TRdelft 18135

Stellingen

behorende bij het
proefschrift van
R.L. Lagendijk

Delft, 26 april 1990.

Het beeldidentificatie en -restauratie probleem kan het best worden aangepakt door een gezonde balans tussen wiskundige hulpmiddelen aan de ene kant, en heuristiek aan de andere kant.

De noodzaak tot regularizatie van het beeldrestauratieprobleem komt veel eerder voort uit het slechte conditiegetal dan uit de singulariteit van de vervormingsoperator.

"Regularization" en "Ringing Reduction" zijn doelstellingen die elkaar slecht verdragen.

Het neurale netwerk ten behoeve van beeldrestauratie voorgesteld door Zhou is niets anders dan een vereenvoudigde versie van het iteratieve "steepest descent" restauratiealgoritme beschreven in dit proefschrift.

Y.T. Zhou, R. Chellapa, A. Vaid, and B.K. Jenkins, "Digital image restoration using a neural net", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 7, pp. 1141–1151, 1988.

Bij algoritmen gebaseerd op projecties op convexe sets kunnen soms ook tegenstrijdige voorwaarden tot bruikbare oplossingen leiden. Zowel het Gerchberg algoritme voor superresolutie als het Papoulis algoritme voor bandbegrensde extrapolatie zijn voorbeelden hiervan.

R.W. Gerchberg, "Superresolution through error energy reduction", *Optica Acta*, vol. 21, pp. 709–720, 1974.
A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation", *IEEE Trans. Circuit and Systems*, vol. 22, pp. 735–742, 1975.

Alhoewel de maximum likelihood methode voor de identificatie van puntspreidingsfuncties geen unieke fase oplevert, is er normaliter voldoende voorkennis beschikbaar om deze fase-ambiguïteit op te lossen.

De parallelle identificatiemethode voor vervormde en verruiste beelden zoals beschreven door Blanc-Féraud levert geen maximum likelihood schatter op.

L. Blanc-Féraud, M. Barlaud, and P. Mathieu, "Image restoration and blur estimation using a constrained maximum likelihood method", *Proc. 3-rd Int. Workshop on Time-Varying Image Processing and Moving Object Recognition*, Florence, 1989.

Met name op het gebied van de beeldcodering worden aan het menselijk visueel systeem vaak vergezochte eigenschappen toegeschreven ter rechtvaardiging van de eigen onderzoeksmethoden.

Het schrijven van een proefschrift kan vooral dan efficient en zelfverzekerd gebeuren indien het te gebruiken materiaal reeds in een eerder stadium gepubliceerd is en dus de toets der wetenschappelijke kritiek doorstaan heeft.

"Beter onderwijs" impliceert niet noodzakelijkerwijs "meer schoolsheid".

Maar al te vaak dient men telefonisch de aankomst van een per elektronische post verstuurd bericht te verifiëren.

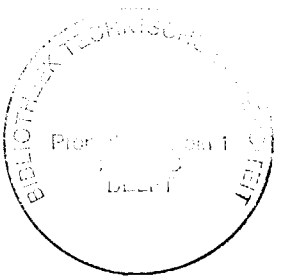
Juist in recreatieve competities is het de arbiter niet toegestaan te recreëren.

SU 694
1809
op 25-10-91 → a fw op Tues
TR diss 1813

Iterative Identification and Restoration of Images

TR diss
1813

Iterative Identification and Restoration of Images



Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus, prof.drs. P.A. Schenck,
in het openbaar te verdedigen
ten overstaan van een commissie
aangewezen door het College van Dekanen
op donderdag 26 april 1990 te 14.00 uur.

door

Reginald Leendert Lagendijk

geboren te Leiden

elektrotechnisch ingenieur

Dit proefschrift is goedgekeurd
door de promotor
prof.dr.ir. J. Biemond

Iteratieve Identificatie en Restoratie van Beelden

Aan Marleen en Annick

Contents

Summary

Images are produced in order to record or display useful information. Due to imperfections in the image formation process (camera, photographic film), however, the recorded image often represents a degraded version of the original scene. Although the degradations may have many causes, two types of degradations are usually dominant: blurring and noise. The field of image identification and restoration is concerned with the problem of undoing the effects of imperfections in the image formation process in order to facilitate the (human) interpretation or further processing (e.g. analysis) of the recorded image. More specifically the goal of image identification is to estimate the properties of the imperfect imaging system (blur) from the observed degraded image, together with some (statistical) characteristics of the noise and the original (uncorrupted) image. On the basis of these properties the image restoration process computes an estimate of the original image. Applications of image identification and restoration are found for instance in astronomy, medical imaging, and forensic science, and for analyzing images of unique events.

In this thesis we are concerned with iterative procedures for identifying and restoring images which have been degraded by a linear spatially invariant blur and additive white observation noise. As opposed to non-iterative methods, iterative schemes are able to solve the image restoration problem when formulated as a constrained and spatially variant optimization problem. It will be shown that in this way restoration results can be obtained which outperform the results of conventional restoration filters. In image identification we will introduce an efficient iterative procedure, known as the expectation-maximization (EM) algorithm, in order to optimize a complicated and nonlinear likelihood function. Most of the work reported in this thesis has been published in scientific journals [14,55,57,59].

In Chapter 2 we will discuss various aspects of modeling the image for-

mation process. These models form the groundwork for the mathematical treatment of the image identification and restoration problem as discussed in the remainder of the thesis.

Chapters 3 through 5 address many aspects of the use of iterative methods in image restoration. In Chapter 3 we will first discuss the ill-conditionedness of the restoration problem, which means that the observation noise that is inevitably mixed with the data is amplified enormously if an inverse filter is employed in the restoration process. Next a concise introduction is given to some well-known restoration filters, such as the Wiener and Kalman filter, constrained least-squares filter, and the method of projections onto convex sets. These methods are collectively called regularized restoration filters. They all make use of either a stochastic, algebraic, or deterministic formulation of *a priori* knowledge about the image to be estimated in order to find a compromise between the amplification of the observation noise and the accuracy of the solution.

Chapter 4 discusses another method to incorporate the fact that the observed blurred image is always contaminated by noise, namely by terminating the iterative implementation of the inverse filter prior to convergence. As a result a partially restored image is obtained which usually does not show any serious noise amplification. Additional advantages are that no matrix inverses need to be implemented, and that the method can be extended to more complex schemes. Several variations on the standard iterative restoration filters are derived. Although initially a basic steepest descent iterative scheme is employed, a number of more efficient implementations are considered as well.

In Chapter 5 we show that as a result of regularizing the restoration process, ringing artifacts are introduced in restored images. The iterative restoration algorithm introduced in this chapter incorporates two methods to suppress these ringing artifacts, namely (i) the use of deterministic *a priori* knowledge, and (ii) the local regulation of the magnification of the observation noise. Experimental results demonstrate that the proposed iterative scheme leads to results which are to be preferred to conventional restoration results in both numerical and visual respects.

Chapters 6 through 8 are concerned with recent methods to solve the image identification problem via a maximum likelihood approach. First, in Chapter 6 several conventional identification methods are discussed which are applicable to a rather restricted class of problems. Next the image identification problem is formulated as a maximum likelihood (ML) prob-

lem. Unfortunately, the computation of the ML estimator turns out to be a highly complicated and nonlinear optimization task. We show that all identification algorithms known sofar in the literature are merely different implementations of this estimator, resulting from different modeling assumptions (such as noisy or noiseless data), and/or considerations about the computational complexity or computer resources available (such as a preference for a recursive or a matrix-vector problem formulation).

In Chapter 7 the iterative expectatation-maximization (EM) algorithm is applied to the ML image identification problem. This leads to a particularly elegant algorithm which simultaneously identifies and restores the noisy blurred image. Whereas the original ML image identification method requires the solving of a nonlinear optimization problem, the proposed iterative identification procedure requires the solving of linear equations only. The derivation and performance of this method is discussed in detail.

Although the algorithms presented in Chapters 6 and 7 are mathematically well defined, they are subject to a number of restrictions in practical applications such as numerical inaccuracies and the sensitivity of the solution with respect to the initial estimate. Therefore in Chapter 8 two more practically oriented image identification techniques are proposed, both of which are based on the identification method developed in Chapter 7. In the first method structural knowledge about the blur and image model is incorporated into the identification process. The second method employs resolution pyramids in order to estimate the blur in a hierarchical manner. Experimental results on both synthetic and photographic motion and defocusing blurs are given. The chapter is concluded with an evaluation of the status quo in image identification and a look into possible future developments.

Chapter 1

The Image Identification and Restoration Problem

1.1 Introduction

Images are produced in order to record or display useful information. Due to imperfections in the electronic or photographic medium, however, the recorded image often represents a degraded version of the original scene (Figure 1.1). The degradations may have many causes, but two types of degradations are often dominant: blurring and noise. *Blurring* is a form of bandwidth reduction of the image due to the imperfect image formation process. It can be caused by relative motion between the camera and the original scene, or by an optical system which is out of focus. When aerial photographs are produced for remote sensing purposes, blurs are introduced by atmospheric turbulence, aberrations in the optical system, and relative motion between the camera and the ground. Such blurring is not confined to optical images, for example electron micrographs are corrupted by spherical aberrations of the electron lenses and CT scans suffer from X-ray scatter.

In addition to these blurring effects, the recorded image is also corrupted by *noises*. These may be introduced by the transmission medium (e.g. a noisy channel), the recording medium (e.g. film grain noise), measurement errors due to the limited accuracy of the recording system, and quantization of the data for digital storage.

The field of *image restoration* (sometimes referred to as image deblurring) is concerned with the reconstruction or estimation of the uncorrupted image from a distorted and noisy one. Essentially, it tries to perform an op-

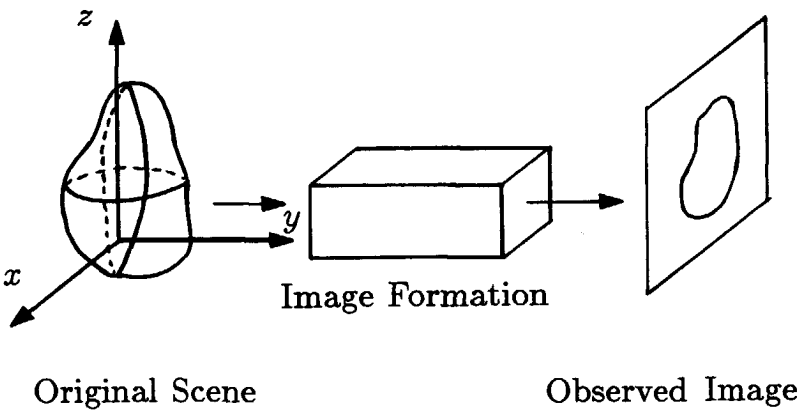


Figure 1.1: Image formation system.

eration on the image which is the inverse of the imperfections in the image formation system. It is important in fields such as astronomy, where the resolution limitations are severe, in medical imaging, where the physical requirements for high-quality imaging are unacceptable from a biological viewpoint, for analyzing images of unique events, and in forensic science, where potentially useful photographic evidence is sometimes of extremely bad quality.

In the use of image restoration methods, the characteristics of the degrading system and the noises are assumed to be known *a priori*. In practical situations, however, one usually has hardly enough knowledge to obtain this information directly from the image formation process. The goal of *image identification* is to estimate the properties of the imperfect imaging system from the observed degraded image itself prior to the restoration process.

In this thesis we will assume that the image formation can be adequately described by a linear spatially invariant relation and that the noises are additive. The observed or recorded image $g(i,j)$ is then given as

$$g(i,j) = d(i,j) * f(i,j) + w(i,j), \tag{1.1}$$

where $d(i,j)$ denotes the point-spread function (PSF) of the image formation system, where $f(i,j)$ is the ideal or original image that would have resulted from a perfect recording of the original scene, and where $w(i,j)$ models the noise in the recorded image. A more detailed discussion about the development of this model will be given in Chapter 2. In terms of the mathematical model (1.1) the purpose of image restoration can now be specified as the computation of an estimate $\hat{f}(i,j)$ of the original image $f(i,j)$ when $g(i,j)$ is observed, $d(i,j)$ is known and some (statistical) knowledge of both $f(i,j)$ and $w(i,j)$ is available. The complementary image identification problem on the other hand focusses on estimating $d(i,j)$ and the parameters of the statistical models for $f(i,j)$ and $w(i,j)$, from the observed image $g(i,j)$.

Noise constitutes an important limitation in the identification and restoration of images. The amount of noise present in an observed image is given by the (blurred-) signal-to-noise ratio (SNR):

$$\text{SNR} = 10 \log_{10} \left(\frac{\text{variance of the blurred image}}{\text{variance of the noise}} \right) \quad \text{(dB)}. \qquad (1.2)$$

Because images are nearly always digitally stored, the signal-to-noise ratio encountered in practical restoration applications is at most 40 to 50 dB. In this case the noise is not visible. On the other hand, for very low SNRs the degrading effect of the noise is more prominent than the blurring. For images that have an SNR less than 10 to 20 dB, the regular image identification and restoration algorithms are no longer useful, because at these SNR levels their effect is mainly to smooth out the noise and not perform any restoration at all. More feasible approaches to restoring such noisy images are provided by image enhancement techniques.

This thesis will deal exclusively with the *digital* identification and restoration of *monochromatic* (black-and-white or gray valued) images. Extension of the presented methods to color images is straightforward if the color image is described by a vector with three components corresponding to the tri-stimulus values red, green and blue. Considering each of these as a monochromatic image in itself, and neglecting mutual relations between the color components, the processing of a color image becomes equivalent to processing three independent monochromatic images.

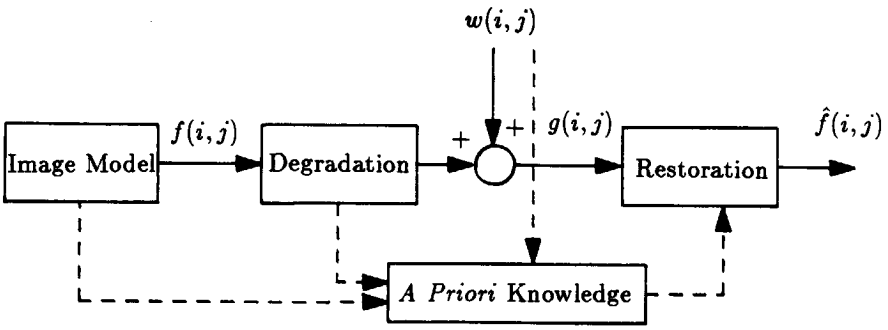


Figure 1.2: *A priori* restoration scheme.

1.2 Restoration Methods

Since the introduction of restoration in digital image processing in the sixties, a variety of image restoration methods have been developed. Nearly all these methods assume that the point-spread function $d(m, n)$ of the image formation process is known, and are therefore called *a priori* restoration. They concentrate on inverting Eq. (1.1) in order to get an estimate $\hat{f}(i, j)$ which is "as close as possible" to the original image $f(i, j)$ (Figure 1.2). One of the first methods used in image restoration was to simply neglect the presence of noise in (1.1) and to invert this equation through a frequency domain approach (inverse filtering) [3]. Since the amount of noise is usually not negligible, it will be amplified enormously by the restoration filter. Therefore this method does not lead to useful results.

Later on less noise-sensitive filters were developed which explicitly incorporate the fact that noise is always mixed with the data. These methods can collectively be called *regularized* restoration filters [99] and make use of either a stochastic or algebraic formulation of *a priori* knowledge about the image to be estimated. An aspect common to all regularized filters is that they trade off the amplification of the noise $w(i, j)$ to the resolution of the restored image. The class of regularized restoration methods includes the classical Wiener filter [3], and the constrained least-squares filter of Hunt [36] — which are usually implemented in the frequency domain —, and the recursive or Kalman filters in two dimensions of Woods *et al.* [112],

Biemond *et al.* [10,11], and Angwin *et al.* [5], which are implemented in the spatial domain.

Another method to explicitly incorporate the presence of noise consists of terminating an iterative restoration scheme prior to convergence. The limiting solution of such an iterative scheme is usually an inverse or regularized filter. The advantages of the iterative approach is that no matrix inverses need to be implemented, that the methods can be extended to more complex iteration schemes, and that the restoration results can be visually inspected as the iterations progress. The use of iterative schemes for image restoration has been proposed by many researchers, among which Kawata *et al.* [49], Katsaggelos *et al.* [45,46] and Maeda *et al.* [64].

