techniques, such as blind deconvolution, are still required [10], [11].

The physical requirements for improved image quality are unrealizable in some applications. For instance, in space exploration, the physical weight of a high resolution camera exceeds practical constraints. Similarly, in x-ray imaging, improved image quality occurs with increased incident x-ray beam intensity, which is hazardous to a patient's health [12].

It is clear that the development of a practical blind deconvolution scheme for

images would benefit many observation facilities.

1.4 Contributions of this Thesis

The objective of this thesis is to address many of the problems associated with existing blind deconvolution schemes for images. Existing methods exhibit a poor compromise among computational complexity, reliability of the solution, portability of the algorithm to different applications, and robustness to noise. Chapter 2 gives an overview of the existing classes of algorithms and discusses their strengths and limitations.

The first contribution of this work is the development of a novel blind deconvolution technique for the restoration of linearly degraded images. Explicit knowledge of either the original image or point spread function is not required. The proposed technique is relevant to applications in which an object of finite extent is imaged against a uniformly grey background. The edges of the object are assumed to be completely or almost completely included within the observed frame. This often occurs in applications such as astronomy, and medical imaging. The only information required for restoration is the nonnegativity of the true image, and support size of the original object. The restoration procedure involves recursively equalizing the blurred image using a convex cost function. Chapter 3 introduces the technique and presents a proof of convexity of the associated cost function. Uniqueness of the solution and suitable methods for optimization are discussed. Chapter 4 examines the behaviour of the proposed technique under practical conditions. The effect of using an FIR inverse filter of arbitrary dimensions on the quality of restoration is discussed. In addition, analytic expressions for the performance of the technique in the presence of noise are derived and methods of compensating for undesirable effects are proposed. The advantage of the proposed technique over existing methods is that convergence to the feasible set of solutions is guaranteed.

The second contribution is the design of a novel support-finding algorithm, for situations in which the size of the original object is unknown. The algorithm is based on the principle of cross-validation. Because it is computationally expensive, a

simpler approach is suggested which does not substantially sacrifice its performance. Chapter 5 provides a discussion of the algorithm

The third contribution of this thesis is a comparative study of the performance of the proposed technique with other methods belonging to the class of nonparametric, finite support blind deconvolution methods for images. The proposed technique is shown to produce more reliable results and to often converge faster than the other methods. In addition, it is more robust to overestimation of the support size. Chapter 6 presents simulations results showing the quality of restored images for each method.

Conclusions and further research are included in Chapter 7.

Chapter 2

Existing Classes of Blind Deconvolution Techniques

The methods developed for blind deconvolution of images apply to the restoration of grey-scale images. They can be grouped into six major classes based on their assumptions about the true image and blur. Various algorithms and implementations for each class exist. The following sections briefly describe these methods and discuss the advantages and disadvantages of implementing each for blind image restoration. Figure 2.1 provides a classification map of existing techniques.

The basis of comparison is broken down into four categories:

1. *Reliability of solution*: the convergence and uniqueness properties of the solution
2. *Computational complexity*: the computational processing required for proper restoration
3. *Portability of algorithm*: the potential for using the algorithm for many different applications
4. *Robustness to noise*: the reliability of the algorithm in the presence of additive noise

The relative importance of each of the above factors depends on the particular application. For example, in real-time image restoration, reducing computational

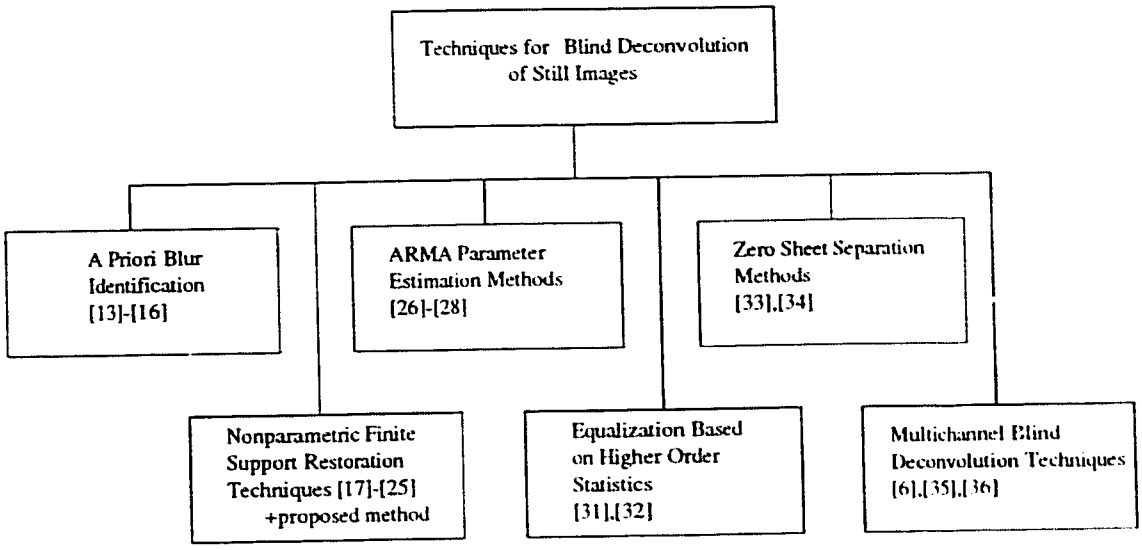

Figure 2.1: Classification Map of Existing Blind Image Restoration Techniques

complexity is of the utmost importance, but for medical imaging applications, the reliability of the solution is the primary consideration.

The general idea behind all blind deconvolution methods is to incorporate the partial information available about the image and blur into an optimality criterion. An estimate of the image and blur is produced by attempting to minimize or maximize this criterion.

2.1 A Priori Blur Identification Methods

In practice, the PSF is often identified from the degraded image characteristics before restoration. Based on this estimate, the image is restored using one of the classical restoration procedures.

Some commonly used identification techniques assume that a parametric model for the PSF is known. In particular, the blurring may be known to result from camera motion or an out-of-focus lens. For each of these PSFs a single parameter is used in the model, and the associated power spectrum of the PSF is known to contain zeros at locations related to this parameter value. Assuming that the power spectrum of the image has no zeros, an estimate of the PSF's parameter value is made

using the location of the zeros of the power spectrum of the degraded image [13]. A major drawback of these methods is their high sensitivity to noise. More robust techniques [14],[15] have been proposed to improve the accuracy of the identification of the spectral nulls in the presence of noise; however, such methods cannot identify PSFs that do not possess any spectral nulls, such as the Gaussian PSF.

Other techniques are based on local image characteristics. A point source, line or edge within the image is isolated, and the PSF is identified assuming radial symmetry [16]. In many cases, however, no special features on the blurred image can be established; moreover, the PSF estimate is dependent on the particular feature isolated within the image.

The main advantage of the techniques of this class is their low computational complexity.

2.2 Nonparametric Finite Support Restoration Techniques

In contrast to the a priori blur identification methods, the algorithms of this class do not assume parametric models for either the image or the blur. The true image is assumed to be positive, and to be comprised of an object with known finite *support* against a uniformly black, grey or white background. The support refers to the smallest rectangle within which the object is completely encompassed. If the background is black, which corresponds to a pixel value of zero, the support is the smallest rectangle within which the true image pixels are nonzero. The proposed blind deconvolution scheme falls into this class of techniques. The three most popular nonparametric finite support restoration methods are discussed in this section, and are used for comparison in Chapter 6.

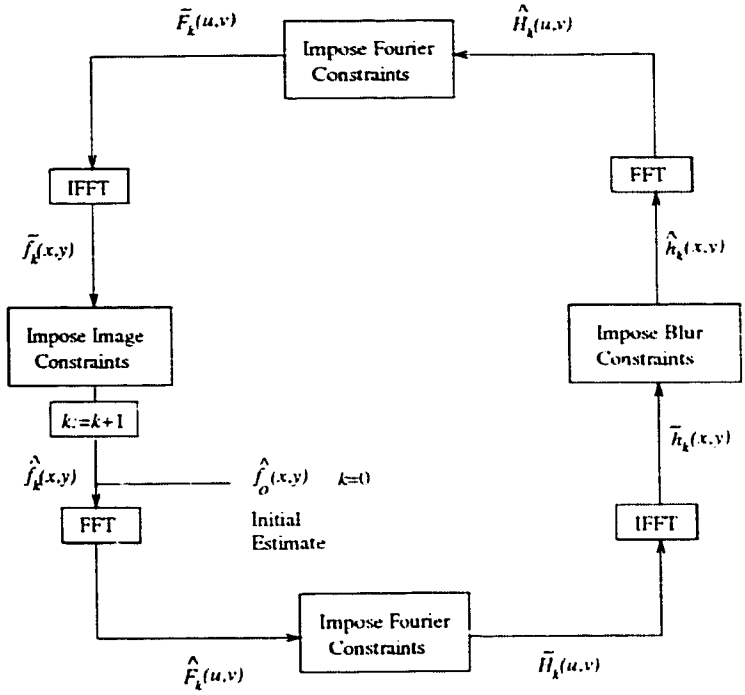

Figure 2.2: Iterative Blind Deconvolution Method

2.2.1 Iterative Blind Deconvolution Method

The iterative blind deconvolution method proposed by Ayers and Dainty [17] is, by far, the most popular method in this class of restoration techniques. The general algorithm is designed to be implementable on a digital computer using Fast-Fourier Transform (FFT) algorithms. The basic structure of the algorithm is presented in figure 2.2. The image estimate is denoted by $\hat{f}(x,y)$, the PSF estimate by $\hat{h}(x,y)$, and the linearly degraded image by $g(x,y)$. The capital letters represent Fast-Fourier Transformed versions of the corresponding signals. Subscripts denote the iteration number of the algorithm.

After a random initial guess is made for the image, the algorithm alternates between the image and Fourier domains, enforcing known constraints in each. The constraints are based on information available about the image and PSF. The image $f(x,y)$ and PSF $h(x,y)$ are both assumed to be positive with finite known support. The image domain constraints are imposed by replacing negative-valued pixels within the region of support and nonzero pixels outside the region of support with zero valued pixels. The Fourier domain constraint involves estimating the PSF (image) using the

FFT of the degraded image and image (PSF) estimate. That is, at the kth iteration,

$$\hat{H}_k(u,v) = \frac{G(u,v)\hat{F}^*_{k-1}(u,v)}{|\hat{F}_{k-1}(u,v)|^2 + \alpha/|\hat{H}_{k-1}(u,v)|^2} \tag{2.1}$$

$$\hat{F}_k(u,v) = \frac{G(u,v)\hat{H}^*_{k-1}(u,v)}{|\hat{H}_{k-1}(u,v)|^2 + \alpha/|\hat{F}_{k-1}(u,v)|^2} \tag{2.2}$$

The real constant α represents the energy of the additive noise and is determined by prior knowledge of the noise contamination level, if available. The value of α must be chosen carefully for reliable restoration. The algorithm is run for a specified number of iterations, or until the estimates begin to converge.

The method is popular for its low computational complexity. Many different implementations of this basic algorithm have been suggested. They differ in their assumptions about the true image and PSF, and how these assumptions are imposed in the image and Fourier domains [17], [18], [19], [21], [22]. Extensions have been proposed for situations where several degraded versions of the same image are available [20], [23]. Another advantage of this technique is its robustness to noise because of the Wiener-like filters (equations 2.1, and 2.2) used in the Fourier domain.

The major drawback of the method is its lack of reliability. The uniqueness and convergence properties are, as yet, uncertain. In addition, the restoration is sensitive to the initial image estimate, and the algorithm often exhibits instability.

2.2.2 Conjugate Gradient Method

The conjugate gradient method was proposed by Lane [24] to alleviate the problems associated with the instability of the iterative blind deconvolution method. The image and PSF are assumed to be nonnegative with known finite support. Essentially, the procedure involves the minimization of the following cost function using the conjugate gradient optimization routine:

$$J(\hat{f}(x,y),\hat{h}(x,y)) = \sum_{(x,y)\in\gamma_f} \hat{f}^2(x,y) + \sum_{(x,y)\in\gamma_h} \hat{h}^2(x,y) + \sum_{\forall(u,v)} |G(u,v) - \hat{F}(u,v)\hat{H}(u,v)|^2 \tag{2.3}$$

where γ_f and γ_h represent pixels for which the image and PSF estimates violate their known constraints.

The algorithm attempts to minimize equation 2.3 by varying the pixels corresponding to the image estimate $\hat{f}(x,y)$ and PSF estimate $\hat{h}(x,y)$, with arbitrary initial conditions for all pixels. Although the algorithm has reasonably low computational complexity and appears to be fairly robust to noise, it suffers from incorrect convergence to local minima. The cost J is multimodal, so the minimization routine is often trapped in local minima. For realistic images, it is almost impossible to achieve proper convergence.

2.2.3 Simulated Annealing Method

The simulated annealing approach by McCallum [25] entails the minimization of the following multimodal cost function:

$$J(\hat{f}(x,y),\hat{h}(x,y)) = \sum_{\forall(x,y)} [f(x,y) * h(x,y) - g(x,y)]^2$$

The image and PSF are assumed to be positive with known finite support. Using these constraints on $\hat{f}(x,y)$ and $\hat{h}(x,y)$, a simulated annealing procedure is employed for the minimization of J. In simulated annealing, estimates of the cost function parameters are iteratively varied to globally minimize J. The parameter values are randomly perturbed. If the perturbation reduces J, then it is accepted; if it increases J, then it is accepted with probability p. The probability p is reduced as iterations progress. In the case of infinite precision and infinitely many iterations, the procedure is guaranteed to reach the global minimum of a multimodal cost function.

The restoration algorithm is reliable and provides reasonable results in the presence of noise.

The major disadvantage is that convergence to the global minimum of the cost function is slow. The speed of convergence of the algorithm depends to a large extent on how quickly p is reduced.

2.3 ARMA Parameter Estimation Methods

Another approach for blind image restoration involves modelling the true image as a two-dimensional autoregressive (AR) process and the PSF as a two-dimensional linear system with finite impulse response. The degraded image is therefore modeled as a noisy observation of an autoregressive moving average (ARMA) process. Therefore, the blind deconvolution problem becomes an ARMA parameter identification problem.

Several approaches have been proposed to estimate the ARMA parameters. Among these are maximum likelihood estimation [26], [27], and generalized cross-validation [28].

Although these methods are reasonably robust to noise, there exist numerous limitations. Both techniques suffer from convergence to local optima. Various starting points may be used in the algorithm, with the hope that at least one initial condition will result in a globally optimal solution. However, as the number of unknown parameters increase with the order of the parametric models, the number of suboptimal solutions increase, and the number of initial guesses required to reach to global minimum grows unacceptably large. To compensate the PSF is assumed to be of a certain type (eg. Gaussian, out-of-focus).

In addition, second order statistics are used to identify the parameters. Therefore, the phase of the PSF is undetermined. To overcome this ambiguity the PSF estimate is constrained to be symmetric, that is, to have zero phase.

Although an AR model for the image can be justified in some situations [29], its use in restoration has shown to result in images with smoothed edges and ringing artifacts because of the model's insensitivity to abrupt changes in local image characteristics [30].

2.4 Equalization Methods Based on Higher Order Statistics

This class of techniques are useful for the restoration of texture images. They are based on the minimization of a given cost function which accounts for the probabilistic nature of the true image [31], [32]. The degraded image is passed through an inverse FIR filter yielding an estimate of the true image. The FIR filter parameters are updated in order to minimize a cost function which incorporates the higher order statistics (HOS) model of the true image.

The main advantage to these methods is that they allow the identification of non-minimum phase PSFs. The primary limitation is that the true image must be accurately modelled by a known non-Gaussian probability distribution.

2.5 Zero Sheet Separation Algorithms

The method of zero sheets has received attention because it provides valuable insight into the blind deconvolution problem in multiple dimensions. Lane and Bates [33] have shown that any degraded image g, formed by convolving several individual components $f_1, f_2, \ldots, f_n$, having compact support, is automatically deconvolvable provided its dimension is greater than one. Their argument rests on the analytic properties of the Laplace Transform (LT) in multiple dimensions. The zeros of the LT of a K-dimensional component f_i is necessarily continuous and lies on a $(2K-2)$-dimensional hyper-surface. By separating these hyper-surfaces, the individual components may be recognized up to a complex scale factor.

Although the technique does not require any a priori information other than the fact that the true image and blur are compact, there exist major drawbacks: the algorithm is highly sensitive to noise, and the fastest algorithm to date [34] has a computational complexity of $O(N^4)$, where N is the number of pixels of the blurred image.

2.6 Multichannel Blind Deconvolution Techniques

Multichannel techniques deal with situations in which differently blurred versions of the same image are available for processing. Some of them attempt to improve reliability of existing single-frame schemes by incorporating some method of averaging in the restoration procedure [35], [6].

The most successful methods of this class are the *cepstrum based high order statistics algorithms* [36]. The approach combines partial, higher order cepstral information from two differently blurred frames to provide information about the individual PSFs. Using this information, the individual PSFs are reconstructed and an estimate of the true image is computed. The PSFs are assumed to be FIR sequences with no zeros in common with each other, in common with the true image, or on the unit circle. The technique is robust to additive noise, and does not require any particular information about the image or PSF. The major limitation of the technique is that it is computationally intensive, and is implemented only for one-dimensional blurs.

Chapter 3

Proposed Recursive Inverse Filtering Algorithm

This chapter introduces the proposed nonnegativity and support constraints recursive inverse filtering (NAS-RIF) algorithm for blind deconvolution of images. A specific definition of the problem in terms of the partial information available on the true image and PSF is provided. The algorithm is described, and a proof of convexity of the associated cost function is given. The chapter concludes with a discussion of the numerical optimization routines used for minimization of the cost function.

3.1 Problem Definition and Assumptions

The objective of blind deconvolution of images is to construct a reliable estimate of the true scene. This task is achieved by using some information about the imaging process as a reference to deconvolve the true image and PSF from the blurred image.

The problem addressed by this thesis may be stated as follows:

> Given a grey-scale image $g(x,y)$ degraded by a linear shift invariant PSF $h(x,y)$, find a reliable estimate of the true image $f(x,y)$ given partial or no information about the PSF and true image.

The goal is to obtain a scaled shifted version of the true image. That is,

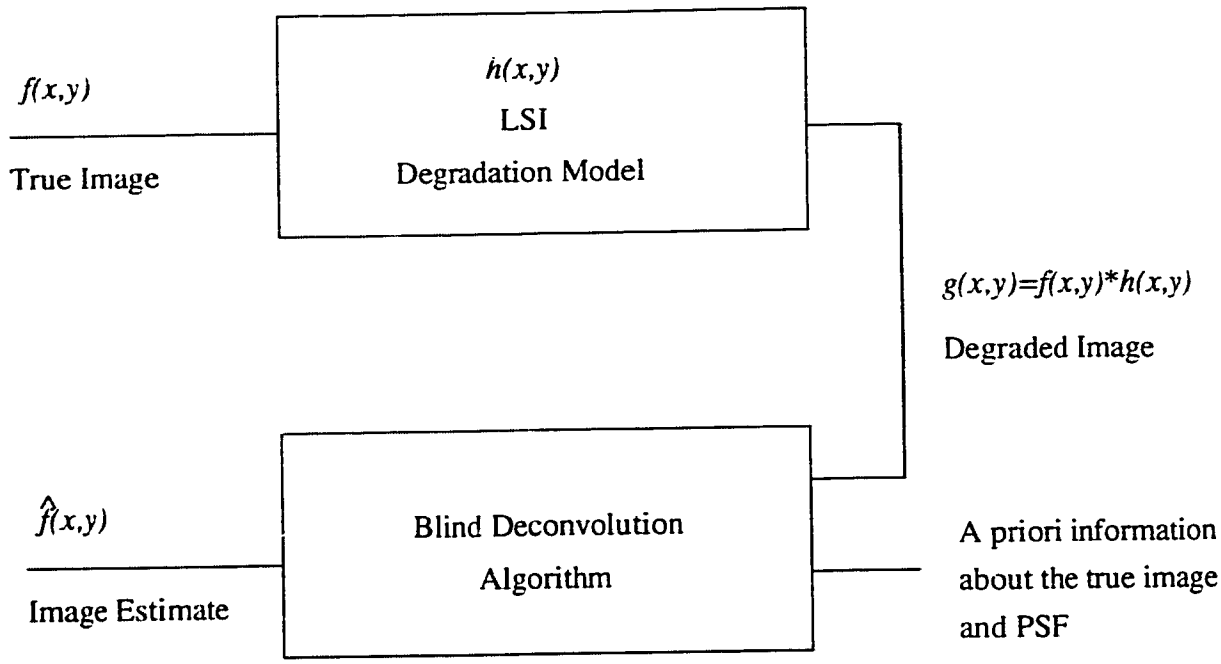

Figure 3.1: Blind Deconvolution Problem

$$\hat{f}(x,y) = Kf(x - D_x, y - D_y)$$

Experience shows that in general K, D_x, and D_y are not recoverable by blind deconvolution algorithms, however, this does not constitute a severe limitation. Figure 3.1 gives an overview of the blind deconvolution problem for images.

The partial information assumed about the true image and PSF distinguishes many of the existing blind deconvolution methods from one another. In this thesis, the following assumptions are made about the imaging process, true image, and PSF.

- The degradation is described by the linear model of equation 1.3.
- The imaging is performed such that the object is entirely encompassed by the observed frame.
- The background of the image is uniformly grey, black or white. This often occurs in such diverse applications as magnetic resonance imaging, fluorescence microscopy, and astronomical speckle imaging.