More recent developments in recursive and iterative image restoration concentrate on implementing filters which are spatially adaptive in order to restore an image dependent on the local image content [4,41,46,47,55,89,98]. In this way restoration artifacts such as "ringing" [55] can be suppressed significantly and the results are typically superior to nonadaptive restoration results in both visual and numerical respect.

Other directions in image restoration, initiated by Youla *et al.* [115] and Sezan *et al.* [87], have involved the use of *deterministic a priori* knowledge about the original image and the noise. Since this knowledge is usually formulated as (nonlinear) constraints on the restored image, an iterative technique, known as the method of convex projections [17], is used to find the image satisfying all the *a priori* constraints.

Restoration methods which combine the use of deterministic knowledge with spatial adaptivity have recently been proposed as well. This combined approach requires the use of a constrained iterative optimization stategy [46,51,55], or the use of the method of convex projections [89]. Although the success of these methods may vary with the image content, they tend to outperform conventional restoration algorithms which are non-adaptive and do not make use of *a priori* constraints.

In most applications we assume that either a human interpreter or another image processing algorithm, such as a segmentation or analysis procedure, will be using the image restoration result. It would be ideal if we had available an evaluation criterion or performance measure that corresponds to either the human visual system or the requirements of the subsequent processing steps. Unfortunately, such criteria are hardly available, and the ones that are known are virtually impossible to use within the context of image restoration. Therefore restored images are usually objectively eval-

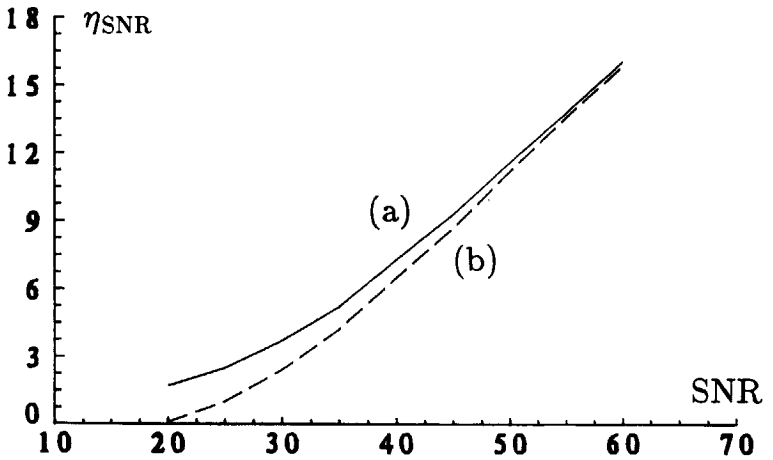


Figure 1.3: Signal-to-noise improvement as a function of the signal-to-noise ratio of the blurred image (a) using the exact image formation model parameters, and (b) using estimated parameters.

uated by using the following quadratic signal-to-noise ratio improvement measure:

$$\eta_{\text{SNR}} = 10 \log_{10} \frac{\sum_{i,j}(g(i,j) - f(i,j))^2}{\sum_{i,j}(\hat{f}(i,j) - f(i,j))^2} \quad \text{(dB)}, \tag{1.3}$$

where $\hat{f}(i,j)$ is the restored image. This performance measure can only be evaluated for controlled experiments in which the blur and noise have been synthetically introduced, since the undistorted image $f(i,j)$ is required as well. The maximally achievable signal-to-noise ratio improvement depends strongly on the content of the image, the type of blur considered and the signal-to-noise ratio of the blurred image. As an example Figure 1.3 shows η_{SNR} as a function of the amount of noise in the blurred image for a typical image which has been degraded by motion blur of moderate severity.

1.3 Identification Methods

Over the last decade the emphasis in image restoration research has been on the development and refinement of filtering algorithms. As has been observed by various researchers in this area, all methods utilized in the image

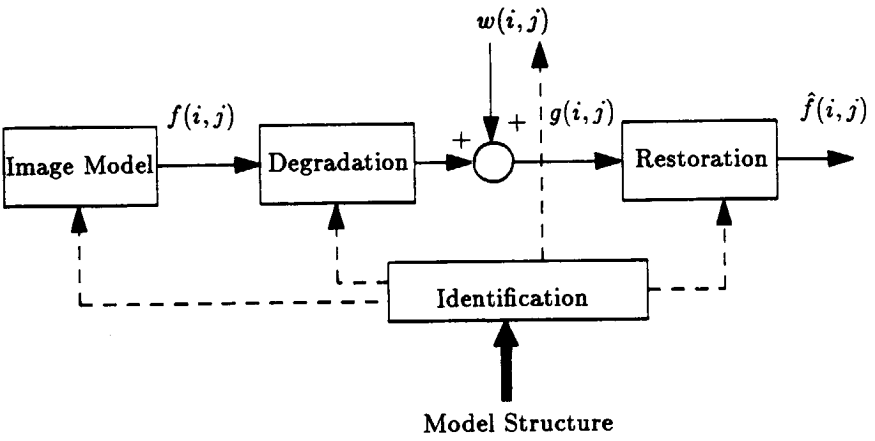


Figure 1.4: *A posteriori* identification and restoration scheme.

restoration literature have a comparable performance. This may indicate that image restoration has received maturity, or that the performance limitations of the current generation restoration methods have been reached and that entirely new directions are required to enforce a breakthrough. On the other hand the related image identification problem has received relatively little attention, but is becoming a research topic of increasing interest.

The general scheme for the combined identification and restoration of images, sometimes referred to as *a posteriori* restoration, is given in Figure 1.4. In this figure it is shown that prior to the restoration filtering, the characteristics of the blur (i.e. the point-spread function $d(m,n)$) must be estimated, as well as some (statistical) properties of the noise and the original image. Since the estimated parameters will always deviate somewhat from their true values, the resulting restored image is usually of poorer quality than the one restored with the exact parameter values (Figure 1.3). In evaluating the quality of the estimated image model, blur and noise characteristics, it is not sufficient to perform a numerical comparison between the estimated and the true parameter values. The reason for this is that the quality of the restoration result varies in a different way for estimation errors in each of the parameters. Only after restoring an image using the estimated parameters, can the performance of the identification

algorithms be evaluated. Therefore image identification cannot be considered as an isolated problem, but must be dealt with in conjunction with the restoration problem.

The earliest work in identifying the characteristics of the image formation system concentrated on PSFs which can be modeled parametrically and of which the spectra show a regular pattern of zero crossings. Since these zeros can also be located in the spectrum of the blurred image, spectral or cepstral techniques can be used to identify the PSF from the distance between spectral zeros of the blurred image [18,93]. The complementary phase estimation problem was addressed for example in [70]. Shortcomings of this method are that PSFs not satisfying these conditions cannot be identified and that the presence of noise in the observed data is not taken into account.

In [96], Tekalp *et al.* introduced a class of image identification methods that can handle a wide range of PSFs. In their work, the undistorted image is first modeled as a 2-D stochastic process and the presence of noise is neglected. Next, the identification problem is formulated as a maximum likelihood problem, which turns out to be equivalent to a 2-D autoregressive moving-average (ARMA) model identification problem. Several solution strategies were proposed to identify the coefficients of this ARMA process, such as the 2-D recursive methods of Tekalp *et al.* [96,97] and Wagner [107], the parallel banks of conventional 1-D ARMA process identification methods used by Biemond *et al.* [13], Katayama *et al.* [44], and Blanc-Féraud *et al.* [15], and gradient-based optimization algorithms proposed by Lagendijk *et al.* [52]. Because the noise is initially neglected, these methods are, however, restricted to blurred images with a relatively high signal-to-noise ratio.

Of recent date are the image identification methods introduced by Lagendijk *et al.* [53,57] and Angwin *et al.* [4] which are based on a maximum likelihood approach as well, but which explicitly take into account the presence of noise in the observed image. Computationally efficient algorithms based on the reduced order model Kalman filter [5] and the expectation-maximization algorithm [22] were proposed to optimize the resulting non-linear likelihood function. As a result of including the observation noise model, these algorithms outperform the earlier ones in identifying blurred images with a moderate to low signal-to-noise ratio.

1.4 Scope of the Thesis

This thesis addresses the use of iterative methods in modern image identification and restoration. The advantages of using iterative optimization algorithms are that they allow for a more flexible, possibly more complex but certainly improved formulation of the solution to the identification and restoration problem. As opposed to non-iterative methods, such as the recursive or frequency domain filters, iterative schemes can also handle problems which do not have an explicit analytical solution, but have been formulated as the optimization of a nonlinear and/or spatially variant objective function [14]. A disadvantage of iterative solution strategies is that the computational load is increased significantly compared to non-iterative methods. However, through the use of suitable parallel or array processing facilities, even complex iterative methods can be run within an acceptable amount of time.

This thesis is divided into three parts. In the first part (Chapter 2) various aspects of modeling the image formation process and of modeling images are considered. This chapter is important because it forms the groundwork for the mathematical treatment of the image identification and restoration problem as discussed in the remainder of the thesis. The image formation model (1.1) will be discussed in detail. Since this model has some problems near the boundaries of an image, remedies to this boundary value problem will be presented as well. Several relevant point-spread functions will be addressed in this chapter. In image restoration, it is common to model an image as a 2-D autoregressive process. A brief review of this type of image model will be given.

The second part of the thesis (Chapters 3 through 5) is concerned with many aspects of the use of iterative methods in image restoration, and the relation of iterative methods to other restoration methods. In Chapter 3 we will show that it is required to regularize the image restoration problem because of its ill-conditionedness. Next, several conventional regularized image restoration algorithms will be presented. The stochastic Wiener and Kalman restoration filter will be briefly reviewed. Next, the class of algebraic restoration methods will be discussed. It will be shown that stochastic and algebraic regularization methods are based on essentially the same concept. The chapter is concluded with the description of the method of convex projections. This class of image restoration methods

can utilize a variety of (possibly nonlinear) knowledge about the original image in order to regularize the restoration process.

In Chapter 4 the application of iterative methods in image restoration will be discussed from an historical point of view. We will show that iterative methods have certain advantages over the conventional restoration methods. More specifically, it will be shown that truncating an iterative scheme to compute the inverse filter is in fact a regularization method in itself. This idea is next extended to the regularized filters of Chapter 3, in this way introducing an additional degree of freedom in some well-known restoration methods. Various more efficient implementations of the basic iterative scheme will be considered.

In Chapter 5 we will work towards solving one of the most common problems in image restoration, namely the introduction of "ringing" artifacts as a result of regularizing the restoration filters. To this end, first the origin of ringing in linear spatially invariant image restoration algorithms is analyzed. Next, two possible methods to reduce these artifacts are discussed. A constrained adaptive iterative restoration algorithm is proposed which includes both ringing reduction methods. As opposed to Chapter 4, where iterative methods were used to regularize the restoration problem, iterative methods are used to optimize a spatially variant objective function subject to a nonlinear constraint in this chapter.

The third and last part of the thesis (Chapters 6 through 8) addresses recent methods to solve the image identification problem via a maximum likelihood (ML) approach. First, in Chapter 6 several conventional methods will be reviewed to estimate the PSF of an image formation system. Next, we will formulate the image identification problem as an ML problem, and consider its properties. From this formulation it follows that ML image identification requires the optimization of a complicated nonlinear function. It will be shown that all identification methods known sofar in the literature are merely different implementations of this estimator, resulting from different modeling assumptions and/or considerations about the computational complexity or computer resources available.

In Chapter 7 the iterative expectation-maximization algorithm that was developed for optimizing likelihood functions in general, will be applied to the image identification problem. It will be shown that this iterative approach yields a particularly elegant algorithm which simultaneously identifies and restores noisy blurred images. The derivation and performance of this method will be discussed in detail in this chapter.

The algorithms presented in Chapters 6 and 7 are mathematically well defined. However, in applying them to practical situations of interest sometimes an "engineer's approach" is required, which means that some heuristics need to be incorporated into the solution method. Two more practically oriented image identification algorithms will be considered in Chapter 8, both of which are based on the iterative identification scheme developed in Chapter 7. In the first method structural knowledge about the PSF and image model is incorporated into the identification process. In the second approach resolution pyramids are employed to estimate the PSF in a hierarchical manner. The performance of the presented image identification and restoration methods will be illustrated by numerous examples.

Chapter 2

Image Formation Models

In order to solve the image identification and restoration problem, mathematical models are required for the real-world processes involved in image generation, formation and recording. It may even be argued that the ultimate goal of image restoration, that is the recovery of the original undistorted image, can only be understood and formulated through the use of mathematical models which reflect in one way or another the *a priori* information one has regarding the observed image, the image formation process and the original image. Without such knowledge, the identification and restoration of a degraded image could probably not be accomplished.

The determination of suitable analytical models is not a trivial problem in general and requires careful considerations with respect to the level of abstraction that is acceptable for the application considered. A model that is too simple probably leads to computationally elegant algorithms which have, however, no practical relevance. On the other hand, models which do describe real-world processes in detail are often too complex to be feasible from an algorithmic viewpoint.

This chapter describes several aspects of modeling noisy blurred images for the purpose of contemporary image identification and restoration. In view of the results obtained by many researchers in this area, including the work described in this thesis, it is arguable that the presented models achieve an acceptable level of abstraction for a wide class of (but certainly not all) image identification and restoration problems.

In Section 2.1 the image formation and recording process will be modeled as a linear process. After discussing the fundamental limitations in establishing the relations between the blurring process in a 3-D continuous world and a 2-D discrete image formation model, we describe both

a discrete input-output relation and a state-space representation for the image formation process. In Section 2.2 an input-output relation and a state-space representation for modeling discrete images will be discussed. Finally, in Section 2.3 several of the most common 1-D and 2-D point-spread functions encountered in practice will be discussed.

2.1 Blur Models

2.1.1 Linear Image Formation

It is appropriate to begin by assuming that a three-dimensional (3-D) object or scene has been imaged onto a 2-D imaging plane by means of a recording system such as a camera (Figure 1.1). Since image recording systems are never perfect, both deterministic and statistical distortions may be introduced in general.

The deterministic degradations introduced by the imaging process may be very complex [34] for several reasons. In the first place the 2-D imaging system may not be able to capture all 3-D phenomena, such as a 3-D rotation of an object, parallax and 3-D geometry effects. Secondly the transfer function of the imaging system itself may be very complex because of diffraction effects in optical systems, system aberrations, atmospheric turbulence, motion blurs and defocused systems, although the effect of these imperfections can usually be considered to be a bandwidth reduction (or blurring) of the recorded scene. Furthermore the severity of the bandwidth reduction may vary as a function of the image coordinates (spatially varying blur). Finally, all kind of nonlinearities due to the response of the sensor may further degrade the recorded image.

Since modeling and restoring 3-D degradations is virtually impossible in an image restoration context, we will restrict ourselves to 2-D degradations in this thesis. The nonlinear response of the sensor on the other hand cannot be discarded without any further discussion. Sometimes it may be approximated by a linear response, for example the logarithmic behavior of a photographic medium becomes linear when the image is of low contrast [3]. If such a linearization is not possible, however, there are two ways to handle the nonlinearity. The mathematically most elegant way is to incorporate the nonlinear response into the image formation model. Unfortunately, such a model yields complex restoration algorithms [12,37,101], and might even be impossible to use within image identifi-

cation algorithms. A second more practical approach assumes that it is allowed to first apply the inverse sensor response to the observed image before any further processing is done. Although this way of dealing with nonlinearities has several faults, it has been shown to be unexpectedly succesful in the practice of image restoration. Throughout this thesis we will assume that nonlinearities introduced by the sensor can be removed either by a linearization approach or by applying the inverse sensor response in advance.

From the above discussion we learn that in many cases of practical interest it is justifiable to restrict the modeling of the image formation system to the usually dominant effect of blurring. The entire process then becomes a linear system characterized by a 2-D point-spread function (PSF) $d(x, y; s, t)$. For both incoherent optical systems [31] and penetrating radiation systems the observed image is given by the following 2-D superposition integral [3]:

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(x, y; s, t) f(s, t) ds dt. \tag{2.1}$$

Both the original image $f(x, y)$ and the observed image $g(x, y)$ represent real-valued intensity distributions and take nonnegative values only. As a result $d(x, y; s, t)$ is real-valued and nonnegative as well.