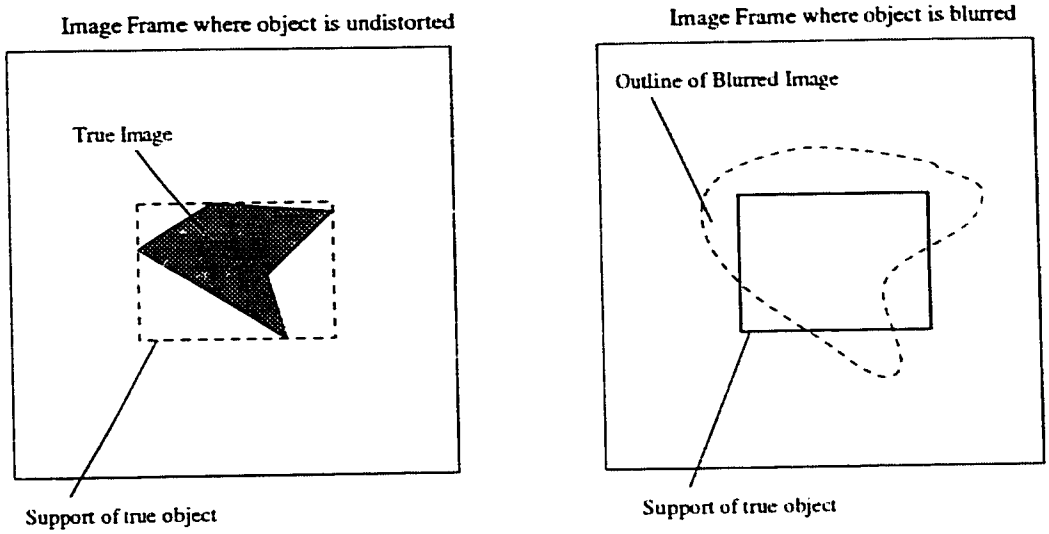

Figure 3.2: Example of a Finite Support Image

- The true image is nonnegative and its support is known a priori. Figure 3.2 illustrates what is meant by the region of support.
- The true image and PSF are *irreducible*[1].
- The inverse of the PSF exists and both the PSF and its inverse are absolutely summable. That is,

$$\sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} |h(x,y)| < \infty$$

and

$$\sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} |h^{-1}(x,y)| < \infty$$

No other constraints are imposed for the PSF. If the actual support of the true image is unknown, a novel support-finding algorithm, described in chapter 5, is used to determine the extent of the object.

Constraints of nonnegativity and support have been used in non-blind restoration problems to improve resolution of gamma-ray spectra [37], [38]. Evidence exists that nonnegativity and support information can extrapolate high frequency components

[1] An irreducible signal is one which cannot be expressed as the convolution of two or more component signals, on the understanding that the two-dimensional delta function is not a component image

lost when distortion is bandlimiting [39]. Therefore, such constraints hold promise in blind image restoration.

The problem is that of finding an estimate of the image $\hat{f}(x,y)$, given $g(x,y)$ by minimizing a criterion that incorporates knowledge of the support and nonnegativity of the true image. A solution which globally minimizes the criterion is termed a *feasible* solution.

3.2 Proposed Approach: Recursive Two-Dimensional Equalization

As described in Chapter 2, existing blind deconvolution techniques for images exhibit a poor compromise among computational complexity, reliability of solution, robustness to noise, and portability. They are, therefore, unsuitable for practical imaging applications.

The objectives of a new blind deconvolution scheme are to have low computational complexity, yet produce reliable results. The algorithm should exhibit improved convergence properties over existing techniques and should provide meaningful solutions in the presence of additive noise.

The proposed NAS-RIF algorithm belongs to the class of nonparametric restoration techniques described in section 2.2. In this class of methods, the object is assumed to be imaged completely within the frame. In addition, the true image and blur are assumed to be nonnegative with known support. Similarly, in the NAS-RIF method, the same assumptions are made about the true image; in contrast the proposed method does not require that the PSF be finite in extent with known dimensions. Instead, the PSF is required to have an inverse. Chapter 6 compares the performance of the proposed technique with the existing techniques of this class.

3.2.1 General Introduction

The novel algorithm presented is globally structured. Globally structured algorithms refer to a set of linear image restoration techniques which use information from a wide data region, if not the entire data set, for restoration of a single pixel [3]. This approach is useful in blind restoration because of the importance of maximizing the use of what little information is available. An iterative approach is chosen to perform blind deconvolution because such techniques do not require statistical knowledge of the image and/or blur. In addition, they are flexible in incorporating a priori information, are free from causality restrictions, and have shown to restore high quality images without degrading detail for situations where the PSF is known [40].

Most of the existing iterative techniques can be grouped into two subgroups: 1) least squares estimation (LSE), which entails the minimization of a squared-error criterion [40], and 2) projection onto convex sets [41],[42]. The latter, projection onto convex sets (POCS), identifies convex sets lying in image parameter space (usually $\mathbf{R}^N$, where N is the number of pixels in the image estimate), corresponding to a priori knowledge about the noise and true image. A given convex set represents the set of all possible image estimates for a corresponding assumption about the image or noise. At each iteration, an estimate of the image is successively projected on every convex set to obtain an updated estimate. Convergence of the procedure is guaranteed by the convexity of the sets under consideration. If there is a large amount of information available, the solution is restricted to lie in a small set containing the desired signal. If the information is sparse, however, the solution set becomes too large to produce a reasonable reconstruction. In this present form, POCS is not suitable for blind image restoration.

A more appropriate method for blind image restoration is least squares estimation. To apply this approach an error criterion or cost function must be chosen to reliably evaluate how appropriate an image estimate is. The proposed cost function is

$$J = \sum_{(x,y)\in D_{neg}} \hat{f}^2(x,y) + \sum_{(x,y)\in \overline{D_{sup}}} (\hat{f}(x,y) - L_B)^2 \tag{3.1}$$

where D_{sup} is the set of all pixel locations of $\hat{f}(x,y)$ within the region of support, D_{neg} is the set of all pixel locations within D_{sup} for which $\hat{f}(x,y)$ is negative, and $\overline{D_{sup}}$ is the set of all pixel locations of $\hat{f}(x,y)$ outside the region of support. Thus, $\overline{D_{sup}}$ is the complement of D_{sup}. L_B is the average pixel value of the background. Since any scaled version of the true image is acceptable, L_B can be assigned to any arbitrary non-zero grey level value if the background is not black[2]. The case where $L_B = 0$ is discussed in Chapter 4.

A major limitation of the existing techniques in the class of nonparametric finite support methods is that the cost functions used for minimization are not convex. The proposed cost function of equation 3.1 is, however, convex[3], which results in much more reliable restoration.

In fact, the proposed blind deconvolution method for images is motivated by the blind equalization algorithms for communication channels that use convex cost functions [43].

3.2.2 Proposed Blind Deconvolution Scheme for Images

The problem is that of constructing a reliable estimate of the true image from the received blurred image. For the remainder of this thesis, $\hat{f}(x,y)$ will denote the image estimate, $f(x,y)$ the true image, $h(x,y)$ the PSF, and $g(x,y)$ the degraded image. The image estimate $\hat{f}(x,y)$ is updated using the cost function of equation 3.1. The following scheme is developed.

The scheme is similar to that of training a solution in system identification. The difference is that a training signal is not available for blind deconvolution; therefore, a method of supplying some sort of training signal is required. In the proposed method a nonlinear transformation of $\hat{f}(x,y)$ is used.

The received blurred image $g(x,y)$ is filtered by a variable FIR filter $u(x,y)$ to produce an estimate of the image $\hat{f}(x,y)$. This estimate is passed through a nonlinear filter which uses a non-expansive mapping to project the estimate onto the convex set

[2]In this context black corresponds to a pixel value of zero.
[3]The proof of convexity is presented in section 3.3.

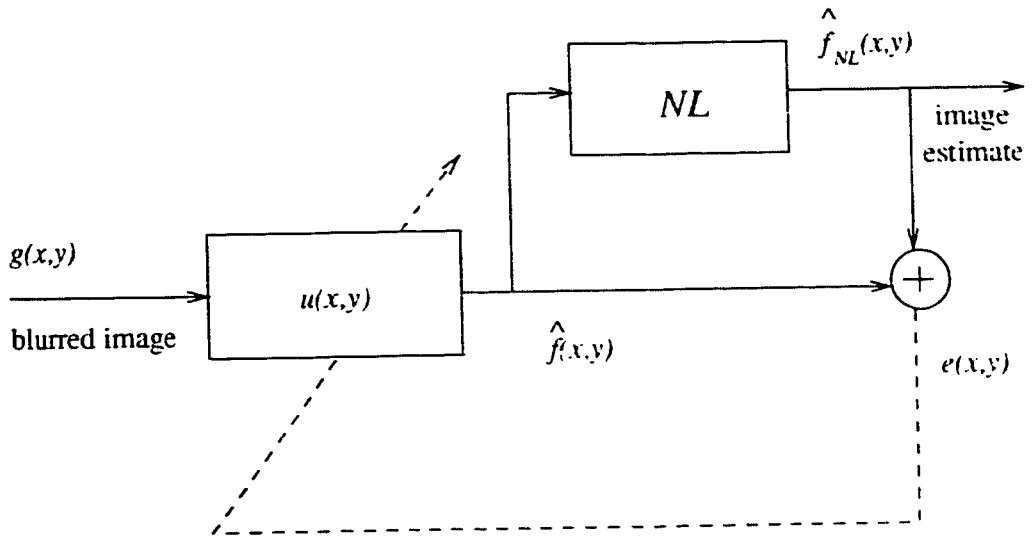

Figure 3.3: Proposed Blind Deconvolution Scheme for Images

corresponding to the known characteristics of the true image. The difference between this projected image, $\hat{f}_{NL}(x,y)$, and $\hat{f}(x,y)$ is used as the error signal to update the variable filter coefficients. Figure 3.3 provides an overview of the scheme.

The NL block represents the non-expansive mapping. For the particular assumptions made about the true image in this thesis, it represents the projection of the estimated image into the set of images that are nonnegative with given finite support. This entails forcing the negative pixel values within the region of support to zero and all pixel values outside the region of support to L_B.

The cost function in terms of its parameters $\{u(x,y)\}$ is presented below. The image estimate $\hat{f}(x,y)$ in equation 3.1 is replaced by its convolved components $u(x,y)$ and $g(x,y)$.

$$J(u(1,1),u(1,2),\ldots,u(N_{xu},N_{yu})) =$$
$$\sum_{(x,y)\in D_{neg}} \left[\sum_{x'=1}^{N_{xu}}\sum_{y'=1}^{N_{yu}} u(x',y')g(x-x'+1,y-y'+1)\right]^2 +$$
$$\sum_{(x,y)\in \overline{D_{sup}}} \left[\left(\sum_{x'=1}^{N_{xu}}\sum_{y'=1}^{N_{yu}} u(x',y')g(x-x'+1,y-y'+1)\right) - L_B\right]^2$$

The pixel indices of $u(x,y)$ and $g(x,y)$ range from $(1,1)$ to (N_{xu}, N_{yu}) and (N_{xg}, N_{yg}), respectively. For compactness, the form of equation 3.1 will used when referring to the cost function; however, its dependence on $u(x,y)$ should be recognized. The next section introduces the concept of convexity and proves that the proposed cost function is convex.

3.3 Convexity of the Proposed Cost Function

3.3.1 Definition and Implications of Convexity

The convergence properties of a blind deconvolution algorithm are important since they dictate to a large extent the reliability of the restoration technique. Convexity of the cost function implies that no local maxima or minima exist. This is highly desirable because it suggests that the global minimum of the cost function can be attained using one of a variety of numerical optimization algorithms from any arbitrary parameter initialization. Convergence to the feasible set of solutions is guaranteed.

Before convexity of a function is defined, a preliminary concept is introduced.

Definition 3.1 (Convex Set) *A set $C \subset \mathbf{R}^N$ is said to be* convex *if $\alpha x + (1-\alpha)x'$ is in C whenever x and x' are in C, and $\alpha \in]0,1[$.*

Geometrically, this implies that the line-segment

$$[x, x'] := \{\alpha x + (1-\alpha)x' : 0 \leq \alpha \leq 1\}$$

is entirely contained in C whenever its endpoints x and x' are in C.

A *convex* function can be defined using the definition of a convex set.

Definition 3.2 (Convex Function) *Let C be a nonempty convex set in $\mathbf{R}^N$. A function $f : C \to \mathbf{R}$ is said to be* convex *on C when, for all pairs $(x, x') \in C \times C$ and all $\alpha \in]0,1[$, there holds*

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x') \tag{3.2}$$

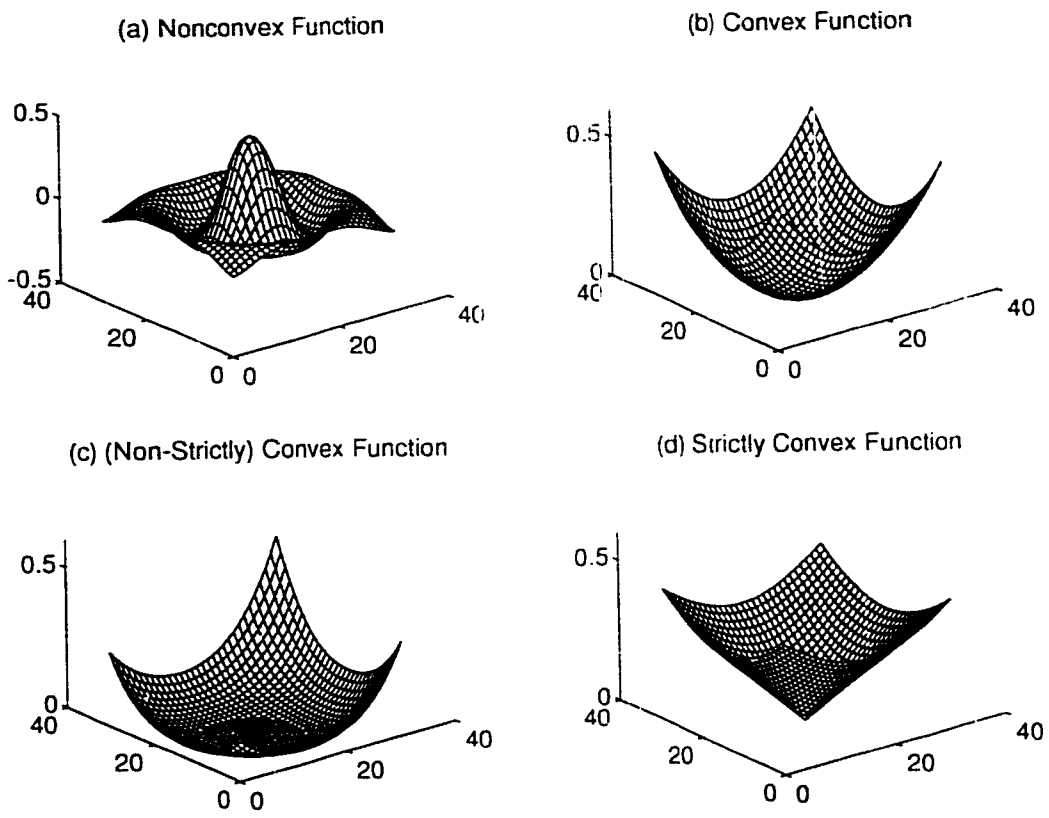

Figure 3.4: Examples of Convex/Nonconvex Functions

We say that f is *strictly convex* on C when the relation of 3.2 holds as a strict inequality for $x \neq x'$. The distinction between convexity and strict-sense convexity is important because it determines the uniqueness of the solution of a minimization routine. Convexity implies that there exists a set of parameter values for which the cost function has a global minimum, whereas strict-sense convexity guarantees unimodality (unique global minimum) of the cost function. Figure 3.4 provides examples of nonconvex, convex and strictly convex functions in $\mathbf{R}^2$.

When dealing with convexity, it is convenient to consider a function f as being defined on the whole space $\mathbf{R}^N$, by allowing the value of $+\infty$ for $f(x)$. Classically, convexity involved a pair (C, f), where C is a nonempty convex set and f is a function from C to $\mathbf{R}$ satisfying inequality 3.2 on C.

Such an f can be extended beyond C via the function

$$f_e(x) := \begin{cases} f(x) & \text{for } x \in C \\ +\infty & \text{for } x \notin C \end{cases}$$

This *extended-value* function f_e sends $\mathbf{R}^n$ to the set $\mathbf{R} \cup \{+\infty\}$. In the remaining chapters all (potentially) convex functions are considered to be of extended-value. The subscript "e" is therefore dropped and definition 3.2 is accordingly replaced by the following definition of convexity [44].

Definition 3.3 (Convex Function) *A function $f : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$, not identically equal to $+\infty$ is said to be convex when, for all $(x, x') \in \mathbf{R}^N \times \mathbf{R}^N$ and all $\alpha \in]0, 1[$, there holds*

$$f(\alpha x + (1-\alpha)x') \leq \alpha f(x) + (1-\alpha) f(x')$$

considered as an inequality in $\mathbf{R} \cup \{+\infty\}$.

The class of such functions is denoted by $Conv\mathbf{R}^N$. The remaining theorems stated in this thesis will refer to the above meaning of $Conv\mathbf{R}^N$.

3.3.2 Proof of Convexity

A useful criterion to determine the convexity of a function entails the function's *Hessian matrix* which incorporates the second partial derivatives of f. The Hessian matrix is defined below for a function of N parameters, $\{x_1, x_2, \ldots, x_N\}$.

Definition 3.4 (Hessian of a function) *The Hessian of a function $f : \mathbf{R}^N \to \mathbf{R}$ is defined as*

$$\nabla^2 f(x_1, x_2, \ldots, x_N) \triangleq \begin{bmatrix} \frac{\partial^2 f(x_1,\ldots,x_N)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x_1,\ldots,x_N)}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x_1,\ldots,x_N)}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f(x_1,\ldots,x_N)}{\partial x_N^2} \end{bmatrix}$$

Here, $f(x)$ where $x \in \mathbf{R}^N$ is replaced with $f(x_1, \ldots, x_N)$ where $x_i \in \mathbf{R}$ for $i = 1, 2, \ldots, N$ for clarity. Another important definition is that of an *open set*.

Definition 3.5 (Open Set) *A set of points S is open if every point that is a member of S has a neighbourhood completely in the set S.*

For example, in a normed space W, a neighbourhood about a point w of radius ε is defined as

$$B_\varepsilon(w) := \{z \in W : \| w - z \| < \varepsilon\}$$

Thus, a set S in a normed space W is open if for every $s \in S$ there exists an $\varepsilon > 0$ such that $B_\varepsilon(s)$ is contained in S. The set $\mathbf{R}^N$ is open.

To prove convexity of a given function, the following theorem is used [44]:

Theorem 3.1 *Let f be twice differentiable on an open convex set $\Omega \subset \mathbf{R}^N$. Then*

1. *f is convex on Ω iff $\nabla^2 f(x_0)$ is positive semi-definite for all $x_0 \in \Omega$*

2. *if $\nabla^2 f(x_0)$ is positive definite for all $x_0 \in \Omega$, then f is strictly convex on Ω.*

To apply theorem 3.1, we must determine if the cost function is twice differentiable over $\mathbf{R}^N$, the space spanned by the associated parameters.

Recalling that $\hat{f}(x,y) = u(x,y) * g(x,y)$, the proposed cost function of equation 3.1 can be represented as

$$J = \sum_{(x,y)\in D_{sup}} \hat{f}^2(x,y) \left[\frac{1 - \mathrm{sgn}(\hat{f}(x,y))}{2}\right] + \sum_{(x,y)\in \overline{D_{sup}}} [\hat{f}(x,y) - L_B]^2 \qquad (3.3)$$

where the following definition for $\mathrm{sgn}(\cdot)$ is used:

$$\mathrm{sgn}(f) := \begin{cases} 1 & \text{if } f \geq 0 \\ -1 & \text{if } f < 0 \end{cases} \qquad (3.4)$$

The difference between equations 3.1 and 3.3 is that the set of summation indices elements D_{neg} in equation 3.1 is replaced with D_{sup}. To compensate, a nonlinearity is introduced in the corresponding summation argument. Equation 3.3 is easier to analyze because the summation indices elements are fixed and do not depend on the

parameters $\{u(x,y)\}$. J can be broken into components J_1 and J_2 as defined below.

$$J_1 = \sum_{(x,y)\in D_{sup}} \hat{f}^2(x,y)\left[\frac{1-\mathrm{sgn}(\hat{f}(x,y))}{2}\right] \tag{3.5}$$

$$J_2 = \sum_{(x,y)\in \overline{D_{sup}}} [\hat{f}(x,y) - L_B]^2 \tag{3.6}$$

Convexity of J may be proven by determining the individual convexities of J_1 and J_2 and using the following theorem [44]:

Theorem 3.2 *Let $f_1, \ldots f_m$ be in $Conv\mathbf{R}^N$, $t_1, \ldots, t_m$ be positive numbers, and assume that there is a point where all the f_js are finite. Then the function*

$$f := \sum_{j=1}^{m} t_j f_j$$

is in $Conv\mathbf{R}^N$.

The presence of the sgn$(\cdot)$ function reveals that J_1 is not twice-differentiable, so theorem 3.1 cannot be directly applied to prove its convexity. Appendix A provides a detailed proof of the convexities of J_1 and J_2. Given that they are convex, it follows from theorem 3.1 that J is convex.

3.4 Uniqueness of the Global Minimum

In the preceding section, no reference to the strict-sense convexity of the proposed cost function was made. This is an important characteristic because it guarantees the uniqueness of the global minimum [44]. From theorem 3.1, part 2, we see that a function is strictly convex if its Hessian is positive definite. Appendix A shows that $\nabla^2 J$ is in general positive semidefinite. The conditions under which $\nabla^2 J$ is positive definite for all possible parameter settings of $u(x,y)$ are sufficient for the solution of the proposed NAS-RIF algorithm to be unique.

Since $\nabla^2 J$ can be written as the sum of $\nabla^2 J_1$ and $\nabla^2 J_2$, it is sufficient for $\nabla^2 J$ to be positive definite if either $\nabla^2 J_1$ or $\nabla^2 J_2$ is positive definite. Therefore, the conditions

on $g(x,y)$ under which either $\nabla^2 J_1$ or $\nabla^2 J_2$ is positive definite are investigated to determine situations for which J is unimodal.

The following theorem provides a method of distinguishing a matrix that is positive definite from one that is positive semidefinite [45].

Theorem 3.3 *A real symmetric matrix A is positive definite (positive semi-definite) if and only if there exists a nonsingular matrix B (a possibly singular matrix B) such that $A = BB^T$*

Theorem 3.3 is useful because a positive semidefinite matrix can be written as

$$A = BB^T = \sum_{i=1}^{N_B} b_i b_i^T \tag{3.7}$$

where N_B is the number of columns in B and b_i is the ith column vector of B.