Unfortunately, the model (2.1) is not very useful for image identification and restoration purposes, because the complexity implied by the possibility of having a different PSF $d(x, y; s, t)$ at each coordinate (s, t) of the image is unacceptable from a computational viewpoint. Furthermore, it is rather unrealistic to assume that one might be able to estimate a different PSF for each location in the image simply because of the lack of sufficient information for the estimation procedures. We therefore have to assume that the PSF of the image formation process is stationary or spatially invariant over (at least a significant portion of) the image, yielding

$$\begin{aligned} g(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(x - s, y - t) f(s, t) ds dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d(s, t) f(x - s, y - t) ds dt \\ &= d(x, y) * f(x, y), \end{aligned} \tag{2.2}$$

where $*$ is used to denote 2-D convolution.

In order to perform image identification and restoration by using a digital computer, the discrete equivalent of (2.2) is required. To this end

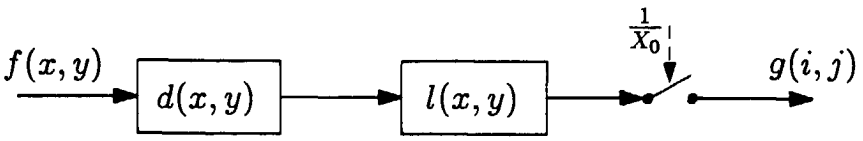


Figure 2.1: Image formation and sampling.

the blurred image $g(x,y)$ is sampled on a 2-D regularly spaced lattice with dimensions $M \times N$ after appropriately restricting its bandwidth (see Figure 2.1). Practical values of M and N are 128, 256, 512 or even larger. The discrete observed image $g(i,j)$ is then given by [25]:

$$\begin{aligned} g(i,j) &= \{l(x,y) * g(x,y)\}_{(x,y)=(iX_0,jX_0)} \\ &= \{l(x,y) * d(x,y) * f(x,y)\}_{(x,y)=(iX_0,jX_0)}, \\ & \qquad i \in [0, M-1], j \in [0, N-1], \quad (2.3) \end{aligned}$$

indicating that the samples of $g(i,j)$ are taken from $l(x,y) * g(x,y)$ at the (vertical and horizontal) coordinates $(x,y) = (iX_0, jX_0)$, where X_0 is the sample distance and $l(x,y)$ is the spatial response of the bandwidth restricting filter.

In order to arrive at an expression that is equivalent to (2.3), but is based on a discrete convolution between the sampled original image $f(i,j)$ and the sampled PSF $d(i,j)$, (2.3) is rewritten as follows (Figure 2.2):

$$\begin{aligned} g(i,j) &= \{l(x,y) * d(x,y) * f(x,y)\}_{(x,y)=(iX_0,jX_0)}, \\ &= \{l(x,y) * d(x,y)\}_{(x,y)=(iX_0,jX_0)} \ * \ \{l(x,y) * f(x,y\}_{(x,y)=(iX_0,jX_0)} \\ &= d(i,j) * f(i,j) \\ &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d(m,n) f(i-m, j-n). \quad (2.4) \end{aligned}$$

The above reformulation of (2.3) is only correct if one of the following two (mutually exclusive) sets of conditions holds:

(1) – $f(x,y)$ and $d(x,y)$ are bandlimited with maximum spatial frequencies Ω_f and Ω_d, respectively,

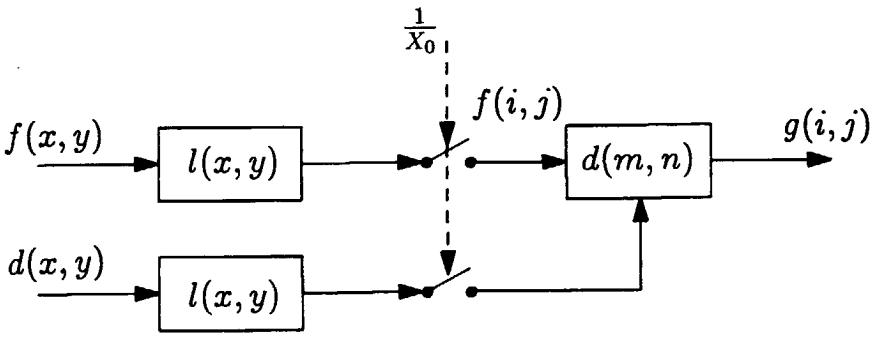


Figure 2.2: Equivalent discrete image formation system.

- the sample distance satisfies $X_0 \leq 1/\{2\max(\Omega_f, \Omega_d)\}$, and
- $l(x,y)$ has a 2-D all-pass character for spatial frequencies less than $\max(\Omega_f, \Omega_d)$,

(2)
- $f(x,y)$ and $d(x,y)$ are not bandlimited,
- $l(x,y)$ is an ideal 2-D low-pass filter with cut-off frequency Ω_c, and
- the sample distance satisfies $X_0 \leq 1/2\Omega_c$.

Under the first set of conditions listed $f(x,y)$ and $d(x,y)$ can be reconstructed perfectly from their samples $f(i,j)$ and $d(i,j)$, respectively, while under the second set of conditions this is no longer possible because a part of the information of $f(x,y)$ and/or $d(x,y)$ is lost due to the required low-pass filtering. In both cases $f(i,j)$, $g(i,j)$ and $d(i,j)$ can be sampled without introducing aliasing.

Unfortunately, neither of the two sets of conditions can be satisfied in practice, because (i) continuous images and PSFs in particular are hardly ever bandlimited, and (ii) the ideal low-pass filter $l(x,y)$ cannot be realized. For instance, consider the use of a (CCD) sensor array camera to record an image. In this case the low-pass filter $l(x,y)$ is implicitly incorporated into the system by the integrating operation of each array element. Since this integration does not approximate the required ideal low-pass filtering very well, the sampled blurred image $g(i,j)$ is aliased. As a general conclusion

we see that (2.4) never describes the actual continuous image formation (2.3) entirely correct.

Hence, if the discrete model (2.4) is used for the purpose of image identification and restoration using real-world data, the parameters and restored images obtained may deviate from their actual values not only due to possible estimation inaccuracies, but also due to the fundamental restrictions in sampling and reconstructing images. Nevertheless, we restrict the discussion in this thesis to the image formation model (2.4), keeping in mind that if we use relatively high sampling rates, the aliasing effects in $g(i,j)$ play a negligible role. In that case the discrete model (2.4) is an acceptable approximation to the continuous image formation process (2.2).

Equation (2.4) is the general expression for a 2-D discrete convolution in which the imaging system has a point-spread function of infinite support. In real life, however, the observed intensity at the position (i,j) is only affected by intensity values in a small neighborhood of (i,j). In other words, we may very well replace (2.4) by a system with a finite 2-D point-spread function (i.e. a 2-D FIR filter):

$$\begin{aligned} g(i,j) &= \sum_{m=-m_1}^{m_2} \sum_{n=-n_1}^{n_2} d(m,n) f(i-m, j-n) \\ &= \sum_{m,n \in S_d} d(m,n) f(i-m, j-n). \end{aligned} \tag{2.5}$$

Here we use S_d to denote the support of the PSF, i.e. the finite extent region where $d(m,n) \neq 0$, which may have any shape but is typically noncausal [3] in the sense that $g(i,j)$ is a function of both past and future pixels with respect to most definitions of causality in two dimensions [112]. For the purpose of simplicity we have replaced the double summation symbol by a single summation symbol in (2.5).

A convenient shorthand notation of (2.5) can be arrived at by row-wise scanning the image and stacking the data into a vector (called lexicographic ordering, see Appendix A) [78], yielding

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{f}, \tag{2.6}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are the lexicographically ordered vectors of size $MN \times 1$. If a circular convolution is assumed in (2.5), the blurring matrix $\boldsymbol{D}$ (of size $MN \times MN$) has a block-circulant structure. We will return to the validity

of this assumption in Section 2.1.3. Because of the properties of $d(x,y)$, $g(x,y)$ and $f(x,y)$ the entries of the matrix $\boldsymbol{D}$, and the vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ are real-valued and always positive.

The advantage of having a circular convolution in (2.5) is that the eigenvalues and -vectors of $\boldsymbol{D}$ can be computed easily, and are in fact given by the coefficients of the discrete Fourier transform of $d(m,n)$ and the discrete Fourier basis functions, respectively (Appendix A). Therefore an alternative formulation of (2.5) and (2.6) is the frequency domain model

$$G(u,v) = D(u,v)F(u,v). \tag{2.7}$$

Here capitals denote the Fourier domain samples and u, v the discrete vertical and horizontal frequency variables.

Besides the deterministic distortions described to this point, recorded images are invariably degraded by stochastic degradations, usually referred to as observation noise. It may originate from the image formation process, the transmission medium, the recording process, quantization of the data or any combination of these [3]. In this thesis we will model the noise contributions as an additive zero-mean white Gaussian noise process with variance σ_w^2, which is statistically uncorrelated with the images. This is a simplification since noises such as film grain noise and noise due to photon statistics are not uncorrelated with the input and may even be non-additive. This simplification nonetheless leads to identification and restoration methods which can be applied to a wide class of problems.

Denoting the noise contribution by $w(i,j)$, $\boldsymbol{w}$ or $W(u,v)$, the complete image formation model becomes (Figure 2.3):

$$g(i,j) = \sum_{m,n \in S_d} d(m,n) f(i-m, j-n) + w(i,j), \tag{2.8}$$

which is alternatively written as the following matrix-vector or frequency domain relation:

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{f} + \boldsymbol{w}, \tag{2.9}$$

$$G(u,v) = D(u,v)F(u,v) + W(u,v). \tag{2.10}$$

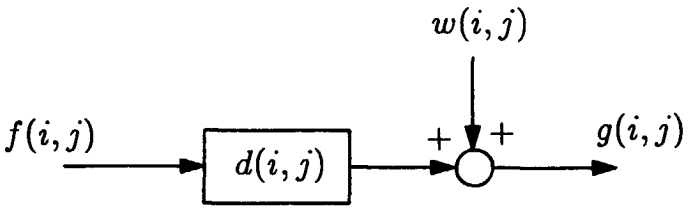


Figure 2.3: Discrete image formation model.

2.1.2 State-Space Representation

The models presented in the previous section describe the image formation system through an input-output relation of a linear system. Another way to describe such a system is through the use of a state-space representation, as is common in for example recursive filtering. Once an ordering of the image data has been chosen (causality condition), the 2-D image formation model can be converted into an equivalent 1-D model by the use of a state. The state at the image coordinates (m, n), denoted by $s^{(m,n)}$, has to be chosen in such a way that it includes all past and present information (pixel values) to determine future responses uniquely. We note here that converting the 2-D FIR model (2.8) into a state-space representation is not very useful by itself. However, in combination with the state-space representation for the image model, which will be introduced in Section 2.2.2, these equations form the basis for Kalman restoration filtering in two dimensions [112].

There are many ways to define the concept "state" in multiple dimensions. Here we follow the approach of Woods *et al.* [111,112] in ordering the image data, which consists of scanning the image row-wise. For a given support S_d of the PSF, the state now includes all image rows over which the PSF extends. In order to obtain a causal support for the state from the usually noncausal PSF the observations are delayed over a suitable vertical and horizontal distance (or in other words, a fixed delay model is used). If the PSF has a support of size $(m_1 + m_2 + 1) \times (n_1 + n_2 + 1)$, a suitable definition of the state is:

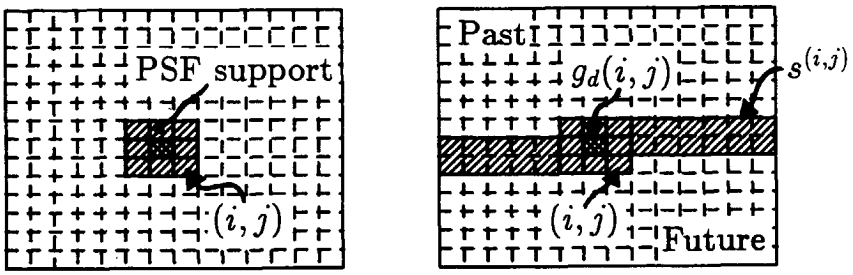


Figure 2.4: Example of a state definition for a 3×3 point-spread function.

$$\begin{aligned} s^{(i,j)} &= [f(i,j), f(i,j-1), \cdots, f(i,0), \\ &\quad f(i-1,N-1), f(i-1,N-2), \cdots, f(i-1,0), \\ &\quad \cdots\cdots, f(i-m_1-m_2, j-n_1-n_2)]^t. \end{aligned} \tag{2.11}$$

As an example Figure 2.4 shows the state of a system related to a noncausal PSF having a 3×3 support S_d.

Based on the definition of the state, the observation model (2.8) can be cast in the following form:

$$g_d(i,j) = \boldsymbol{d}^t \boldsymbol{s}^{(i,j)} + w(i,j). \tag{2.12}$$

Here $g_d(i,j)$ denotes the observations $g(i,j)$ which are delayed over a fixed vertical and horizontal distance m_1 and n_1, respectively. Further, the blur vector $\boldsymbol{d}$ is of the same dimension as the state $\boldsymbol{s}^{(i,j)}$, and contains the appropriately arranged PSF coefficients $d(m,n)$.

2.1.3 Boundary Value Problem

In nearly any practical situation an image is recorded by a device of finite spatial extent. As a result not all the data $f(i,j)$ that is required to model $g(i,j)$ through (2.8) is available, since the convolution extends beyond the boundaries of the image. This problem is known as the boundary value

problem in image processing, and requires special attention in image identification and restoration when real-life data is used. Although at a first glance the boundary value problem appears to have a (negligible) effect because only the boundaries are incorrectly modeled, a more detailed analysis reveals that it has a global effect on the identification and restoration task [113], and deteriorates the performance of these significantly.

In order to avoid the modeling errors near the boundaries from affecting the interior of the (restored) image, the blurred image needs to be preprocessed prior to the identification and restoration process. The missing data just outside the boundaries of $g(i,j)$, called boundary values, are substituted by values which are a close approximation of the true values at these locations. Several methods have been proposed to assign boundary values [4,15,112,113]:

- the boundary values may be fixed at the (local) image mean,
- the boundary values can be given as random values with a mean and variance the same as those measured from the image, or they can be obtained by a model based extrapolation,
- repeat or mirror the P first and last rows and columns to obtain the boundary values,
- the effective size of the image to be restored can be reduced so that the data which is no longer a part of this reduced image can be used as boundary values,
- a circulant approximation of the image can be made where the left and right, and the upper and lower boundaries are thought to be connected.

The last two methods have shown to be most useful in an image restoration environment [4,113]. Reducing the effective image size seems to be tailored best to recursive methods, especially with regard to the implementation aspects of linear convolutions in (2.8) [4]. In restoration methods that employ discrete Fourier transforms, the circulant approximation of an image is most popular, because the evaluation of (circular) convolutions using DFTs relies on the same assumption [30].

If, however, the blurred image is assumed to be circulant, intensity jumps may occur at the boundaries due to possible intensity differences

between the left and right and the upper and lower boundaries of the image. These intensity jumps introduce leakage frequencies into the blurred image which are not present in the actual data. During the deconvolution action of the restoration filter certain leakage frequencies may be amplified enormously, which will give rise to spurious oscillations dominating the entire restored image. Therefore, if an image is either implicitly or explicitly regarded as being circulant, the leakage frequencies must be suppressed. In practice this can be done to a sufficient degree by a linear or third order interpolation of the intensity difference between the facing boundaries [113].

Summarizing, in those regions of the image where the observation model does not fit the actual data very well due to boundary effects, we do not adapt the model, but preprocess the data such that (2.8) is consistent with either a linear or circular convolution. In this thesis we will assume that (2.8) indicates a circular convolution when we consider DFT implementations of image identification and restoration algorithms, so that $\boldsymbol{D}$ can be assumed to be block-circulant. In the recursive implementations (2.8) represents a linear convolution, so that $\boldsymbol{D}$ has a block-Toeplitz structure. In both cases, it will be assumed that the image data has been preprocessed appropriately.

2.2 Image Models

Many image identification and restoration methods make use of *a priori* knowledge about the structure of the original image. In this section we will restrict the discussion to modeling statistical characteristics of discrete images through the use of 2-D autoregressive models. A comprehensive survey of these images models can be found in [9,40,43], including extensive discussions about fitting image models, the model quality and model stability. In Chapter 3 other ways of formulating *a priori* knowledge about images will be addressed.

2.2.1 2-D AR Modeling

The development of a suitable model for discrete images requires a tradeoff between the accuracy of representation and its utility in image identification and restoration. In view of the experimental results obtained by many researchers in the area of image restoration, we use a 2-D autoregressive

(AR) model driven by an independent zero-mean[1] white noise process with variance σ_v^2:

$$\begin{aligned} f(i,j) &= \sum_{k,l \in S_a} a(k,l) f(i-k, j-l) + v(i,j) \\ &= a(i,j) * f(i,j) + v(i,j). \end{aligned} \tag{2.13}$$

Here $a(k,l)$ denote the minimum mean squared error (MMSE) image model coefficients which are computed by minimizing the variance of the noise process $\sigma_v^2 = \mathrm{E}[v(i,j)^2]$. In this respect $v(i,j)$ may also be regarded as the modeling error between a complex real-world image and the relatively simple autoregressive process fitted onto this data.

Different models result from different choices of the image model support S_a. Some common choices for the model support are:

$$S_a = \begin{cases} \{(k,l) : (k \geq 0, l \geq 0) \cap (k+l > 0)\} & \text{, quarter plane} \\ \{(k,l) : (k > 0, l \leq 0) \cup (k \geq 0, l > 0)\} & \text{, nonsymmetric halfplane} \\ \{(k,l) : (k > 0, \forall l) \cup (k = 0, \forall l \neq 0\} & \text{, semi-causal} \\ \{(k,l) : \forall (k,l) \neq (0,0)\} & \text{, noncausal.} \end{cases} \tag{2.14}$$

Experimental evidence shows that the order of the image model, that is the size of the support S_a, is not very critical in an image restoration setting. For this reason we restrict ourselves to first order models (see Figure 2.5). A remark about the statistical properties of $v(i,j)$ is required here. Assume that we have available a real-life image, and we fit an image model onto this image by minimizing σ_v^2 (see Section 2.2.3). Then the assumption of $v(i,j)$ being a white noise process holds only if the support of the image model used is causal and has an infinite order [110]. However, if we use a finite model order, the causal image model will not lead to a white noise process driving the autoregressive model, unless the image to be modeled is a realization of a 2-D homogeneous discrete Markovian field [110]. Since this requirement is rarely met, (2.13) is a suboptimal model with respect to the statistical properties of the driving noise $v(i,j)$.

The image model (2.13) can only be of use in image identification and restoration if it is BIBO (bounded-input, bounded-output) stable. A general criterion for the image model stability is given by the following z-transform of the image model [9,40]:

[1] This assumption is conflicting with the fact that image intensities are nonnegative. Therefore, the observed noisy blurred image is usually corrected for its mean value in order to satisfy the zero-mean assumption.

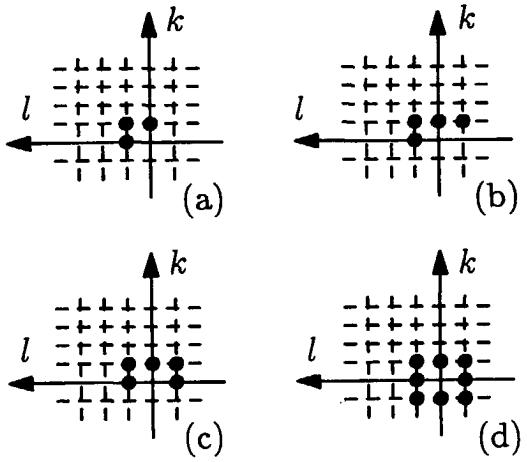


Figure 2.5: Model support for various first order image models (a) quarter plane model, (b) nonsymmetric halfplane model, (c) semi-causal model, and (d) noncausal model.

$$
\begin{aligned}
B(z_1, z_2) \;=\; & 1 - \sum_{k,l \in S_a} a(k,l) z_1^{-k} z_2^{-l} \neq 0, \\
& \text{for } |z_1| \geq 1 \text{ and } |z_2| \geq 1, \text{ causal models,} \\
& \text{for } |z_1| \geq 1 \text{ and } |z_2| = 1, \text{ semi-causal models,} \\
& \text{for } |z_1| = 1 \text{ and } |z_2| = 1, \text{ noncausal models.} \quad (2.15)
\end{aligned}
$$

When a set of image model coefficients is computed by fitting a causal image model onto a given image, the stability of the resulting image model is assured [40].

An alternative more compact notation of (2.13) is given by

$$\boldsymbol{f} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{v}, \tag{2.16}$$

where $\boldsymbol{f}$ and $\boldsymbol{v}$ are the lexicographically ordered images $f(i,j)$ and $v(i,j)$, respectively, and where $\boldsymbol{A}$ is the image model matrix of size $MN \times MN$. If the boundary value problem is solved by assuming that $f(i,j)$ is circulant, the convolution in (2.13) is circular and $\boldsymbol{A}$ has a block-circulant structure.

In this situation DFTs can be used to yield the frequency domain relation

$$F(u,v) = \frac{1}{1 - A(u,v)} V(u,v). \tag{2.17}$$

Observe that because we have assumed that the image model (2.13) is stable, $B(z_1, z_2)$ in (2.15) does not have zeros on the unit bi-circle and hence the reciprocal of $(1 - A(u,v))$ exists for all (u,v). Because of the same reason the matrix $(\boldsymbol{I} - \boldsymbol{A})$ is always nonsingular.

2.2.2 State-Space Representation

If 2-D recursive procedures are used to identify and restore a noisy blurred image, the stochastic image model (2.13) needs to be converted into a state-space model through a procedure similar to the one described for converting the observation model (2.5) into a state-space representation. Although the choice between the various image model supports typically depends on the model quality, algorithmic complexity and a preference for causal, semi-causal or noncausal supports, the nonsymmetric halfplane (NSHP) support has various advantages over the others, as becomes apparent in for instance Kalman filtering in two dimensions [111,112] and the factorization of 2-D spectra [24,111]. Further, the NSHP image model immediately allows for a 2-D recursion without the necessity to delay the observations. Eq. (2.13) is written in a state-space form as:

$$\boldsymbol{s}^{(i,j)} = \boldsymbol{T}\boldsymbol{s}^{(i,j-1)} + [1,0,0,\cdots,0]^t v(i,j), \tag{2.18}$$

where $\boldsymbol{T}$ is the state transition matrix containing the corresponding NSHP image model coefficients $a(k,l)$. If we restrict ourselves to first order NSHP image models, the following pixels from $\boldsymbol{f}$ are included in the state $\boldsymbol{s}^{(i,j)}$ at the coordinates (i,j) [111] (Figure 2.6):

$$\begin{aligned} \boldsymbol{s}^{(i,j)} \;=\; & [f(i,j), f(i,j-1), \cdots, f(i,0), \\ & f(i-1,N-1), f(i-1,N-2), \cdots, f(i-1,j-1)]^t. \end{aligned} \tag{2.19}$$

The dimension of the state vector equals the number of columns in the image plus 2. For higher order NSHP models the state will extend over multiple image rows.

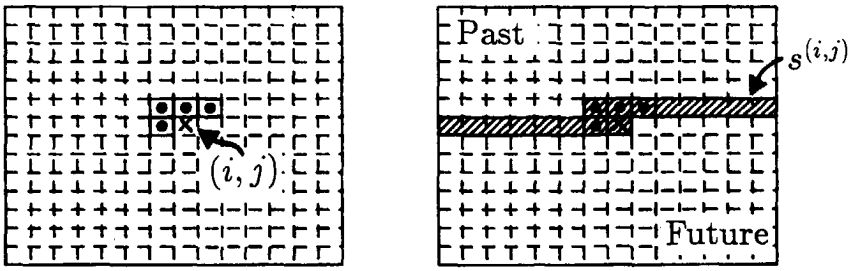


Figure 2.6: State definition for a 1-st order NSHP image model.

2.2.3 Model Fitting Problem

In this section we will briefly review the computation of the MMSE image model coefficients $a(k,l)$ based on the autocorrelation function of a given (original) image. The MMSE image model coefficients are obtained by minimizing the variance of the driving noise, given by

$$\sigma_v^2 = \mathrm{E}[\{f(i,j) - \sum_{k,l \in S_a} a(k,l) f(i-k, j-l)\}^2], \tag{2.20}$$

with respect to $a(p,q)$ for $(p,q) \in S_a$. Straightforward evaluation of this minimization problem yields:

$$\mathrm{E}[\{f(i,j) - \sum_{k,l \in S_a} a(k,l) f(i-k, j-l)\} f(i-p, j-q)] = 0. \tag{2.21}$$

Rewriting these equations leads to the Yule-Walker equations:

$$r_{ff}(p,q) = \sum_{k,l \in S_a} a(k,l) r_{ff}(p-k, q-l), \qquad \forall (p,q) \in S_a, \tag{2.22}$$

$$\sigma_v^2 = r_{ff}(0,0) - \sum_{k,l \in S_a} a(k,l) r_{ff}(k,l), \tag{2.23}$$

where the autocorrelation coefficients of the original image are defined by $r_{ff}(p,q) = \mathrm{E}[f(i,j) f(i-p, j-q)]$. From the above set of linear equations the image model coefficients can be determined immediately.

2.3 Common Point-spread Functions

In this section several of the most common point-spread functions encountered in practice are discussed. The point-spread functions will be specified as parametric functions $d(s,t;\psi)$, where the parameter (-vector) ψ determines the severity of the blur.

Since the blurring of images is basically a continuous process, we will present the PSFs in their continuous formulation. In digital image identification and restoration, however, we make use of the discrete model (2.8). As was discussed in Section 2.1.1, an exact relation between the discrete PSF $d(m,n)$ and the continuous PSF $d(s,t;\psi)$ cannot be established due to aliasing problems. As an approximating approach we assume that the sample rate is chosen sufficiently high in order to minimize the modeling errors, and that a simple square sensor array is used to sample the PSF. In mathematical form the discretized PSF $d(m,n)$ can then be given as:

$$d(m,n) = \int_{\Omega(m,n)} d(s,t;\psi)\, ds\, dt, \tag{2.24}$$

$$\Omega(m,n) = \{s,t | m-\frac{1}{2} \leq s \leq m+\frac{1}{2}, n-\frac{1}{2} \leq t \leq n+\frac{1}{2}\}. \tag{2.25}$$

A point-spread function cannot take arbitrary values. In the observation model (2.1) it was observed that the original and recorded image are nonnegative real-valued quantities because of the physics of the underlying image formation process. As a consequence, the PSFs $d(s,t;\psi)$ and $d(m,n)$ need to be nonnegative and real-valued as well.

In addition to this, the imperfections in an image formation system normally act as passive operations on the data, i.e. they do not absorb or generate energy. Consequently, all energy arising from a specific point in the original image should be preserved, yielding

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d(s,t;\psi)\, ds\, dt = 1.0. \tag{2.26}$$

A discrete PSF is constrained to satisfy

$$\sum_{m,n\in S_d} d(m,n) = 1.0. \tag{2.27}$$

In the following we will present four common point-spread functions, which are encountered regularly in practical situations of interest.

2.3.1 Linear Motion Blur

Many types of motion blur can be distinguished [34], all of which are due to relative motion between the recording device and the object. This can be in the form of a translation, a rotation, a sudden change of scale, or some combinations of these. Here only the important case of a translation will be considered.

When the object translates at a constant velocity V under an angle of ϕ radians with the horizontal axis during the exposure interval $[0, T]$, the distortion is one dimensional. Defining the "length of motion" by $L = VT$, the PSF is given by:

$$d(s,t;L,\phi) = \begin{cases} \frac{1}{L} & \text{, if } \sqrt{s^2+t^2} \leq \frac{L}{2} \text{ and } s/t = -\tan\phi, \\ 0 & \text{, elsewhere.} \end{cases} \tag{2.28}$$

The PSF in this case is space-invariant (stationary), but if only a part of the image is subject to translational motion, the overall distortion is obviously spatially variant.

2.3.2 Uniform Out-of-Focus Blur

When a 3-D scene is imaged by a camera onto a 2-D imaging plane, some parts of the scene are in focus while other parts are not. If the aperture of the camera is circular, the image of any point source is a small disk, known as the circle of confusion (COC). The degree of defocus (diameter of the COC) depends on the focal length $\mathcal{F}$ and the aperture number n of the lens, and the distance $\mathcal{P}$ between camera and object. If the camera is focused sharply at an object at distance $\mathcal{S}$, the diameter of the COC, denoted by $C(\mathcal{P})$, is given by [77]

$$C(\mathcal{P}) = \begin{cases} \frac{\mathcal{F}\mathcal{S}}{n(\mathcal{S}-\mathcal{F})} - \frac{\mathcal{F}^2\mathcal{S}}{n\mathcal{P}(\mathcal{S}-\mathcal{F})} - \frac{\mathcal{F}}{n}, & \text{for} \quad \mathcal{S} < \mathcal{P} < \infty, \\ -\frac{\mathcal{F}\mathcal{S}}{n(\mathcal{S}-\mathcal{F})} + \frac{\mathcal{F}^2\mathcal{S}}{n\mathcal{P}(\mathcal{S}-\mathcal{F})} + \frac{\mathcal{F}}{n}, & \text{for} \quad \mathcal{F} < \mathcal{P} < \mathcal{S}. \end{cases} \tag{2.29}$$

To obtain a complete model for defocusing we need to know the intensity distribution within the COC caused by a point object. An accurate model includes the effect of diffraction [31]. If, however, the degree of defocusing is large relative to the wavelengths considered, a geometrical approach can be followed [94], resulting in a uniform intensity distribution within the COC. The PSF of this uniform out-of-focus blur with a radius of $R = \frac{C(\mathcal{P})}{2}$

is given by:

$$d(s,t;R) = \begin{cases} \frac{1}{\pi R^2} & \text{, if } \sqrt{s^2+t^2} \leq R, \\ 0 & \text{, elsewhere.} \end{cases} \tag{2.30}$$

2.3.3 Atmospheric Turbulence Blur

Atmospheric turbulence is a severe limitation in remote sensing and aerial imaging as used in for example weather predictions. Though the blur introduced by atmospheric turbulence depends on a variety of factors (such as temperature, windspeed, exposure time), for long-term exposures the point-spread function can reasonably well be described by a Gaussian function [66]:

$$d(s,t;\sigma_G) = C \exp\left\{-\frac{s^2+t^2}{2\sigma_G^2}\right\}. \tag{2.31}$$

Here σ_G determines the severity of the blur. The constant C is chosen in such a way that (2.26) is satisfied. Observe that since the above blur model does not have a finite support, it has to be truncated properly.

2.3.4 Scatter Blur

X-ray images show detail by the varying amount of radiation that was absorbed by the object being radiated. Unfortunately, the X-ray quanta are also scattered from their incident paths, resulting in a distribution of radiation about a point [3]. Again, there are many factors which influence the PSF resulting from this scatter, but within diagnostic energy ranges the PSF may be described sufficiently accurately by the following radially symmetric PSF [108]:

$$d(s,t;\beta_s) = \frac{C}{(\beta_s^2 + (s^2+t^2))^{\frac{3}{2}}}. \tag{2.32}$$

Here β_s determines the severity of the blur, and is a function of the distance between the radiated object and the detector. The above PSF needs to be truncated properly to obtain a support S_d of finite extent.

Chapter 3

Regularized Image Restoration

In the previous chapter we have established the mathematical framework for image restoration by developing relevant models for both real-world images and the image formation process. The purpose of image restoration can now be formulated as the estimation of an improved image $\hat{\boldsymbol{f}}$ of the original image $\boldsymbol{f}$ when a noisy blurred version $\boldsymbol{g}$, given by

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{f} + \boldsymbol{w}, \tag{3.1}$$

is observed. In Chapters 3 through 5 we assume that the blurring matrix $\boldsymbol{D}$ is known. Further, some statistical knowledge about $\boldsymbol{w}$ and $\boldsymbol{f}$ is assumed to be available. Specifically, we will assume that the image model

$$\boldsymbol{f} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{v} \tag{3.2}$$

is feasible. Methods to obtain the parameters of these models will be addressed in Chapters 6 through 8.