For A to be positive definite, B must be full rank, ie., $\{b_i\}$ should be linearly independent. Therefore, a general $N \times N$ positive definite matrix can be represented as the sum of single rank matrices $p_i p_i^T$, where N of the column vectors $\{p_i\}$ are linearly independent.

The Hessian matrix of J can be written as

$$\begin{aligned} \nabla^2 J &= \nabla^2 J_1 + \nabla^2 J_2 \\ &= \lim_{r \to \infty} \sum_{(x,y) \in D_{sup}} A_r(x,y) g_{xy} g_{xy}^T + 2 \sum_{(x,y) \in \overline{D_{sup}}} g_{xy} g_{xy}^T \end{aligned} \tag{3.8}$$

$$\text{where } g_{xy} \triangleq \begin{bmatrix} g(x,y) \\ g(x,y-1) \\ \vdots \\ g(x-N_{xu}+1, y-N_{yu}+1) \end{bmatrix}$$

The coefficient $A_r(x,y)$ is a function of the inverse filter parameters $\{u(x,y)\}$ and is defined in equation A.2 of Appendix A. The parameter r is the regularization constant also discussed in Appendix A.

To evaluate the conditions for which $\nabla^2 J_1$ is positive definite, the dependence of $\lim_{r\to\infty} A_r(x,y)$ on $u(x,y)$ must be taken into account. From Appendix A,

$$\lim_{r\to\infty} A_r(x,y) = 0 \text{ for all } (x,y) \in D_{sup}$$

if $\hat{f}(x,y) \geq 0$ for all $(x,y) \in D_{sup}$. Since the grey-scale blurred image $g(x,y)$ is an intensity distribution, its pixels are nonnegative. Any all-positive setting of $u(x,y)$ results in $\hat{f}(x,y) \geq 0$ for all (x,y), which implies $\lim_{r\to\infty} A_r(x,y) = 0$ for all $(x,y) \in D_{sup}$. According to theorem 3.3, $\nabla^2 J_1$ is in general positive semidefinite, but not positive definite, for all nonnegative blurred images $g(x,y)$. Therefore, if $\nabla^2 J$ is to be positive definite, $\nabla^2 J_2$ must be positive definite.

From theorem 3.3, it is necessary for the number of elements in $\overline{D_{sup}}$ denoted by $\| \overline{D_{sup}} \|$ to be greater or equal to $N_{xu}N_{yu}$, for $\nabla^2 J_2$ to be positive definite. Allowing $N_{xg} \times N_{yg}$, $L_{xf} \times L_{yf}$, and $N_{xf} \times N_{yf}$ to represent the dimensions of the blurred image, dimensions of the support of the true image, and dimensions of the restored image, respectively, the inequality $\| \overline{D_{sup}} \| \geq N_{xu}N_{yu}$ is proven.

$$\begin{aligned} \| \overline{D_{sup}} \| &= N_{xf}N_{yf} - L_{xf}L_{yf} > N_{xf}N_{yf} - N_{xg}N_{yg} \\ &= N_{xu}N_{yu} + (N_{xu}-1)(N_{yg}-1) + (N_{yu}-1)(N_{xg}-1) - 1 \\ &> N_{xu}N_{yu} - 1 \end{aligned}$$

where the relations $N_{xf} = N_{xg} + N_{xu} - 1$ and $N_{yf} = N_{yg} + N_{yu} - 1$ are used.

$$\therefore \| \overline{D_{sup}} \| \geq N_{xu}N_{yu} \tag{3.9}$$

Figure 3.5 shows the region corresponding to $\overline{D_{sup}}$ in terms of the other parameters and size of the blurred image. Regardless of the number of elements in the equalizer, $\| \overline{D_{sup}} \|$ will always contain at least $N_{xu}N_{yu}$ elements. From theorem 3.3, it follows that $\nabla^2 J_2$ is positive definite if there are $N_{xu}N_{yu}$ linearly independent vectors g_{xy} for $(x,y) \in \overline{D_{sup}}$. For any fixed inverse filter dimensions, the rank of $\nabla^2 J_2$ depends on the elements of $g(x,y)$.

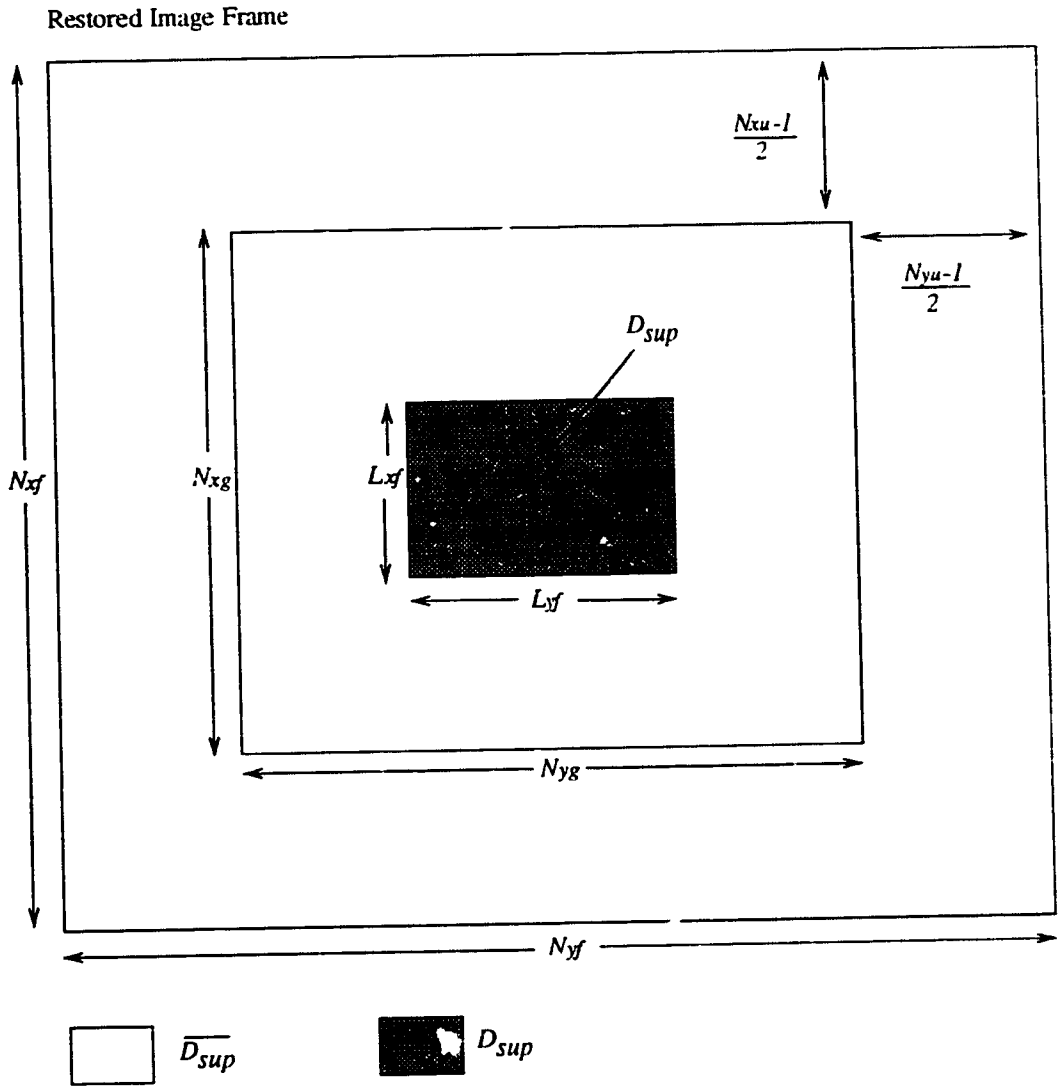

Figure 3.5: Region of Support of the True Image

Thus, a sufficient condition for the solution to the proposed algorithm to be unique is that there exists $N_{xu}N_{yu}$ linearly independent vectors g_{xy} for $(x,y) \in \overline{D_{sup}}$. Linear dependence of the g_{xy} vectors occur if $g(x,y)$ is composed solely of exponentials sinusoids, or damped sinusoids [46]. Experience shows that for practical images, the solution to proposed algorithm is almost always unique. For the simulations conducted, the algorithm always converged to the true image.

Under ideal situations, of infinite extent filter $u(x,y)$ and absence of noise, the solution is not unique. For example, if the blur is smaller than the image, there are infinite number of taps, the background of the true image is black and the true image is also invertible, a cost of zero can be obtained if the inverse filter represents the inverse of the true image or the inverse of the PSF. Thus, the solution of the algorithm is not unique. Many erroneous intermediate solutions can also exist. A non-unique solution also occurs if the image is reducible, that is, it can be represented by the convolution of two or more components. In such a situation, the result can be any of the convolutional components of the true image, and intermediate solutions. The possibility of these erroneous solutions is one of the dilemmas of blind deconvolution algorithms. With the lack of sufficient information, it is difficult to overcome this problem, so such solutions are acceptable for blind deconvolution purposes.

3.5 Optimization Algorithms

Various methods exist in the literature for the minimization of convex cost functions [44], [47]. Since the proposed cost function is nonlinear, its minimizer is difficult to determine analytically, so numerical techniques are used. A variety of numerical algorithms are available for unconstrained optimization of convex functions [48]. The methods differ in such aspects as the assumptions made about the structure of the cost function, necessity of the gradient of the cost, and storage requirements.

An optimization algorithm which makes efficient use of available information is desired. Since the gradient of J is easily computed, the class of *descent algorithms* are considered. They are popular for their simplicity to implement and understand.

Table 3.1: Summary of a general descent routine

- An initial guess $u_0 \in \mathbf{R}^N$ and tolerance $\delta > 0$ are set.
- Begin loop; set $k = 0$.

  1) (Test stopping criterion) If $J(u_k) \leq \delta$, stop.

  2) (Determine the direction of descent) With the help of a model of the problem around u_k, find $d_k \in \mathbf{R}^N$ for which $\exists t > 0$ such that $J(u_k + td_k) < J(u_k)$

  3) (Perform line-search). Evaluate $J(u_k + td_k)$ for various values of t; find a convenient $t_k > 0$ satisfying in particular
  $$J(u_k + t_k d_k) < J(u_k)$$

  4) (Update solution estimate). Set $u_{k+1} = u_k + t_k d_k$; replace k by $k + 1$ and loop to 1).

In such methods, a sequence $\{u_k\}$ is constructed such that

$$u_k \rightarrow \overline{u} \quad \text{where } J(\overline{u}) \leq J(u) \text{ for all } u \in \mathbf{R}^N$$

or at least

$$\nabla J(u_k) \rightarrow 0 \text{ when } k \rightarrow \infty \tag{3.10}$$

The term *descent* is used because J is forced to decrease at each iteration. A minimization algorithm is said to be *convergent* if relation 3.10 holds. Table 3.1 gives a summary of the general descent routine. The variable J represents the cost, u_k is the estimate of the solution at the kth iteration, d_k is the direction taken by the line-search routine to obtain u_{k+1}, and t_k is the current step-size. The quantities $J(u)$ and $\nabla J(u)$ are assumed to exist and be computable.

Two different classes of descent algorithms are investigated: steepest descent methods and conjugate gradient methods. They differ from one another in the selection of d_k.

3.5.1 Steepest Descent Methods

The most popular descent routine is the steepest-descent method. It is commonly used because it is straightforward to understand and implement. In this method, the

direction of steepest descent of J at the current location u_k is selected as the direction to travel in order to minimize the cost J. Specifically, a direction is chosen such that

$$\min_d \{< \nabla J(u_k), d >: |||d||| = |||\nabla J(u_k)|||\} \tag{3.11}$$

where d is the direction of travel, $< \cdot, \cdot >$ is a predefined inner product defining ∇J and a predefined norm $|||d|||$, not necessarily $< \cdot, \cdot >^{1/2}$. Minimization of equation 3.11 corresponds to the *steepest-descent direction*. The choice of $||| \cdot |||$ has little influence on whether u_k converges or not to the minimizer $\overline{u}$. However, it does influence the speed of convergence as discussed in [44]. For the situation where $||| \cdot ||| = \| \cdot \|$ induced by a scalar product $< \cdot, \cdot >$ defining ∇J, the steepest descent algorithm becomes the well-known *gradient* algorithm. The convergence of such algorithms are in general extremely slow compared to the conjugate gradient method, and no indication of the optimal step-size t_k is given.

Table 3.2 gives a summary of the NAS-RIF method using the steepest descent gradient algorithm for minimization of the proposed cost function of equation 3.1. The gradient $\nabla J(u_k)$ is an $N_{xu}N_{yu} \times 1$ column vector and is computed using the following equation

$$\begin{aligned} [\nabla J(u_k)]_{j+(i-1)N_{xu},1} &= \frac{\partial J(u_k)}{\partial u(i,j)} \\ &= 2 \sum_{(x,y)\in D_{neg}} \hat{f}_k(x,y) g(x-i+1, y-j+1) \\ &+ 2 \sum_{(x,y)\in \overline{D_{sup}}} [\hat{f}_k(x,y) - L_B] g(x-i+1, y-j+1) \end{aligned} \tag{3.12}$$

where $\hat{f}_k(x,y) = u_k(x,y) * g(x,y)$.

3.5.2 Conjugate Gradient Methods

The concept of conjugate directions has been introduced successfully in optimization theory, and was motivated by the need to accelerate the steepest descent methods [44]. One of the advantages of this method is that convergence in a finite number of iter-

Table 3.2: Summary of the NAS-RIF method using the steepest descent gradient algorithm for optimization

- Select the largest step-size t (experimentally) such that the minimization algorithm is stable.
- Set initial conditions ($k = 0$):

$$\begin{aligned} u_k^T &= [u_k(1,1), \ldots, u_k(N_{xu}+1)/2, (N_{yu}+1)/2, \ldots, u_k(N_{xu}, N_{yu})] \\ &= [0, \ldots, 1, \ldots, 0] \\ \delta &> 0 \text{ is set.} \end{aligned}$$

- At iteration (k): $k = 0, 1, 2, \ldots$
  1) If $J(u_k) \leq \delta$, stop.
  2) $d_k = -\nabla J(u_k)$
  3) $u_{k+1} = u_k + t_k d_k$

where $u_k \in \mathbf{R}^{N_{xu} N_{yu}}$ is the current estimate of the inverse filter, and $\nabla J(u_k)$ is calculated as in equation 3.12.

ations is guaranteed when a quadratic cost function is used and exact arithmetic is assumed. Even for nonquadratic costs, the method shows considerably increased convergence speed [49]. The algorithm is based on the premise that information about the curvature of J with each new iteration can accelerate convergence. The "best" direction to move is not always the instantaneously steepest direction.

The rationale behind the conjugate gradient methods is algebraic. The set of directions $\{d_k\}$ should include "non-interfering" directions with the property that minimization along one direction is not "spoiled" by subsequent minimization along another. When the cost function is quadratic, these directions are called *conjugate directions*, and are related by the expression [44]

$$< Qd_i, d_j >= 0 \quad \text{for } i \neq j$$

where Q is the cost function's Hessian matrix.

The cost function is, therefore, minimized by proceeding in directions conjugate

Table 3.3: Summary of the NAS-RIF method using the conjugate gradient algorithm for optimization

---

- Set initial conditions ($k = 0$):

$$\begin{aligned} u_k^T &= [u_k(1,1), \ldots, u_k(N_{xu}+1)/2, (N_{yu}+1)/2, \ldots, u_k(N_{xu}, N_{yu})] \\ &= [0, \ldots, 1, \ldots, 0] \\ \delta &> 0 \text{ is set.} \end{aligned}$$

- At iteration (k): $k = 0, 1, 2, \ldots$

  1) If $J(u_k) \leq \delta$, stop.

  2) If $k = 0$, $d_k = -\nabla J(u_k)$. Otherwise, $d_k = -\nabla J(u_k) + \beta_{k-1} d_{k-1}$

  3) Perform a line minimization to find t_k such that

$$J(u_k + t_k d_k) \leq J(u_k + t d_k) \quad \text{for } t \in \mathbf{R}$$

  4) $u_{k+1} = u_k + t d_k$

where $u_k \in \mathbf{R}^{N_{xu} N_{yu}}$ is the current estimate of the inverse filter, $\nabla J(u_k)$ is calculated as in equation 3.12, and β_k is given by equation 3.14.

---

to all previous directions traversed. From [44], these directions are related by

$$d_{k+1} = -\nabla J(u_{k+1}) + \beta_k d_k \tag{3.13}$$

where k is the iteration, and

$$\beta_k = \frac{< \nabla J(u_{k+1}) - \nabla J(u_k), \nabla J(u_{k+1}) >}{\| \nabla J(u_k) \|^2} \tag{3.14}$$

and is called the Polak-Ribière formula. There exists evidence [50] that this choice of β_k provides fast convergence for general non-quadratic cost functions, as well.

For cost functions that are not exact quadratics, one can "pretend" that J is quadratic, and hope that the directions of equation 3.13 are more efficient than the plane gradient used in the steepest descent methods. Table 3.3 gives a summary of the NAS-RIF methods using conjugate gradient minimization proposed cost function.

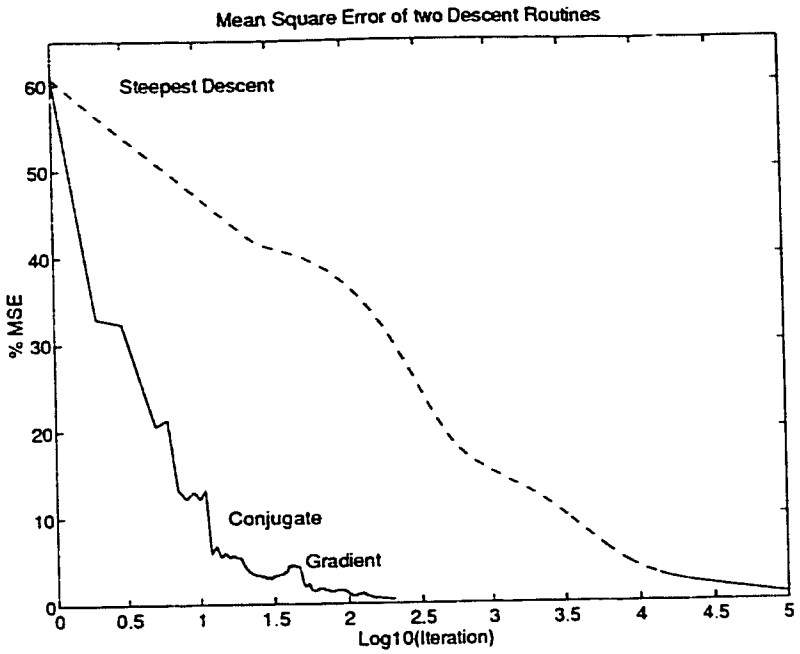

Figure 3.6: Convergence rates of the steepest descent and conjugate gradient methods

3.5.3 Comparison of Convergence Rates of Optimization Algorithms

The proposed cost function of equation 3.1 is composed of the two components presented below:

$$J_1 = \sum_{(x,y)\in D_{sup}} \hat{f}^2(x,y)\left[\frac{1-\mathrm{sgn}(\hat{f}(x,y))}{2}\right]$$

$$J_2 = \sum_{(x,y)\in \overline{D_{sup}}} [\hat{f}(x,y) - L_B]^2$$

J_2 is exactly quadratic, and J_1 is similar in form to a quadratic function, the difference being the presence of the sgn$(\cdot)$ function. Therefore, the conjugate gradient algorithm appears to be an intuitively more efficient choice for the optimization of $J_1 + J_2$, than the steepest descent algorithm. Experimentally, it does not exhibit the slow convergence of the steepest descent algorithm. Figure 3.6 compares the algorithms presented in tables 3.2, and 3.3 for a blurred image using a 13×13 equalizer.

The conjugate gradient algorithm converges approximately 3 orders of magnitude

faster than the steepest descent method. In all simulations conducted for this thesis, the conjugate gradient method exhibited much faster convergence than the steepest descent. The conjugate gradient method has been successfully used in iterative non-blind deconvolution algorithms [51], [52], [53], [54].

Chapter 4

Performance Improvement Techniques

Although convergence to the feasible set of solutions[1] is guaranteed with the proposed method, the question arises as to whether this solution is sufficiently close to the true image. In the development of the algorithm, two practical considerations were overlooked: a finite-dimensional equalizer must be implemented to approximate the (generally) infinite extent inverse PSF, and additive noise is present in the degraded image. This chapter addresses the effects of finite equalizer parameters and noise on the restoration algorithm. Performance improvement techniques are proposed to alleviate the resulting ill-effects.

4.1 Finite Equalizer Parameters

4.1.1 Effect of the Finite Equalizer on the Global Minimum

In general, the true inverse of a PSF is infinite in extent. Therefore, attempting to invert the effect of the blurring process using a two-dimensional FIR filter only approximates the true image. Convexity of the cost function is preserved in the finite

[1]The feasible set of solutions is defined as the set of all solutions which globally minimize the associated cost function of a problem.

parameter case, however, the question which naturally arises is whether the global minimum of the proposed cost function occurs sufficiently close to that for the infinite parameter case.

A measure of the reliability of the image estimate is derived by exploiting the continuity of the cost function with respect to parameters $u(x,y)$. Appendix B provides analysis showing that the global minimum of the cost function in the finite parameter case is close to the global minimum of the cost function in the infinite parameter case u^*. The following inequality is derived for noiseless conditions:

$$J(u^{(N_x,N_y)}) \leq J^* + \varepsilon^2 G_{max}^2 S_u^2 [4 \parallel D_{neg} \parallel + \parallel \overline{D_{sup}} \parallel]$$

The significant parameters are briefly defined below:

J^* is the global minimum of the cost function for an infinite extent equalizer.

$J(u^{(N_x,N_y)})$ represents the cost for the truncated inverse filter of dimension $(2N_x + 1) \times (2N_y + 1)$.