The direct inversion of the matrix $\boldsymbol{D}$ does not lead to useful restoration results, because of the ill-conditionedness of $\boldsymbol{D}$, as will be discussed in the first section of this chapter. Procedures to stabilize the inversion of an ill-conditioned matrix are called regularization methods, and make nearly always use of *a priori* knowledge about the original image and the noise. In Section 3.2 through 3.4 we will examine some particular approaches to regularize the ill-conditioned image restoration problem. In Section 3.2 linear methods will be discussed which make use of stochastic models to formulate *a priori* knowledge. An alternative, though related approach is the algebraic regularization approach which will be presented in Section 3.3. This chapter is concluded with the review of a nonlinear method that

is probably most intimately related to "the use of *a priori* knowledge", namely to explicitly enforce the solution of the restoration problem to satisfy multiple *a priori* constraints. In the next chapter we will investigate the use of iterative methods to achieve regularization.

3.1 Ill-Conditionedness of the Image Restoration Problem

The simplest approach to the inverse problem of restoring images is to use a filter whose response, formulated as the matrix operator $\boldsymbol{D}^{-1}$, is the inverse of $\boldsymbol{D}$ [3] in the sense that

$$\boldsymbol{D}^{-1}\boldsymbol{D} = \boldsymbol{I}, \tag{3.3}$$

where $\boldsymbol{I}$ is the identity matrix. If such an inverse filter exists, its output will be equal to

$$\hat{\boldsymbol{f}} = \boldsymbol{D}^{-1}\boldsymbol{g} = \boldsymbol{f} + \boldsymbol{D}^{-1}\boldsymbol{w}. \tag{3.4}$$

The restored image is thus equal to the desired image plus inverse filtered noise. Unfortunately, there are several problems with the above approach, which we can analyze best through the use of an eigenvalue expansion of the above linear equations.

Let $\{\boldsymbol{z}_{uv}; u \in [0, M-1], v \in [0, N-1]\}$ denote the eigenvectors of unit length associated with the blurring matrix $\boldsymbol{D}$ and let λ_{uv} denote the corresponding (complex-valued) eigenvalues (see Appendix A). For the purpose of simplicity, we assume that $\boldsymbol{D}$ and $\boldsymbol{D}^t$ have the same set of eigenvectors. As a result the eigenvectors $\boldsymbol{z}_{uv}$ are mutually orthogonal, and $\hat{\boldsymbol{f}}$ can be expanded in terms of these eigenvectors as follows:

$$\begin{aligned}
\hat{\boldsymbol{f}} &= \sum_{u,v}(\hat{\boldsymbol{f}}, \boldsymbol{z}_{uv})\boldsymbol{z}_{uv} = \sum_{u,v}(\boldsymbol{D}^{-1}\boldsymbol{g}, \boldsymbol{z}_{uv})\boldsymbol{z}_{uv} \\
&= \sum_{u,v}\frac{1}{\lambda_{uv}}(\boldsymbol{g}, \boldsymbol{z}_{uv})\boldsymbol{z}_{uv} \qquad (3.5) \\
&= \sum_{u,v}\frac{1}{\lambda_{uv}}(\boldsymbol{D}\boldsymbol{f} + \boldsymbol{w}, \boldsymbol{z}_{uv})\boldsymbol{z}_{uv} \\
&= \sum_{u,v}(\boldsymbol{f}, \boldsymbol{z}_{uv})\boldsymbol{z}_{uv} + \sum_{u,v}\frac{1}{\lambda_{uv}}(\boldsymbol{w}, \boldsymbol{z}_{uv})\boldsymbol{z}_{uv}. \qquad (3.6)
\end{aligned}$$

The first term on the right hand side in (3.6) represents the original image $\boldsymbol{f}$ expanded in terms of the eigenvectors of $\boldsymbol{D}$, and the second term is

an observation noise related term. Based on the above expansion the restoration error $\hat{\boldsymbol{f}} - \boldsymbol{f}$ can be bounded by

$$\|\hat{\boldsymbol{f}} - \boldsymbol{f}\| \leq \sum_{u,v} \frac{1}{|\lambda_{uv}|} |(\boldsymbol{w}, \boldsymbol{z}_{uv})|, \tag{3.7}$$

where $\|\cdot\|$ denotes the regular Euclidean norm, i.e.:

$$\|\boldsymbol{f}\| = [\boldsymbol{f}^t \boldsymbol{f}]^{\frac{1}{2}} = \sqrt{\sum_{i,j} |f(i,j)|^2}. \tag{3.8}$$

In practice the above eigenvalue expansion may be hard to realize when dealing with images, primarily because computing the eigenvectors of a huge matrix of size $MN \times MN$ is virtually impossible. However, recalling that $\boldsymbol{D}$ has a block-circulant structure when it is related to a spatially-invariant circular convolution, the eigenvalues λ_{uv} are given by the 2-D discrete Fourier transform coefficients $D(u,v)$ of $d(m,n)$, and the eigenvectors are given by the basis functions of the 2-D DFT. Relations such as (3.6) may therefore immediately be interpreted as a frequency domain relation, e.g.

$$\hat{F}(u,v) = F(u,v) + \frac{1}{D(u,v)} W(u,v). \tag{3.9}$$

Based on (3.6) or (3.9) we see that if the blurring matrix $\boldsymbol{D}$ has eigenvalues which are equal to zero, the inverse filter may not exist.

Even if $\boldsymbol{D}$ has non-zero eigenvalues only, there are usually problems due to excessive noise amplification at higher frequencies as represented by the second term on the right hand size of (3.6) or (3.9). This is because the transfer function of the blur (defined as $|D(u,v)|$) is typically highest at low frequencies (or eigenvectors number) and rolls off significantly for higher ones (the blur acts like a low-pass filter). The spectrum of the observation noise, on the other hand, typically contains relatively much high frequency components. Thus at high frequencies, $W(u,v)/D(u,v)$ may take very large values so that $\hat{\boldsymbol{f}}$ is dominated by the inverse filtered noise, yielding useless solutions. The restoration error $\|\hat{\boldsymbol{f}} - \boldsymbol{f}\|$ will practically be unbounded in this case.

The dilemmas involved in solving inverse problems of the form (3.1) were first studied by Hadamard [33]. He observed that the solution $\hat{\boldsymbol{f}}$ could differ by an arbitrary large amount from the true solution due to small

errors (noise) in measuring the observed signal. Based on his investigations and that of later mathematicians, the term "ill-posed" or "ill-conditioned"[1] problem was introduced to denote the class of problems which behaved in a similar manner. It has widely been recognized that the (image) restoration problem as addressed in this thesis is always an ill-posed problem, though its severity may vary from moderate (Gaussian PSFs) to very severe (strongly bandlimiting PSFs such as motion blur) [1,3,7,42,46,61,73,76,81,91].

The main objective in solving ill-posed problems is [74,99] "the construction of a physically acceptable and meaningful approximation of the true solution that is sufficiently stable from a computational point of view". If we are to obtain a useful approximate solution to (3.1), we must modify the problem in such a way that the solution becomes less sensitive to noise in the observed image (well-posed problem). At the same time the solution of the modified problem must be close to the solution of the original problem. The process of achieving a compromise between these conflicting goals is referred to as "regularization", and is typically controlled by one or more regularization parameters. Nearly all concepts used in regularization are based on incorporating *a priori* knowledge about either the true solution or the noise into the algorithm which solves the image restoration problem. In the next sections we will address the stochastic and algebraic approaches to regularization, which yield essentially linear restoration methods, and a nonlinear regularization approach based on the use of deterministic constraints.

3.2 Stochastic Restoration

3.2.1 Bayesian Methods

Bayesian estimation methods are used when the *a posteriori* probability density function (PDF) of the original image, given the observed image, can be established. To this end the PDFs of $\boldsymbol{f}$ and $\boldsymbol{w}$ are required. Although these PDFs are hardly ever known exactly and might take any shape, it is often justified in practice to assume that both the noise process $\boldsymbol{w}$ and the

[1] The term ill-conditionedness is reserved for ill-posed problems in finite dimensional spaces (matrix operators). In general, ill-posedness in an infinite dimensional space is harder to tackle than ill-posedness in a finite dimensional setting. Similar solution techniques are however being used for ill-posed and ill-conditioned problems.

original image $\boldsymbol{f}$ are multivariate Gaussian with zero-mean and correlation matrices $\boldsymbol{R}_{ww} = \sigma_w^2 \boldsymbol{I}$ and $\boldsymbol{R}_{ff}$, respectively. These assumptions essentially form the *a priori* knowledge used to regularize the restoration problem in a Bayesian approach.

The Bayesian estimator of $\boldsymbol{f}$ maximizes the *a posteriori* PDF $p(\boldsymbol{f}/\boldsymbol{g})$ with respect to $\boldsymbol{f}$. Using Bayes' law and some basic mathematics we get [3]

$$\begin{aligned} p(\boldsymbol{f}/\boldsymbol{g}) &= \frac{p(\boldsymbol{g}/\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{g})} \\ &= C \exp\left\{(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})^t \boldsymbol{R}_{ww}^{-1}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f}) + \boldsymbol{f}^t \boldsymbol{R}_{ff}^{-1}\boldsymbol{f}\right\}. \end{aligned} \tag{3.10}$$

Optimization of (3.10) with respect to $\boldsymbol{f}$ yields the MAP (maximum *a posteriori*) estimator:

$$\hat{\boldsymbol{f}} = \left[\boldsymbol{D}^t\boldsymbol{D} + \boldsymbol{R}_{ww}\boldsymbol{R}_{ff}^{-1}\right]^{-1}\boldsymbol{D}^t\boldsymbol{g}. \tag{3.11}$$

This restoration filter can easily be implemented using DFTs because all matrices involved are block-circulant. With hindsight we note that (3.11) has a form similar to the Wiener filter and the Tikhonov-Miller regularized filter. The analysis of the properties of (3.11) will therefore be deferred to Section 3.3.

Associated with the MAP estimator is the ML (maximum likelihood) estimator which considers $\boldsymbol{f}$ as a nonrandom vector, and maximizes the likelihood function $p(\boldsymbol{g}/\boldsymbol{f})$ with respect to $\boldsymbol{f}$. This yields the following solution:

$$\hat{\boldsymbol{f}} = \left[\boldsymbol{D}^t\boldsymbol{D}\right]^{-1}\boldsymbol{D}^t\boldsymbol{g}, \tag{3.12}$$

which reduces to the inverse filter in this case [3]. Thus, if $\boldsymbol{f}$ is regarded as being nonrandom, no model is assumed for the original image, which yields a nonregularized (and therefore useless) solution when the observed image is noisy.

3.2.2 Wiener Filtering

The Wiener filter or linear minimum variance estimator is designed to minimize the mean-squared error between the original image and the restoration result. Denoting the Wiener filter by a matrix $\boldsymbol{H}_w$ we get:

$$\boldsymbol{H}_w \leftarrow \min_{\boldsymbol{H}_w} \mathrm{E}[(\boldsymbol{f}-\hat{\boldsymbol{f}})^t(\boldsymbol{f}-\hat{\boldsymbol{f}})], \tag{3.13}$$

$$\text{where} \qquad \hat{\boldsymbol{f}} = \boldsymbol{H}_w\boldsymbol{g}. \tag{3.14}$$

Here the ill-conditionedness of the restoration problem is handled by the use of the mean-squared restoration error criterion. The solution to this rather classical problem is given by

$$\hat{\boldsymbol{f}} = \boldsymbol{H}_w \boldsymbol{g} = \left[\boldsymbol{D}^t \boldsymbol{D} + \boldsymbol{R}_{ww} \boldsymbol{R}_{ff}^{-1}\right]^{-1} \boldsymbol{D}^t \boldsymbol{g}, \tag{3.15}$$

which is identical to the MAP estimator under the multivariate Gaussian assumptions. If we do not incorporate the fact that the blurred image is noisy, i.e. $\sigma_w^2 = 0$ and hence $\boldsymbol{R}_{ww} = \boldsymbol{O}$, the Wiener filter reduces to the inverse filter.

3.2.3 Kalman Filtering

Another solution to linear mean-squared error image restoration uses a Kalman filter. To this end we replace the image model (3.2) and the observation equation (3.1) by their state-space counterparts (2.12) and (2.18):

$$\boldsymbol{s}^{(i,j)} = \boldsymbol{T}\boldsymbol{s}^{(i,j-1)} + [1,0,0,\cdots,0]^t v(i,j), \tag{3.16}$$

$$g_d(i,j) = \boldsymbol{d}^t \boldsymbol{s}^{(i,j)} + w(i,j). \tag{3.17}$$

The state at the coordinates (i,j), denoted by $\boldsymbol{s}^{(i,j)}$, contains all the pixels that are required to determine future responses of the above dynamic system uniquely. It can be formed by considering the union of the states described in the Sections 2.1.2 and 2.2.2. Figure 3.1 shows the state definition for the above system when the image model has a 1-st order NSHP support $\mathcal{S}_a$, and the blur has a 3×3 support $\mathcal{S}_d$.

Since the above model equations are essentially one dimensional, a standard 1-D Kalman filter, such as [50], can be applied to restore the image [111,112]. It consists of two stages: a prediction step (denoted by the subscript "b") and an update step (denoted by the subscript "a"):

$$\hat{\boldsymbol{s}}_b^{(i,j)} = \boldsymbol{T}\hat{\boldsymbol{s}}_a^{(i,j-1)}, \tag{3.18}$$

$$\hat{\boldsymbol{s}}_a^{(i,j)} = \hat{\boldsymbol{s}}_b^{(i,j)} + \boldsymbol{k}^{(i,j)}[g_d(i,j) - \boldsymbol{d}^t \hat{\boldsymbol{s}}_b^{(i,j)}]. \tag{3.19}$$

Here $\boldsymbol{k}^{(i,j)}$ denotes the Kalman gain, computed via

$$\boldsymbol{k}^{(i,j)} = \boldsymbol{P}_b^{(i,j)} \boldsymbol{d}[\boldsymbol{d}^t \boldsymbol{P}_b^{(i,j)} \boldsymbol{d} + \sigma_w^2]^{-1}, \tag{3.20}$$

$$\boldsymbol{P}_b^{(i,j)} = \boldsymbol{T}\boldsymbol{P}_a^{(i,j-1)}\boldsymbol{T}^t + \sigma_v^2[1,0,\cdots 0]^t[1,0,\cdots 0], \tag{3.21}$$

$$\boldsymbol{P}_a^{(i,j)} = [\boldsymbol{I} - \boldsymbol{k}^{(i,j)}\boldsymbol{d}^t]\boldsymbol{P}_b^{(i,j)}. \tag{3.22}$$

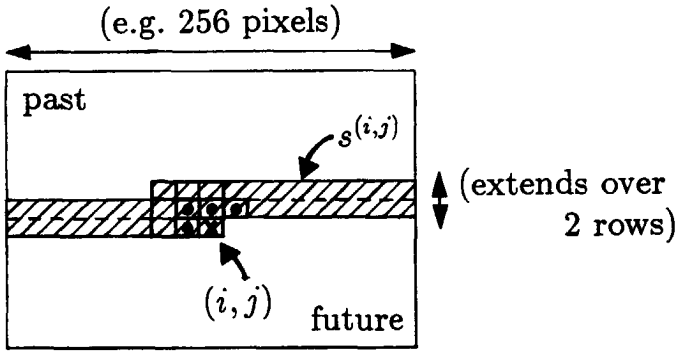


Figure 3.1: Example of a state definition for Kalman filtering in two dimensions.

Here $\boldsymbol{P}_b^{(i,j)}$ and $\boldsymbol{P}_a^{(i,j)}$ denote the predicted and updated covariance matrices of the state:

$$\boldsymbol{P}_b^{(i,j)} = \mathrm{E}[(\boldsymbol{s}^{(i,j)} - \hat{\boldsymbol{s}}_b^{(i,j)})(\boldsymbol{s}^{(i,j)} - \hat{\boldsymbol{s}}_b^{(i,j)})^t], \tag{3.23}$$

$$\boldsymbol{P}_a^{(i,j)} = \mathrm{E}[(\boldsymbol{s}^{(i,j)} - \hat{\boldsymbol{s}}_a^{(i,j)})(\boldsymbol{s}^{(i,j)} - \hat{\boldsymbol{s}}_a^{(i,j)})^t]. \tag{3.24}$$

The combination of (3.20) - (3.22) is usually referred to as the Ricatti equation for the computation of the Kalman gain.