$u^*(x,y)$ represents the global minimizer of J, such that it is the true inverse PSF, and $J(u^*) = J^*$

$u^{(N_x,N_y)}$ is a truncated version of $u^*(x,y)$, expressed as

$$u^{(N_x,N_y)}(x,y) = \begin{cases} u^*(x,y) & \text{for } |x| \leq N_x \text{ and } |y| \leq N_y \\ 0 & \text{for } |x| > N_x \text{ or } |y| > N_y \end{cases}$$

ε is a positive constant such that

$$\sum_{|x| \geq N_x+1} \sum_{|y| \geq N_y+1} |u^*(x,y)| \leq \varepsilon \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} |u^*(x,y)| \tag{4.1}$$

G_{max} represents the maximum value of $|g(x,y)|$ for all (x,y).

$\parallel D_{neg} \parallel$ and $\parallel D_{sup} \parallel$ are the number of elements in D_{neg} and D_{sup}, respectively

and

$$S_u = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} |u^*(x,y)|$$

Appendix B provides detailed definitions.

The inequality of 4.1 shows that ε can be made sufficiently small by selecting (N_x, N_y) to be appropriately large. This suggests that $J(u^{(N_x,N_y)})$ can be made arbitrarily close to J^* through the selection of (N_x, N_y). Because of the continuity of J with respect to $u(x,y)$, the global minimum of J for finite extent parameters can be made "close" to $u^*(x,y)$. Thus, the restored image can be made arbitrarily close to the true image.

4.1.2 Possible All-Zero Global Minimum

For applications such as magnetic resonance imaging and astronomy, where the background is often black[2], an assignment of $L_B = 0$ must be made in equation 3.1. Subsequently, the cost function becomes:

$$J = \sum_{(x,y)\in D_{neg}} \hat{f}^2(x,y) + \sum_{(x,y)\in \overline{D_{sup}}} \hat{f}^2(x,y) \tag{4.2}$$

Intuitively, one can infer that for noiseless conditions, availability of an infinite extent equalizer and infinite precision, any nonnegatively scaled version of the inverse PSF $ku^*(x,y)$ will globally minimize the cost function. For this parameter setting,

$$\hat{f}(x,y) = ku^*(x,y) * g(x,y) = kf(x,y)$$

where $f(x,y)$ is the true image; the cost becomes

$$\begin{aligned} J^* &= \sum_{(x,y)\in D_{neg}} k^2 f^2(x,y) + \sum_{(x,y)\in \overline{D_{sup}}} k^2 f^2(x,y) \\ &= 0 \end{aligned}$$

[2]In this context black corresponds to a pixel value of 0.

because

$$kf(x,y) \geq 0 \Rightarrow D_{neg} = \{\emptyset\}$$

and

$$kf(x,y) = 0 \text{ for all} (x,y) \in \overline{D_{sup}}$$

for $k \geq 0$. Any positively scaled version of the true image is an acceptable solution to the minimization problem. Unfortunately, the all-zero solution which occurs for $k = 0$ must be avoided.

In practice, when the equalizer is finite in extent and noise exists, there is only one minimum and it is straightforward to see that it occurs when

$$\hat{f}(x,y) = 0 \Rightarrow u(x,y) = 0 \quad \text{for all } (x,y)$$

Thus, the cost function is uniquely globally minimized at the trivial all-zero solution. Unfortunately, the desired set of solutions lie on a ray corresponding to any positively scaled version of the PSF inverse, where the cost is nonzero, but small. Figure 4.1 shows the surface plot of the cost function for two-dimensional parameters. The ray of desired solutions resembles a valley.

The existing nonparametric finite image support restoration techniques concentrate on the situation in which the background is black, and it is therefore important to address this practical limitation of the proposed algorithm. The next section discusses techniques which can be used to successfully constrain the parameters away from the trivial all-zero solution.

4.1.3 Methods of Constraining Parameters from the Trivial Solution

To constrain the global minimum in equation 4.2 away from the trivial solution, two issues must be addressed:

1. What type of constraint should be imposed on the parameters of the equalizer?

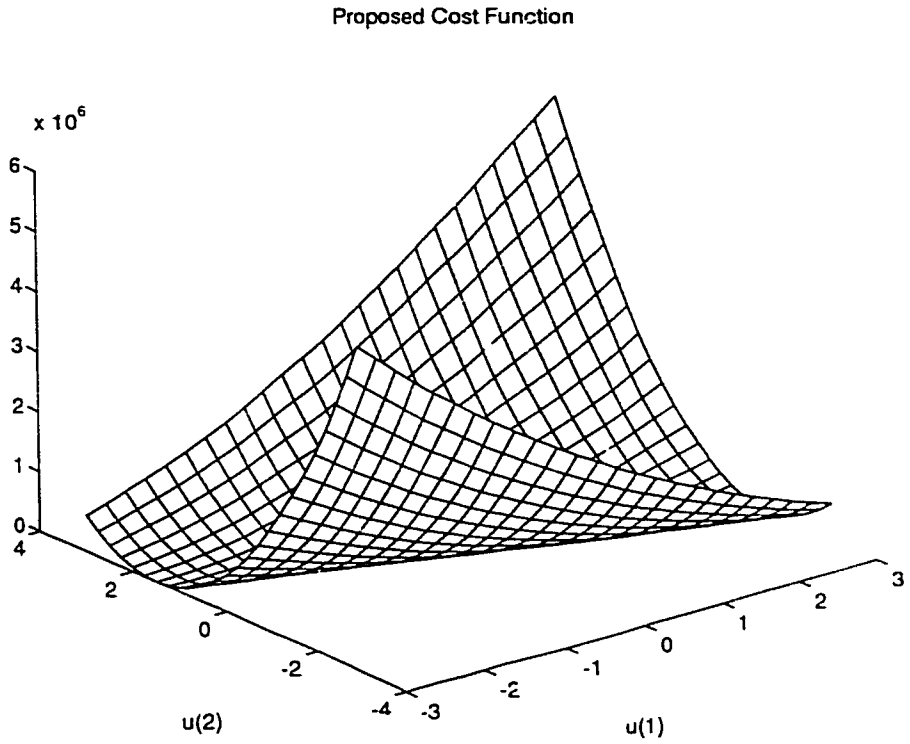

Figure 4.1: Proposed cost function in two-dimensions for $L_B = 0$.

2. How should the constraint be incorporated into the optimization routine?

Selection of Constraint

The trivial solution can be avoided by constraining $\{u(x,y)\}$ to a set excluding the origin.

Unfortunately, there is no guarantee that the cost function is convex under an arbitrarily constrained parameter space. The following theorem provides insight [44].

Theorem 4.1 *Let C_1 and C_2 be nonempty convex sets in $\mathbf{R}^N$, such that $C_2 \subseteq C_1$. A convex function f on C_1 is also convex on C_2.*

If the constrained set of possible parameters forms a convex set, then convexity of J is preserved. Therefore, the problem becomes that of constraining the parameters $\{u(x,y)\}$ to a convex set, excluding the origin, yet including a desirable solution. One

method of accomplishing this is to constrain the parameters using a linear constraint

$$\underline{a}^T \underline{u} = 1 \tag{4.3}$$

where

$$\underline{a}^T = [a(1,1)\ a(1,2) \cdots a(N_{xu}, N_{yu})]$$

$$\underline{u}^T = [u(1,1)\ u(1,2) \cdots u(N_{xu}, N_{yu})]$$

are chosen to exclude the origin and include a desirable solution. Equation 4.3 forms a convex set with respect to the parameters $\{u(x,y)\}$.

To select $\underline{a}$, some additional information about the PSF or its inverse may be used.

Peak of the Inverse PSF For instance, if the inverse of the PSF is known to have a unique peak within a certain range $(x_M \pm \Delta, y_M \pm \Delta)$, then $u(x_M, y_M) = 1$ is used as a constraint. That is,

$$\underline{a} = \begin{cases} 1 & \text{if } (x,y) = (x_M, y_M) \\ 0 & \text{otherwise} \end{cases}$$

The assumed true image support must be made larger to provide room for error in the location of the peak. This location may be available in situations where the physics of the imaging system allows some assessment of the spatial concentration of the PSF. If this information is not available, other constraints based on the assumptions stated in section 3.1 may be applied.

Nonnegativity of the PSF For example, the nonnegativity of the PSF can be exploited. Since the PSF is nonnegative, its sum is a positive constant, and represents its discrete time Fourier transform (DTFT) coefficient $H(\omega)$ at $\omega = 0$. Similarly, the sum of the inverse PSF is also positive since its DTFT coefficient is the reciprocal of $H(\omega)$. In many situations, the imaging system is assumed to preserve the mean value of the true image, in which case $\sum h(x,y) = 1$, and subsequently, $\sum h^{-1}(x,y) = 1$. Therefore, the equalizer coefficients can be

constrained such that their sum is equal to one. That is

$$\underline{a}^T = [1\ 1\ 1 \cdots 1\ 1\ 1]$$

or $\{u(x,y)\}$ is constrained to the convex set

$$\sum_{x=1}^{N_{xu}} \sum_{y=1}^{N_{yu}} u(x,y) = 1 \tag{4.4}$$

Nonnegativity of the True Image If the linear degradation occurs from transmission, then the effective blurring is not necessarily nonnegative, and the sum of the blur pixels could potentially be negative. In this situation, equation 4.4 cannot be used, and the following constraint, which makes use of the knowledge that the true image is nonnegative, can be applied:

$$\sum_{(x,y)\in D_{sup}} \hat{f}(x,y) = 1$$

or, in terms of $\{u(x,y)\}$

$$\sum_{x=1}^{N_{xu}} \sum_{y=1}^{N_{yu}} u(x,y) \left[\sum_{(x',y')\in D_{sup}} g(x'-x, y'-y) \right] = 1 \tag{4.5}$$

The constraint of equation 4.4 is preferred to that of equation 4.5 because of the added computational complexity of incorporating equation 4.5. Both constraints produce comparable restorations in simulations.

Method of Constraining Parameters

Many methods exist in the literature [44], [47] for constrained optimization. In this section, few of the methods used in the applications of iterative image restoration and blind equalization are discussed.

Projection Method The first method, commonly referred to as the *projection method* [39], [55] involves projecting the parameters onto the convex constraint set after

each parameter update of the descent routine. This is one of the simplest and computationally efficient methods of incorporating deterministic constraints. The application of constraints in this manner is not consistent with the conjugate direction methods, but has nonetheless been used in conjunction with them [51], [52]. Experimental results show that some computational power is wasted by projecting the parameters back to the constraint set after each iteration.

Parameter Rotation Method The second method, which has been proposed for blind equalization [43], only works for linear constraints. It involves rotating the set of parameters in the constraint set from $\{u(x,y) : \underline{a}^T\underline{u} = 1\}$ to $\{\alpha(x,y)\}$ such that

$$\underline{u} = \underline{u_0} + V\underline{\alpha}$$

where

$$\underline{\alpha}^T = [\alpha(1,1)\ \alpha(1,2)\cdots\alpha(N_{xu}, N_{yu})]$$

and $\underline{u_0}$ is any solution to the constraint $\underline{a}^T\underline{u} = 1$. The matrix V consists of columns (usually orthonormal) that span the space orthogonal to $\underline{a}$. The minimization must be accomplished with respect to $\{\alpha(x,y)\}$, so the cost function and gradient calculations require, in general, $N_{xu}N_{yu}$ more calculations on every call than originally.

Penalty Method The final method consists of adding a *penalty* term to the cost function to constrain the parameters. It is a common technique of converting a constrained optimization problem to an unconstrained one [56]. The new cost function becomes

$$J = \sum_{(x,y)\in D_{neg}} \hat{f}^2(x,y) + \sum_{(x,y)\in \overline{D_{sup}}} \hat{f}^2(x,y) + \gamma[\sum_{x=1}^{N_{xu}}\sum_{y=1}^{Nyu} a(x,y)u(x,y) - 1]^2$$

This approach does not necessarily constrain the parameters to the constraint set. Instead, it finds a compromise between minimizing the function and constraining the parameters. Because we require that the constraint only eliminate

the trivial all-zero solution, this is not a disadvantage of the proposed algorithm. In addition, the technique does not interfere with the conjugate direction methods nor does it require the computational expense of the parameter rotation method. Most of the simulations of Chapter 6 use the penalty method with $\gamma = 1$ in conjunction with the following constraint:

$$\sum_{x=1}^{N_{xu}} \sum_{y=1}^{N_{yu}} u(x,y) = 1$$

4.2 Effects of Noise

In practical situations, the blurred image is corrupted by additive noise in the degradation process. Common types of noise are:

Electronic noise resulting from the thermal motion of electrons in the electronic components of the imaging system.

Photoelectric noise due to the statistical nature of light and photoelectric conversion process in the image sensor.

Film grain noise from the randomness of silver halide grains in the film that records the image.

Quantization noise which occurs during image digitization.

In practice, exact restoration of the original scene from the observed data may not be possible, even with knowledge of the degrading system characteristics, because of the ill-posed nature of the image restoration problem by the the presence of observation noise. Image restoration methods must attempt to improve the image to be visually as close as possible to the original scene.

4.2.1 Ill-conditioning of the Blind Image Restoration Problem by the Presence of Noise

The *ill-posed* nature of a problem is observed by recognizing that a small perturbation of the given data produces large deviations in the resulting solution. Since the process of equalization attempts to restore the image by direct-inversion of the PSF, the problem is often ill-posed owing to the presence of additive noise. This follows because the direct inverse of the blur transfer function usually has a large magnitude at high frequencies, so excessive amplification of the noise at these frequencies results.

In [57], Nashed suggests that the classification of a problem as ill-posed depends on the notion of the solution, the admissible data, and the measure of continuous dependence. A problem is considered to be *well-posed* if for the set of possible admissible data a solution exists, is unique, and depends continuously on the data. Otherwise, the problem is *ill-posed*.

In section 3.4 the uniqueness of the solution of the proposed algorithm is shown to depend on the Hessian matrix presented in equation 3.8. Since the additive noise perturbs the pixels of the degraded image $g(x,y)$, the Hessian given in equation 3.8 can lose rank, thus loosing uniqueness of the solution. The problem can, therefore, be viewed as ill-posed. However, this is a pathological case that occurs with probability zero in practical situations.

Appendix C shows that the minimum of the support constraint cost function J_2 can be represented as a sum of terms, each inversely proportional to the singular values $\lambda_{m,n}$ related to the following matrix:

$$\underline{G} = \begin{bmatrix} g(1,1) & g(1,0) & \dots & g(-N_{xu}+2,-N_{yu}+2) \\ g(1,2) & g(1,1) & \dots & g(-N_{xu}+2,-N_{yu}+3) \\ \vdots & \vdots & & \vdots \\ g(N_{xf},N_{yf}) & g(N_{xf},N_{yf}+1) & \dots & g(N_{xg},N_{yg}) \end{bmatrix}$$

where $N_{xu} \times N_{yu}$ are the dimensions of the equalizer, $N_{xg} \times N_{yg}$ are the dimensions of the blurred image, and $N_{xf} \times N_{yf}$ are the dimensions of the restored image.

The global minimum of J_2 occurs at

$$u^* = \sum_{m=1}^{N_{xu}} \sum_{n=1}^{N_{yu}} \frac{\underline{v}_{m,n}^T \underline{L}\, \underline{w}_{m,n}}{\lambda_{m,n}}$$

where $\underline{v}_{m,n}$, $\underline{L}$, and $\underline{w}_{m,n}$ are $N_{xf}N_{yf} \times 1$, $N_{xf}N_{yf} \times 1$ and $N_{xu}N_{yu} \times 1$ real vectors, respectively, defined in Appendix C.

Any small perturbation of the elements of this matrix will slightly perturb the singular values, which in turn results in a potentially large perturbation of the global minimum, if $\lambda_{m,n}$ is small. Appendix C shows that on average, zero-mean additive white Gaussian noise (AWGN) has the effect of perturbing the singular values by a term proportional to the variance of the noise σ_n^2. Therefore, the solution is effected by noise to a greater degree if the ratio of $\lambda_{m,n}$ (in the noiseless situation) to σ_n^2 is small.

4.2.2 Effect of Noise on Cost Function

Analysis of the effect of noise on the cost function provides valuable insight into the behaviour of the NAS-RIF algorithm in practical situations. Because the cost function J is nonlinear, its global minimum in the presence of noise u^*_{noise} is difficult to characterize in terms of the minimum in the noiseless case u^*. However, because of the continuity of J with respect to $u(x,y)$, the value of the cost function in noisy conditions at parameter setting u^* is an effective indicator of the degree of bias introduced in the restored image. We present the results for the case of zero-mean stationary additive white Gaussian noise (AWGN) of variance σ_n^2 in this section. Appendix C provides a detailed analysis. The expected value of the cost at u^* assuming infinite extent parameters is

$$\begin{aligned} E\{J(u^*(x,y))\} &= \sum_{(x,y)\in D_{sup}} (f^2(x,y)+\sigma^2)\left(1 - Q\left(\frac{-f(x,y)}{\sqrt{2}\sigma}\right)\right) \\ &\quad - \sum_{(x,y)\in D_{sup}} \sigma \frac{f(x,y)}{\sqrt{2\pi}} e^{-f^2(x,y)/2\sigma^2} + \sigma^2 \parallel \overline{D_{sup}} \parallel \end{aligned} \tag{4.6}$$

$E\{J(u^*(x,y))\}$ represents the bias term resulting from the presence of noise, $f(x,y)$ is the true image, and $u^*(x,y)$ is the desired equalizer setting in the noiseless situation, $\| \overline{D_{sup}} \|$ is the number of elements in $\overline{D_{sup}}$, and

$$\sigma^2 = \sigma_n^2 \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} [u^*(x_1, y_1)]^2 \tag{4.7}$$

and $Q(x)$ is defined as

$$Q(s) \triangleq \frac{1}{\sqrt{2\pi}} \int_s^{\infty} e^{-x^2} dx$$

The global minimum of the cost function is biased by the noise. The bias is a function of the true image, the variance of the noise σ_n^2, and the energy of the optimal parameters u^*.

The first two terms on the right hand side of equation 4.6 correspond to biasing of the nonnegativity constraint, and the last term to the support constraint. The first two terms imply that the effect of noise is small if the quantity $\frac{f(x,y)}{\sigma}$ is large for all $(x,y) \in D_{sup}$. Intuitively, this makes sense because if most pixels of an image are highly positive, the noise added to it makes little difference to its nonnegativity properties. However, if there are many regions of black, the noise has more effect on the restoration with respect to nonnegativity. For the biasing term corresponding to the support constraint, a smaller σ^2 implies a smaller bias. The quantity σ^2 is proportional to the variance of the noise σ_n^2 as seen in equation 4.7. For a fixed noise level, one sees that the noise has less effect on the restoration if the energy of the inverse of the ideal PSF is small. However, recall that the imaging system is often assumed to preserve the mean value of the true image, in which case

$$\sum_{\forall(x,y)} u^*(x,y) = 1$$

Thus, for a fixed noise level, PSFs with inverses exhibiting smaller energy with the above constraint should produce better results in the presence of noise. In practical situations, such inverse PSFs would tend to exhibit less variance, and have more energy concentrated at the low frequency range. Unfortunately, experience shows

that most PSFs are low pass resulting in inverse PSFs with energy concentrated at the higher frequencies.

4.2.3 Forms of Regularization

Regularization encompasses various methods for dealing with the numerical instability inherent in ill-posed inverse problems. It encompasses a class of solution techniques which involve the study of an associated well-posed problem, provided that the analysis this new problem produces a meaningful solution for the original ill-posed problem. To obtain an acceptable solution, some strategy for controlling the numerical instability must be employed.

One technique is to modify the given system to improve its condition but maintain as much fidelity as possible. Some methods incorporate a stabilizing function to suppress the amplification of noise in the restored image [58]. Others impose constraints based on the statistics of the noise and a priori information about the image [59]. Regularized deconvolution techniques using information about the degradation process attempt to roll-off the transfer function of the inverse filter in an attempt to limit the noise amplification during the signal restoration process. Deterministic regularization techniques are usually implemented by iterative procedures to alleviate the computational problems.

Regularization can also take the form of terminating an iterative restoration procedure before it converges to the inverse solution [55], [60], [61], [62]. Nearly all concepts used in regularization are based on incorporating knowledge about either the true solution or the noise into the algorithm. In this sense the procedures used for truncating the number of iterations are called regularization as well [40].

4.2.4 Regularizing Stopping Criterion

Iterative restoration techniques attempt to enhance the high frequency components of the degraded image. Thus, the procedure amplifies whatever noise is present. As an iterative restoration process progresses, the error due to blurring decreases as the

error due to noise amplification increases. At some point in the algorithm, this total error reaches a minimum point and the procedure should be stopped before it reaches convergence. This phenomenon is explained and experimentally investigated in [40].

A drawback of the proposed NAS-RIF algorithm is that the convergence point does not necessarily produce the best estimate of the original image in the presence of noise. The iterative implementation of the inverse filter has the advantage that it can be terminated prior to convergence, resulting in a partially blurred image which will often exhibit less noise amplification. Previous research in the determination of the optimal termination point for iterative image restoration assumes that the PSF is known a priori. In [55], Trussell proposes terminating the iterative process when the average residual error is equivalent to the noise variance. The problem is that this terminates the process too quickly, because the restoration is biased by the noise. Sullivan and Katsaggelos use singular value decomposition in combination with a spectral filter function to evaluate a measure used to decide termination at each iteration [60]. Although this appears to work well, it requires the computation of singular values of the blur and regularization operators at each iteration. Sullivan and Chang [61], and Reeves and Perry [62] propose using cross-validation to determine the best stopping point.

The existing termination methods make use of the PSF in the stopping rule. Since this thesis assumes that this information is not available, only an estimate for the termination point can be made. For example, equation 4.6 can be used to determine a good the stopping point, where the variance of noise σ_n^2 is estimated from the degraded image. If the cost is within the range $(E\{J(u^*)\} - \epsilon, E\{J(u^*)\} + \epsilon)$, the restoration procedure is terminated. At this stopping point there is a chance that the parameters $u_T(x,y)$ are reasonably close to the inverse PSF. Assuming that $u_T(x,y)$ is, indeed, the inverse PSF an optimal filter in the presence of noise is applied on $g(x,y)$ to produce the image estimate. Unfortunately, the quality of the restoration is highly dependent on ϵ, which cannot be estimated a priori. The quantity must be chosen by monitoring the restoration procedure and terminating it at a visually appealing result.