Unfortunately, the dimension of the state may be very large for 2-D blurs. For example, in Figure 3.1 the state contains more than 500 elements for an image with 256 columns. Therefore, the use of the optimal Kalman filter implies extensive computation, which hampers its practical applicability. A suboptimal but efficient alternative to the above 1-D Kalman filter in two dimensions was presented by Woods *et al.* [111,112] who essentially reduced the update of the state in (3.19) to those elements of the state which are within a certain distance of the current pixel (i,j) (reduced update Kalman filter: RUKF). More recently Angwin *et al.* [5] proposed to replace the exact state-space representation of the NSHP image model by an approximate one which uses a state of much lower dimension. As a result they are able to employ the optimal 1-D Kalman for their approximate image model (reduced order model Kalman filter: ROMKF).

In contrast to these scalar Kalman filters Biemond *et al.* developed a Kalman filter for vector observations in which the image is restored line-by-line [10,11]. By using a decorrelating row transform, such as a DFT [10] or DCT [15], this filter for vector observations reduces to a set of parallel 1-D scalar filters. Comparable restoration results were reported for the various (approximating) implementations of the Kalman filter in two dimensions.

3.3 Algebraic Restoration

3.3.1 Tikhonov-Miller Regularization

Tikhonov and Arsenin [99] were the first to study the concepts of regularization in a general setting, though some important prior work had been performed by Phillips [75] and Twomey [105] and a number of Russian mathematicians. Their work was applied later on in image restoration by Hunt [36] and Katsaggelos [46]. The idea is to define a criterion to select an approximate solution from a set of feasible solutions. On the basis of the observation model (3.1) it is plausible that the class of feasible solutions is described by

$$\Phi(\hat{\boldsymbol{f}}) = \|\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}\| \leq \|\boldsymbol{w}\| = \epsilon. \tag{3.25}$$

The bound ϵ is related to the uncertainty or noise in the observed image $\boldsymbol{g}$, and can usually be estimated from a smooth image region. The set of feasible solutions is primarily populated with unacceptable solutions because of the ill-conditioned nature of the restoration problem. Tikhonov defined the regularized solution as the one which minimizes a stabilizing functional $\Omega(\hat{\boldsymbol{f}})$ on the set of feasible solutions.

Although a wide class of different stabilizing functionals is available, including for example the maximum power [102] and maximum entropy [32] measures, usually a stabilizing functional of the following form is chosen to facilitate the mathematical analysis of the problem [7,74,99]:

$$\Omega(\hat{\boldsymbol{f}}) = \|\boldsymbol{C}\hat{\boldsymbol{f}}\|. \tag{3.26}$$

Here $\boldsymbol{C}$ is a real-valued matrix of size $MN \times MN$, known as the regularizing operator. In order to be able to evaluate (3.26), $\boldsymbol{C}$ is restricted to have a block-circulant structure. The required properties of this operator will be discussed in Section 3.3.2.

The computation of the regularized solution now reduces to the minimization of (3.26) subject to (3.25). Using the method of undetermined Lagrange multipliers we need to minimize the objective function

$$\Phi(\hat{\boldsymbol{f}}) = \|\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}\|^2 + \alpha\|\boldsymbol{C}\hat{\boldsymbol{f}}\|^2 \tag{3.27}$$

with respect to $\hat{\boldsymbol{f}}$. The regularization parameter α is chosen so that (3.25) is satisfied with equality.

Another related approach was suggested by Miller in [67], who suggested to replace the minimization of $\Omega(\hat{\boldsymbol{f}})$ by the following bound:

$$\Omega(\hat{\boldsymbol{f}}) = \|\boldsymbol{C}\hat{\boldsymbol{f}}\| \leq E. \tag{3.28}$$

Eqs. (3.25) and (3.28) can be combined into a single quadrature formula, yielding

$$\Phi(\hat{\boldsymbol{f}}) = \|\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}\|^2 + (\epsilon/E)^2\|\boldsymbol{C}\hat{\boldsymbol{f}}\|^2 \leq 2\epsilon^2. \tag{3.29}$$

The result is identical to the Tikhonov result with $\alpha = (\epsilon/E)^2$. If a solution $\hat{\boldsymbol{f}}$ satisfies the bounds in (3.25) and (3.28), then it will also satisfy (3.29). Conversely, if a solution $\hat{\boldsymbol{f}}$ satisfies (3.29), then (3.25) and (3.28) are also satisfied except for a factor of at most $\sqrt{2}$ [67], which is insignificant for practical cases. Other ways to select feasible values for α are reported in for example [29,92,109].

Among the solutions satisfying (3.29) a reasonable choice is the one which minimizes $\Phi(\hat{\boldsymbol{f}})$, called the Tikhonov-Miller regularized solution. This minimization is straightforward and leads to the following solution:

$$(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C})\hat{\boldsymbol{f}} = \boldsymbol{D}^t\boldsymbol{g}, \tag{3.30}$$

which is equivalent to

$$\hat{\boldsymbol{f}} = (\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C})^{-1}\boldsymbol{D}^t\boldsymbol{g}, \tag{3.31}$$

provided that $(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C})$ is invertible. An alternative to the direct computation of $\hat{\boldsymbol{f}}$ in (3.31) is the iterative implementation that will be discussed in Chapter 4.

The Tikhonov-Miller regularized solution is a linear filter that has the same form as both the MAP estimator and the Wiener filter. We will elaborate on the formal relation between these methods in Section 3.3.2. The restoration filter (3.31) formulated in the discrete frequency domain was first introduced by Hunt in [36] as the constrained least-squares filter. For $\boldsymbol{C} = \boldsymbol{I}$ and $\alpha \to 0$ the constrained least-squares filter becomes the pseudo-inverse filter [3], which minimizes the norm of the restored image subject to the condition (3.25).

3.3.2 Choice of the Regularizing Operator

In order to be effective the Tikhonov-Miller regularization has to reduce the noise sensitivity of the restoration process. In this section we will analyze the effect of regularization through an eigenvector expansion of (3.30), and give a qualitative description of the intended effects of the regularizing operator.

Let $\{\boldsymbol{z}_{uv}; u \in [0, M-1], v \in [0, N-1]\}$ denote the eigenvectors associated with the blurring matrix $\boldsymbol{D}$, the regularizing operator $\boldsymbol{C}$, and their transposed version $\boldsymbol{D}^t$ and $\boldsymbol{C}^t$, and let λ_{uv} and μ_{uv} denote the eigenvalues of $\boldsymbol{D}$ and $\boldsymbol{C}$, respectively. In this thesis we restrict the discussion to matrices $\boldsymbol{D}, \boldsymbol{D}^t, \boldsymbol{C}$ and $\boldsymbol{C}^t$ that have the same eigenvectors. This is a valid assumption, for example, for the important case when the associated convolutions are circulant and spatially invariant. Similar expressions as the ones that will be given in the following can be derived for the more general case when $\boldsymbol{D}, \boldsymbol{D}^t, \boldsymbol{C}$ and $\boldsymbol{C}^t$ have different eigenvectors. Such a more complex analysis does, however, not provide more insight in the effects of regularization.

The Tikhonov-Miller solution in (3.31) can now be given by [46,72,73,74]

$$\hat{\boldsymbol{f}} = \sum_{u,v} \frac{\lambda_{uv}^*}{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2} (\boldsymbol{g}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv}. \tag{3.32}$$

Clearly the effect of regularization is to modify the denominator of the inverse filter in (3.5). To decide in which way the denominator should be modified, i.e. how the regularization parameter α and the eigenvalues μ_{uv} of $\boldsymbol{C}$ should be chosen, it is appropriate to consider the difference between the true and regularized solution. This error can be bounded by:

$$\begin{aligned} \|\hat{\boldsymbol{f}} - \boldsymbol{f}\| &\leq \sum_{u,v} \frac{\alpha|\mu_{uv}|^2}{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2} |(\boldsymbol{f}, \boldsymbol{z}_{uv})| + \sum_{u,v} \frac{|\lambda_{uv}|}{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2} |(\boldsymbol{w}, \boldsymbol{z}_{uv})| \\ &= \mathrm{E}_r(\alpha) + \mathrm{E}_n(\alpha). \end{aligned} \tag{3.33}$$

The first term on the right hand side of (3.33) denotes the error due to regularization. For a fixed $\boldsymbol{D}$ and $\boldsymbol{C}$ it can be minimized by setting $\alpha = 0$. The second term, which measures the noise magnification error, however, becomes extremely large as $\alpha \to 0$ if any of the $|\lambda_{uv}|$ are close to zero. For $\alpha \to \infty$ the noise magnification error approaches zero, but now the regularization error becomes large. In the limiting situation, the noise magnification error equals zero, and the regularization error equals

$E_r(\infty) = \sum_{u,v} |(f, z_{uv})|$. Obviously the choice of α requires the trade-off between these two errors.

Another way to trade off the regularization and noise magnification error is through the selection of the (eigenvalues of the) regularizing operator C. Since the original signal should not be overly corrupted by the process of regularization, it is reasonable to choose $|\mu_{uv}| \ll |\lambda_{uv}|$ when $|(f, z_{uv})| \gg |(w, z_{uv})|$. This means that the regularization process has hardly any effect on those (spectral) components where there is relatively little noise. On the other hand components dominated by noise should strongly be subjected to regularization, i.e. $|\mu_{uv}| \gg |\lambda_{uv}|$ when $|(f, z_{uv})| \ll |(w, z_{uv})|$. The interpretation of this choice in terms of frequency domain arguments is that since (i) the signal energy is concentrated in the low-frequency range, (ii) the noise is broad-banded, and (iii) the blur acts like a form of low-pass filter, the regularizing operator C should act like a high-pass filter.

A popular choice for the regularizing operator is the second-derivative operator. In a discrete implementation this high-pass filter is known as the 2-D Laplacian filter [36], and is given in the spatial domain as:

$$c(i,j) = \begin{bmatrix} 0.00 & -0.25 & 0.00 \\ -0.25 & 1.00 & -0.25 \\ 0.00 & -0.25 & 0.00 \end{bmatrix}. \tag{3.34}$$

Although other choices may be more feasible in a particular situation, the above general discussion for choosing C will remain valid.

By choosing the properties of the regularizing operator appropriately, the Tikhonov-Miller regularization approach can be related to a restoration method known as the truncated singular value decomposition (SVD) [35]. In this method only those eigenvalues λ_{uv} are inverted which do not lead to any serious noise magnification. In terms of Eq. (3.32) we set $\mu_{uv} = 0$ for eigenvalues $|\lambda_{uv}| > \mathcal{T}$, and we entirely reject using the eigenvalues for which $|\lambda_{uv}| < \mathcal{T}$ by letting $\alpha|\mu_{uv}| \to \infty$. Equation (3.32) then becomes:

$$\hat{f} = \sum_{\{u,v|\,|\lambda_{uv}|>\mathcal{T}\}} \frac{1}{\lambda_{uv}} (g, z_{uv}) z_{uv}. \tag{3.35}$$

Clearly the value of the threshold $\mathcal{T}$ determines how many eigenvalues of D are used to compute the restoration result. If $\mathcal{T}$ is too large, not enough spectral components of g will be used to compute the restoration result

and the regularization error in (3.35) will be dominating. On the other hand, if T is taken too small the noise magnification will be dominant. In fact, in the limit for T approaches zero, (3.35) equals the (pseudo-)inverse filter. Since the threshold T controls the trade-off between the two types of errors, it can be considered a regularization parameter. The optimal value of T is usually selected by visual inspection of the restoration result.

Example of Regularized Restoration

In this example we illustrate the effects of tuning the regularization parameter α on the regularization and noise magnification error. We consider the original "Cameraman" image in Figure 3.3a (256×256 pixels, each pixel uniformly quantized in 8 bits), which was synthetically blurred by defocusing blur (R=3) and white Gaussian noise was added to the result at an SNR of 40 dB (Figure 3.3b). This image was restored using the Tikhonov-Miller regularized restoration filter. The discrete Laplacian operator was used as the regularizing operator, and α was varied between 10^{-6} and 10^{1}. Figure 3.2 shows how the two components of the restoration error change as the value of α varies. The optimal result is obtained for $\alpha = 0.01$ and has a SNR improvement of $\eta_{\text{SNR}} = 6.1$ dB. Observe that these curves and the improvement in signal-to-noise ratio can be computed only in a controlled experiment, because they are functions of the original (undistorted) image. In Figure 3.4 an arrangement is shown of the restored image and the regularization error for various values of α. It is noted that the regularization error mainly effects the regions near edges, while the noise magnification affects the entire image. We will study this behavior in more detail in Chapter 5.

3.3.3 Formal Relation with Stochastic Restoration

In the previous section we have noted that several restoration filters (MAP, Wiener, Tikhonov-Miller) have the same mathematical form although they were motivated by entirely different arguments. The generic form of these filters is

$$\hat{\boldsymbol{f}} = (\boldsymbol{D}^t\boldsymbol{D} + \boldsymbol{M})^{-1}\boldsymbol{D}^t\boldsymbol{g}, \tag{3.36}$$

where $\boldsymbol{M}$ is chosen as follows:

- $\boldsymbol{M} = \alpha\boldsymbol{I}$ $(\alpha \to 0)$, for the (pseudo-)inverse filter,
- $\boldsymbol{M} = \boldsymbol{R}_{ww}\boldsymbol{R}_{ff}^{-1}$ for the MAP and Wiener filter (stochastic regularization),

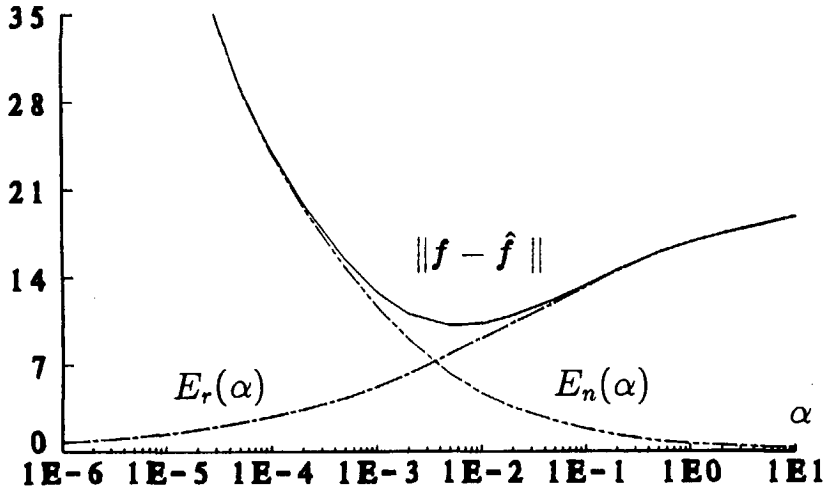


Figure 3.2: The noise magnification, regularization, and total restoration error as a function of the regularization parameter α.

(a) (b)

Figure 3.3: (a) Original Cameraman image; (b) Noisy defocused image.

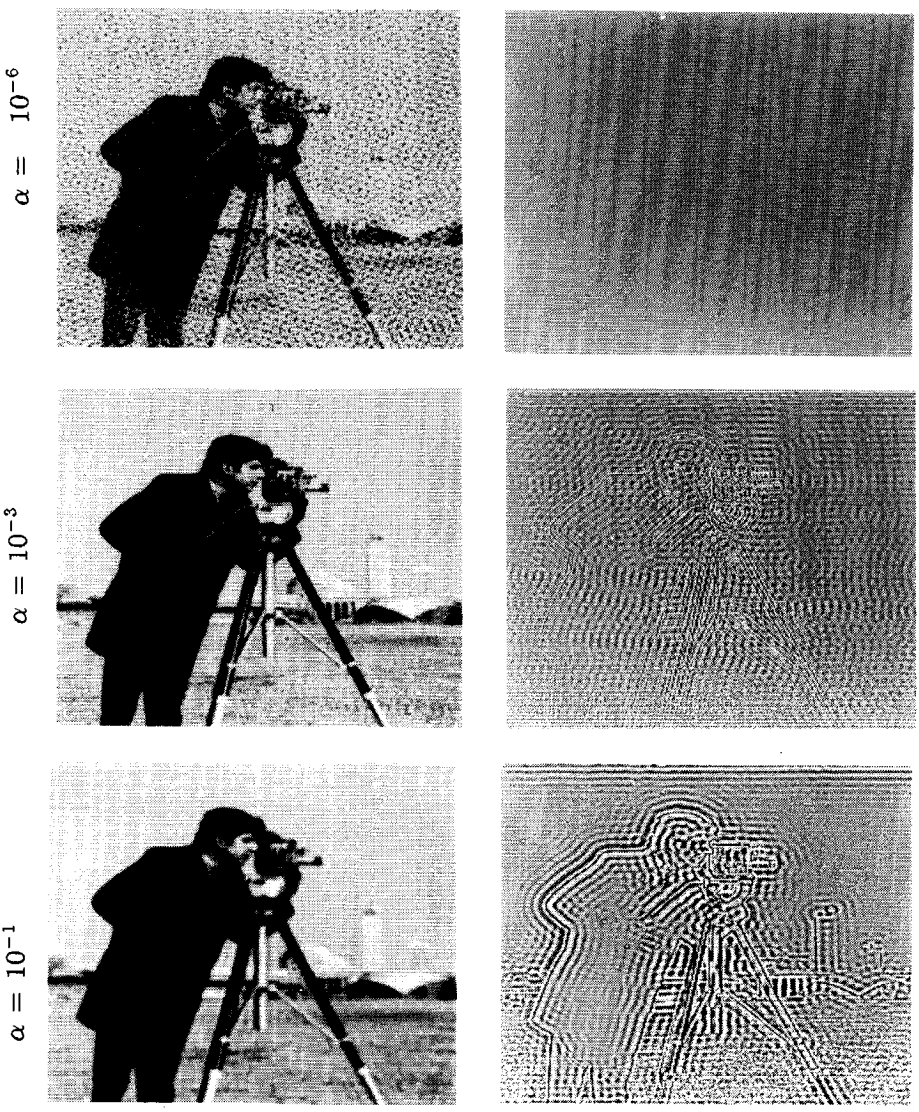


Figure 3.4: (a) Restoration result and (b) Regularization error (stretched for visibility) as a function of α.

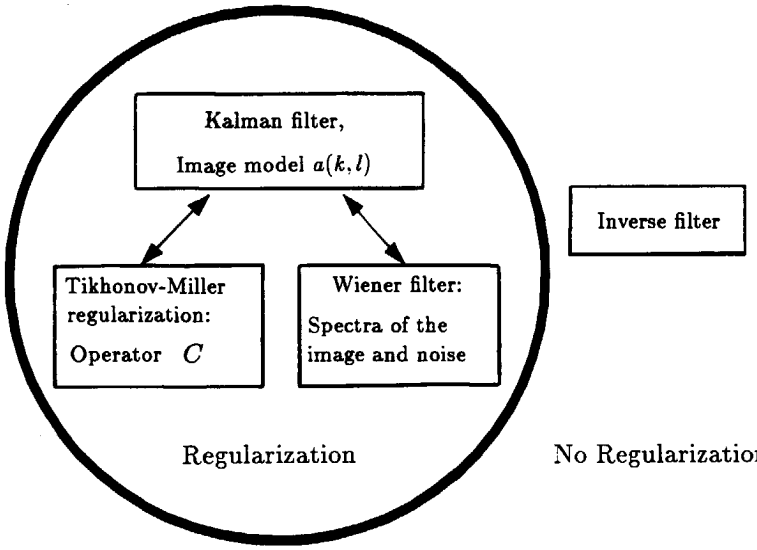


Figure 3.5: Relation between the various regularized filters

- $M = \alpha C^t C$, for the Tikhonov-Miller regularized filter (algebraic regularization).

We can formally relate the stochastic and algebraic restoration methods by relating the expressions for M through the use of the image model (3.2) (see Figure 3.5).

Since the image model (3.2) is assumed to be stable, the matrix $(I - A)$ is assured to be invertible. Therefore (3.2) can be rewritten as follows:

$$(I - A)f = v, \tag{3.37}$$

or

$$f = (I - A)^{-1}v. \tag{3.38}$$

From (3.38) the correlation matrix R_{ff} can be obtained:

$$R_{ff} = \mathrm{E}[ff^t] = \sigma_v^2 (I - A)^{-1}(I - A)^{-t}, \tag{3.39}$$

where we have assumed that v is white noise with variance σ_v^2. With the assumption of white observation noise w, the correlation matrix product as appearing in the Wiener filter is given by

$$M = R_{ww}R_{ff}^{-1} = \frac{\sigma_w^2}{\sigma_v^2}(I - A)^t(I - A). \tag{3.40}$$

Alternatively we can take the norm of both sides of (3.37), yielding

$$\|(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{f}\| = \|\boldsymbol{v}\| = \sigma_v. \tag{3.41}$$

Comparing (3.41) with (3.26) shows that $(\boldsymbol{I}-\boldsymbol{A})$ can very well serve as a regularizing operator $\boldsymbol{C}$. Substitution of the particular choice $\boldsymbol{C} = (\boldsymbol{I}-\boldsymbol{A})$ into $\boldsymbol{M}$ leads to

$$\boldsymbol{M} = \alpha \boldsymbol{C}^t\boldsymbol{C} = \alpha(\boldsymbol{I}-\boldsymbol{A})^t(\boldsymbol{I}-\boldsymbol{A}). \tag{3.42}$$

Clearly Eqs. (3.40) and (3.42) are identical if we set $\alpha = \epsilon^2/E^2 = (\sigma_w/\sigma_v)^2$, i.e. we use the regularization parameter according to Miller's method. This demonstrates that stochastic and algebraic restoration methods are in fact related concepts, merely differing by their choice of the regularizing operator.

It should be noted, however, that although any image model will lead to a valid regularizing operator via the relation $\boldsymbol{C} = (\boldsymbol{I}-\boldsymbol{A})$, not every regularizing operator will lead to a valid image model. This is due to the fact that image models need to be stable, while no such restriction was imposed on the regularizing operator. For example, the image model derived from the 2-D Laplacian operator in (3.34) is given by:

$$a(k,l) = \begin{bmatrix} & 0.25 & \\ 0.25 & 0.00 & 0.25 \\ & 0.25 & \end{bmatrix}. \tag{3.43}$$

By substitution of (3.43) into (2.15) it is straightforward to see that this is an unstable image model, because $B(z_1, z_2)$ has a zero at $(z_1, z_2) = (1, 1)$.

3.4 Multiple Constraints Restoration

3.4.1 Deterministic *A Priori* Constraints

In many image restoration problems there is *a priori* knowledge available about the original image that cannot be expressed in the form of a stabilizing functional such as (3.26). This deterministic knowledge, however, is often very useful to reduce the set of feasible solutions, in this way achieving an alternative form of regularization. We will shortly describe an iterative algorithm that is able to use such knowledge, provided that it can be expressed in the form of a closed convex set C_i [115], called a

constraint. A nonexpansive projection P_i can be associated with each constraint (see Appendix B). It maps all images which violate the constraint onto C_i, i.e.

$$P_i \boldsymbol{f} = \begin{cases} \boldsymbol{f}, & \text{if } \boldsymbol{f} \in C_i \\ \boldsymbol{h}, & \text{if } \boldsymbol{f} \notin C_i, \end{cases} \tag{3.44}$$

where

$$\|\boldsymbol{h} - \boldsymbol{f}\| \leq \|\boldsymbol{x} - \boldsymbol{f}\|, \qquad \forall \boldsymbol{x} \in C_i. \tag{3.45}$$

Some examples of deterministic constraints falling in this class are the following:

- The image has a bounded (e.g. nonnegative) intensity range,
- The object in the image has a finite support,
- There is a maximum value for the image energy,
- The structure of an entity in the image may be known *a priori*,
- The norm of the residual $\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}$ is bounded,
- An image model is known: the norm of the modeling error $\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{f}$ can be bounded.

Note that the Equations (3.25) and (3.28), on which the Tikhonov-Miller regularization (and therefore the entire class of linear regularized filters) is based, actually form two closed convex sets by themselves [88]. The proof of convexity and the derivation of the related projections of the above constraints can be found in several references, for example [87,88,91,100,103,104,115]. The power of the constraints is also discussed in these references.

3.4.2 Projections onto Convex Sets

In the multiple constraints restoration approach the inverse restoration problem (3.1) is regularized by enforcing the restoration result to satisfy as many constraints as possible, where each of the constraints has been defined as tightly as possible. The regularized restoration results can then be selected from the intersection C_0 of all constraint sets C_i. Clearly any image in that intersection will exhibit all of the features associated with all of the sets. If those convex sets all reflect desirable properties for the restored image, then any image in the intersection should be reasonable.

It is generally not possible to find an analytical expression for an image which satisfies a number of possibly nonlinear constraints. The theory of the projections onto convex sets [17,115] was developed to find an image in the intersection C_0 of m closed convex sets C_i, yielding the following iterative scheme:

$$\hat{f}_k = (P_1 P_2 \cdots P_m)^k \hat{f}_0. \tag{3.46}$$

The iterations (3.46) converge to a solution in C_0 for any initial estimate $\hat{f}_0$, unless the intersection C_0 is empty. In that case conflicting constraints are being enforced onto the restoration results, resulting in iterations that will not stabilize. The exact properties of the limiting solution $\hat{f}_\infty$ depend on the initial estimate, unless the intersection C_0 contains only a single element (which is extremely rare). Besides its use in image restoration, the iterations (3.46) have found wide applications in various other signal processing problems such as band-limited and space-limited extrapolation, phase and magnitude retrieval, and reconstruction from incomplete data [91].

Recent research has led to the extension of the method of projections onto convex sets to projections onto fuzzy sets [19]. In this method the "hard" boundaries of the deterministic constraint sets are replaced by fuzzy boundaries. As a consequence the sets to be used in the restoration procedure are easier to define and less sensitive to erroneous assumptions.

Chapter 4

Iterative Image Restoration

In the previous chapter we have described the problems involved in solving the ill-conditioned image restoration problem, and we have shown that several classical restoration filters can be classified as Tikhonov-Miller regularized methods. In this chapter we will consider the use of iterative methods in image restoration. Iterative procedures offer the advantage that no matrix inverses need to be implemented, and that additional deterministic constraints can be incorporated into the solution algorithms. Section 4.1 introduces the basic iterative restoration algorithm that forms the basis for most of the algorithms discussed in Chapters 4 and 5. It will be shown in Section 4.2 that terminating the iterations prior to convergence is a means for regularizing the restoration process. Section 4.3 presents a variation on the basic scheme which asymptotically produces the Tikhonov-Miller regularized solution. Finally, Section 4.4 is concerned with procedures for increasing the convergence speed of the iterative algorithms.

4.1 VanCittert's Iteration

4.1.1 Formulation of the Algorithm

The simplest of the iterative restoration methods has a long history. It goes back at least to the work of VanCittert [106] in the 1930's, and may in fact have even older antecedents. Iterative solution techniques have been applied to the (image) restoration problem by many researchers in recent years [38,39,45,46,47,49,51,55,64,65,68,69,79,81,83,84,90,103,104]. Although originally formulated for spatially invariant restoration, it can be applied to solve spatially variant problems as well (see Chapter 5). If we

neglect for a moment the noise contribution in (3.1), the following identity can be introduced which must hold for any real value of the parameter β:

$$\boldsymbol{f} = \boldsymbol{f} + \beta(\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}). \tag{4.1}$$

Applying the method of successive approximations to this equality yields the following iteration to compute a solution to (3.1):

$$\begin{aligned} \hat{\boldsymbol{f}}_0 &= 0, \\ \hat{\boldsymbol{f}}_{k+1} &= \hat{\boldsymbol{f}}_k + \beta(\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}_k) \\ &= \beta\boldsymbol{g} + (\boldsymbol{I} - \beta\boldsymbol{D})\hat{\boldsymbol{f}}_k. \end{aligned} \tag{4.2}$$

Different researchers refer to this iteration as the VanCittert [106], Bially [8], or Landweber [60,95] iteration, presumably because it has been independently discovered many times.

With an iterative method there are two important concerns — does it converge, and if so, to what limiting solution? By direct enumeration it is seen that[1]

$$\hat{\boldsymbol{f}}_{k+1} = \beta \sum_{r=0}^{k} (\boldsymbol{I} - \beta\boldsymbol{D})^r \boldsymbol{g}, \tag{4.3}$$

which can also be written as

$$\begin{aligned} \hat{\boldsymbol{f}}_{k+1} &= \beta(\boldsymbol{I} - (\boldsymbol{I} - \beta\boldsymbol{D}))^{-1}(\boldsymbol{I} - (\boldsymbol{I} - \beta\boldsymbol{D})^{k+1})\boldsymbol{g} \\ &= \boldsymbol{D}^{-1}(\boldsymbol{I} - (\boldsymbol{I} - \beta\boldsymbol{D})^{k+1})\boldsymbol{g}, \end{aligned} \tag{4.4}$$

provided that $\boldsymbol{D}$ is invertible. If we now assume that

$$\lim_{k \to \infty} (\boldsymbol{I} - \beta\boldsymbol{D})^{k+1} = \boldsymbol{O}, \tag{4.5}$$

which is a sufficient condition for convergence, the limiting solution is seen to be equal to

$$\hat{\boldsymbol{f}}_\infty = \boldsymbol{D}^{-1}\boldsymbol{g}. \tag{4.6}$$

This is the inverse filter of which we have said earlier that it yields useless solutions in the presence of noise. The iterative implementation of the inverse filter has, however, two advantages over the direct implementation.

[1]The analytical expressions given here and in the remainder of this chapter for $\hat{\boldsymbol{f}}_k$ assume that the initial estimate of the iterative process is equal to $\hat{\boldsymbol{f}}_0 = 0$. For nonsingular matrices, this assumption can be made without loss of generality.

First, it can be terminated prior to convergence, resulting in a restored image that will often not exhibit too much noise magnification. This aspect of iterative restoration will be analysed more thoroughly in the next section. The second advantage is that the inverse operator does not need to be implemented. Each iteration requires only that the blurring matrix itself be implemented.

4.1.2 Convergence Analysis

As with the Tikhonov-Miller regularization, we can gain greater understanding of the iteration (4.2) through an eigenvalue analysis. This will not only lead to a better understanding of the convergence condition (4.5), but will also explain why more satisfactory restoration results can be achieved by truncating the iterations prior to convergence.

As in Chapter 3, let $\{z_{uv}; u \in [0, M-1], v \in [0, N-1]\}$ denote the eigenvectors with unit length associated with $\boldsymbol{D}$, and let λ_{uv} denote the corresponding eigenvalues. By expanding $\hat{\boldsymbol{f}}_{k+1}$ in terms of these eigenvectors, and by substitution of the iteration (4.2) we get

$$\begin{aligned}
\hat{\boldsymbol{f}}_{k+1} &= \sum_{u,v} (\hat{\boldsymbol{f}}_{k+1}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv} \\
&= \sum_{u,v} \beta(\boldsymbol{g}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv} + (\boldsymbol{I} - \beta \boldsymbol{D}) \sum_{u,v} (\hat{\boldsymbol{f}}_k, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv} \\
&= \sum_{u,v} \left\{ \beta(\boldsymbol{g}, \boldsymbol{z}_{uv}) + (1 - \beta\lambda_{uv})(\hat{\boldsymbol{f}}_k, \boldsymbol{z}_{uv}) \right\} \boldsymbol{z}_{uv}. \qquad (4.7)
\end{aligned}$$

The restoration result after $k+1$ VanCittert iterations can now be written in terms of these eigenvectors and eigenvalues as

$$\begin{aligned}
\hat{\boldsymbol{f}}_{k+1} &= \beta \sum_{u,v} \left[\sum_{r=0}^{k} (1 - \beta\lambda_{uv})^r \right] (\boldsymbol{g}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv} \\
&= \sum_{u,v} \frac{1}{\lambda_{uv}} (1 - (1 - \beta\lambda_{uv})^{k+1}) (\boldsymbol{g}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv}. \qquad (4.8)
\end{aligned}$$

Here we have assumed that $\boldsymbol{D}$ is invertible so that none of its eigenvalues equals zero. As $k \to \infty$, the iterations converge to

$$\hat{\boldsymbol{f}}_{\infty} = \sum_{u,v} \frac{1}{\lambda_{uv}} (\boldsymbol{g}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv}, \qquad (4.9)$$

provided that

$$|1 - \beta\lambda_{uv}| < 1, \qquad \forall u, v. \qquad (4.10)$$

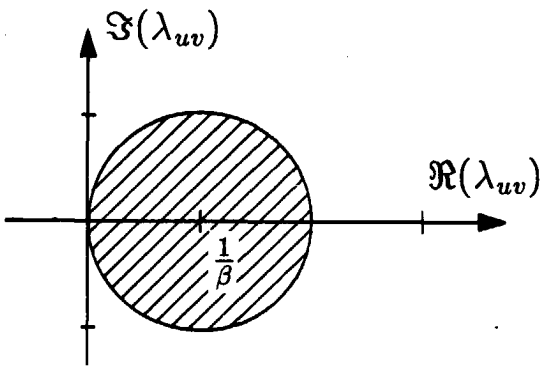


Figure 4.1: Region of the complex plane where the eigenvalues of the blurring matrix must lie for the VanCittert iteration to converge.

The result in (4.9) is the inverse filter (c.f. Eq. (3.5)). The convergence condition in (4.10) is equivalent to that given in (4.5). The interpretation of (4.10) is that all the eigenvalues of $\boldsymbol{D}$ must lie within the shaded area of the complex plane shown in Figure 4.1. It is recalled here that since we have assumed that $\boldsymbol{D}$ represents a circular convolution of the original image with the PSF of the image formation system, the eigenvalues λ_{uv} are equal to the coefficients of the DFT of $d(m,n)$. Therefore the condition (4.10) is very easy to check in practical situations.

From condition (4.10), it follows that a necessary convergence condition is:

$$\Re(\beta\lambda_{uv}) > 0, \qquad \forall u, v, \tag{4.11}$$

where $\Re(\cdot)$ denotes the real part operator. Since β is assumed to take real values only, this condition cannot be satisfied by blurs which have both positive and negative eigenvalues. Therefore the two important cases of motion and defocusing blur are excluded from the iterative process (4.2).

4.1.3 Reblurring Procedure

To overcome the convergence problem observed in the previous section, several authors have proposed to use a process called "reblurring" in the iterations (4.2) [49,84]. This is equivalent to applying the VanCittert procedure to the identity

$$\boldsymbol{f} = \boldsymbol{f} + \beta \boldsymbol{D}^t(\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}), \tag{4.12}$$

yielding the iteration

$$\begin{aligned} \hat{f}_0 &= 0, \\ \hat{f}_{k+1} &= \hat{f}_k + \beta D^t(g - D\hat{f}_k). \end{aligned} \tag{4.13}$$

An alternative way to derive the iteration (4.13) is from the viewpoint of minimizing the norm of the residual $g - D\hat{f}$, because for a good restoration result the blurred estimate $D\hat{f}$ should be approximately equal to the observed image g. Iterative minimization of the objective function

$$\Phi(\hat{f}) = \|g - D\hat{f}\|^2 \tag{4.14}$$

by the method of steepest descent [28] yields

$$\begin{aligned} \hat{f}_{k+1} &= \hat{f}_k + \beta r_k = \hat{f}_k - \frac{1}{2}\beta \nabla_f \Phi(f)\big|_{\hat{f}_k} \\ &= \hat{f}_k + \beta D^t(g - D\hat{f}_k), \end{aligned} \tag{4.15}$$

which is identical to (4.13). In (4.15) $r_k = -\frac{1}{2}\nabla_f \Phi(f)\big|_{\hat{f}_k}$ is called the steepest descent direction which points in the direction of the negative gradient of the objective function at $\hat{f}_k$. The parameter β controls the convergence of the iterations.

If we now apply a similar convergence analysis to the iteration (4.15), we get

$$\hat{f}_{k+1} = \sum_{u,v} \frac{1}{\lambda_{uv}}(1 - (1 - \beta|\lambda_{uv}|^2)^{k+1})(g, z_{uv})z_{uv}. \tag{4.16}$$

For this iteration scheme convergence is seen to require

$$|1 - \beta|\lambda_{uv}|^2| < 1, \qquad \forall u, v, \tag{4.17}$$

which is equivalent to the requirement that $\lambda_{uv} \neq 0$, or in other words, D must be invertible. Since we have already restricted the discussion to invertible matrices because hardly any blurring matrix has exact spectral zeros, we may conclude that the iteration (4.15) can be used for practically any restoration problem, including motion and defocusing blurs.

From (4.17) we can also determine for which values of β the iterations (4.15) are convergent:

$$0 < \beta < \frac{2}{|\lambda_{max}|^2}, \tag{4.18}$$

where λ_{max} is the eigenvalue λ_{uv} of maximal norm. From (4.16) it is seen immediately that if the iterations (4.15) converge, the limiting solution is again the inverse filter. Therefore algorithms of the type (4.15) are called iterative inverse restoration filters.

4.2 Regularization via Truncated Iterations

In order to gain insight in the behavior of the iterations (4.15) with regard to the ill-posedness of the restoration problem, we consider the properties of the regularization and noise magnification error for a finite number of iterations, i.e. when the iterative process is truncated prior to convergence.

On the basis of the expansion (4.16), we can determine an upperbound for the difference between the restoration result after k iterations and the original image. To this end we substitute the image formation model (3.1) into (4.16), yielding

$$\begin{aligned}\hat{\boldsymbol{f}}_{k+1} &= \sum_{u,v} \frac{1}{\lambda_{uv}}(1-(1-\beta|\lambda_{uv}|^2)^{k+1})(\boldsymbol{D}\boldsymbol{f}+\boldsymbol{w},\boldsymbol{z}_{uv})\boldsymbol{z}_{uv} \\ &= \sum_{u,v}(1-(1-\beta|\lambda_{uv}|^2)^{k+1})\left[(\boldsymbol{f},\boldsymbol{z}_{uv})+\frac{1}{\lambda_{uv}}(\boldsymbol{w},\boldsymbol{z}_{uv})\right]\boldsymbol{z}_{uv}. \quad (4.19)\end{aligned}$$

From this expression we obtain the following error bound:

$$\begin{aligned}&\|\hat{\boldsymbol{f}}_{k+1}-\boldsymbol{f}\| \leq \\ &\sum_{u,v}|1-\beta|\lambda_{uv}|^2|^{k+1}|(\boldsymbol{f},\boldsymbol{z}_{uv})| + \sum_{u,v}\frac{1}{|\lambda_{uv}|}|1-(1-\beta|\lambda_{uv}|^2)^{k+1}|\,|(\boldsymbol{w},\boldsymbol{z}_{uv})| \\ &= \mathrm{E}_r(k)+\mathrm{E}_n(k). \quad (4.20)\end{aligned}$$

This error bound has two terms, of which the first term on the right hand side is related only to the original image $\boldsymbol{f}$, and the second term is related only to the observation noise $\boldsymbol{w}$. Both terms are a function of the iteration index k. The first term in (4.20) can be made arbitrarily small by letting $k \to \infty$. This term represents the degree of restoring the image, or in other words the regularization error. The second term in (4.20) is the noise magnification error, which grows to infinity if $k \to \infty$ because of the small eigenvalues of $\boldsymbol{D}$. On the other hand, if k is set to zero (no iterations performed), the noise magnification error is minimized, but the regularization error will take a large value. Since $\mathrm{E}_r(k)$ decreases as a function of k and $\mathrm{E}_n(k)$ increases, their sum attains a minimum after a certain number of iterations. Unfortunately, this number is usually not known in advance and needs to be determined by visual inspection of the restored images as the iterations progress. From the above discussion it can also be seen that the reciprocal value of the iteration index k has the same function as the regularization parameter α in the Tikhonov-Miller regularization approach.

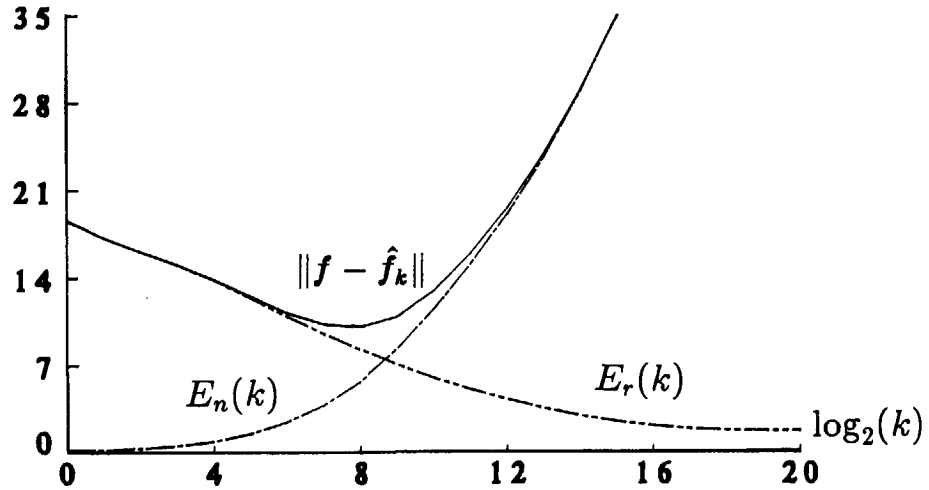


Figure 4.2: The restoration, regularization and noise magnification error as a function of the number of iterations.

Example of Truncated Iterations

To illustrate the effect of truncating the iterations prior to convergence and to show the effect of the "regularization parameter" $1/k$, we consider the restoration of the noisy defocused image in Figure 3.3b using the iterations (4.15) with $\beta = 1$. The two components $\mathrm{E}_r(k)$ and $\mathrm{E}_n(k)$ of the error bound in (4.20) are plotted in Figure 4.2 as a function of the iteration index k. The total restoration error is also shown in this figure. For this example the optimum restoration result occurs after approximately 250 iterations, and has an improvement in SNR of $\eta_{\mathrm{SNR}} = 5.7$ dB.

The restoration results and the regularization errors are also shown in Figure 4.3 as a set of images. The images on the left show the restoration results after 15, 250 and 4000 iterations, while the corresponding regularization errors are shown in the right column. It is interesting to note that the better visual result seems to appear after 4000 iterations, indicating that the improvement in SNR does not correlate well with the subjective judgement of the image quality. Another observation to make is that the regularization errors seem to occur mainly in the regions near edges in

the image. The noise magnification error on the other hand degrades the whole image but is not related to the structures in the original image. We will come back to these effects in Chapter 5.

4.3 Iterative Tikhonov-Miller Solution

So far we have encountered two means for regularizing the image restoration problem, namely by Tikhonov-Miller's method and by truncating the iterative computation of the inverse filter. Several authors have proposed to combine these regularization methods for three purposes [1,46,47,55]: (i) the iterative computation of the Tikhonov-Miller regularized solution might be easier than the direct evaluation of (3.31), (ii) the use of an iterative scheme allows for additional constraints to be imposed on the solution (see Chapter 5), (iii) the combination of two different regularization techniques provides an additional degree of freedom.

To derive the iterative Tikhonov-Miller regularized restoration algorithm, we use the method of steepest descent to minimize the objective function $\Phi(\hat{f})$ in (3.27). This gives the following iterations:

$$\begin{aligned}\hat{f}_{k+1} &= \hat{f}_k + \beta r_k = \hat{f}_k - \frac{1}{2}\beta \nabla_f \Phi(f)|_{\hat{f}_k} \\ &= \hat{f}_k - \beta((D^t D + \alpha C^t C)\hat{f}_k - D^t g) \\ &= (I - \alpha\beta C^t C)\hat{f}_k + \beta D^t(g - D\hat{f}_k). \end{aligned} \tag{4.21}$$

This iteration reduces to the (reblurred) VanCittert iteration if $\alpha = 0$ (no Tikhonov-Miller regularization).

The regularized solution after $k+1$ iterations is given in terms of the eigenvalues λ_{uv} and μ_{uv} of D and C, respectively, as

$$\begin{aligned}\hat{f}_{k+1} &= \beta \sum_{u,v} \left[\sum_{r=0}^{k} \lambda_{uv}^* (1 - \beta(|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2))^r \right] (g, z_{uv}) z_{uv} \\ &= \sum_{u,v} \frac{\lambda_{uv}^*}{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2} (1 - (1 - \beta(|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2))^{k+1}) (g, z_{uv}) z_{uv}. \end{aligned} \tag{4.22}$$

From this the convergence conditions for (4.21) follow immediately:

$$|1 - \beta(|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2)| < 1, \qquad \forall u, v. \tag{4.23}$$

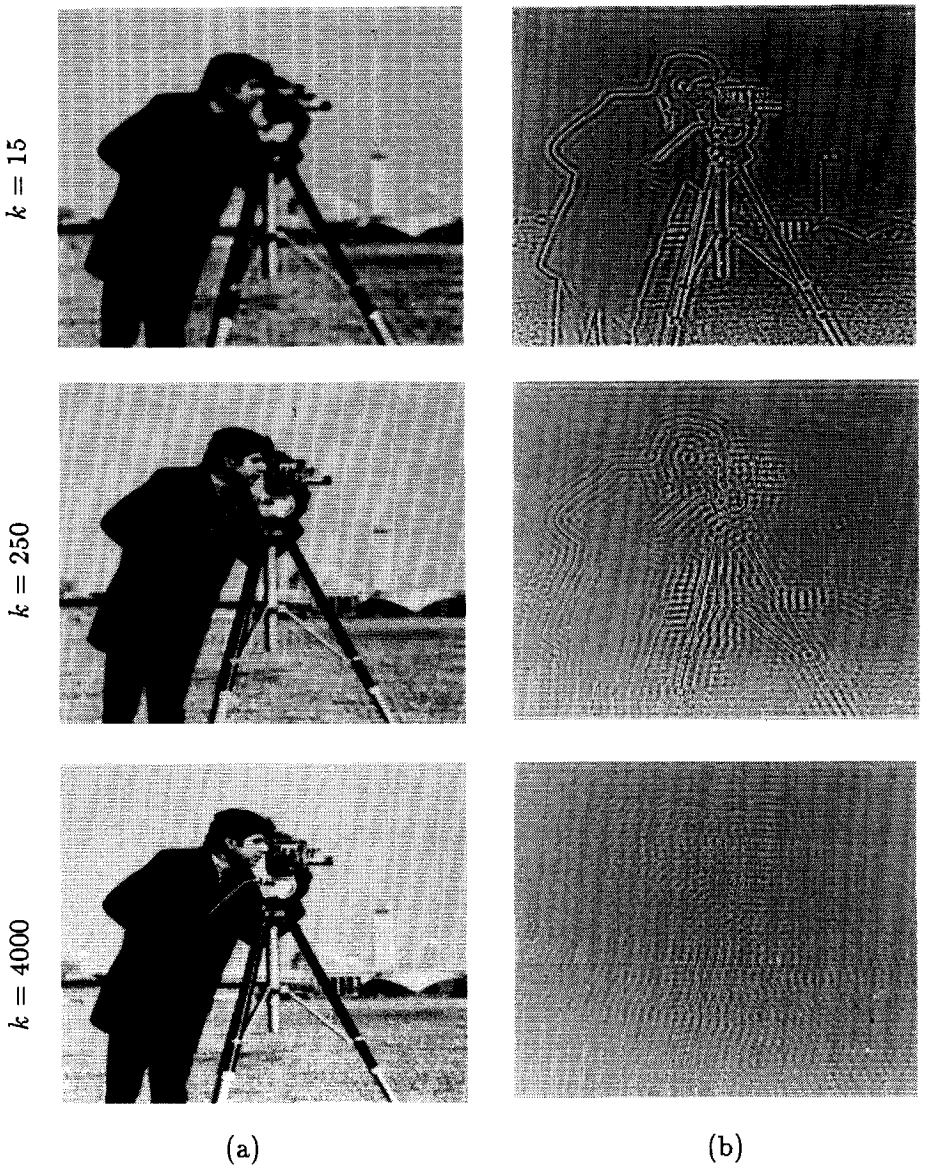


Figure 4.3: Example showing the effect of truncating the iterations prior to convergence: (a) Restoration result and (b) the regularization error as a function of the number of iterations.

In order to ensure convergence, β must satisfy

$$0 < \beta < \frac{2}{|\rho_{max}|}, \tag{4.24}$$

where ρ_{max} is the largest eigenvalue of the matrix $\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C}$, i.e. $\rho_{max} = \max_{u,v}\{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2\}$.

If the iterations converge, the limiting solution is given by (3.32). Again, if the iterations are terminated prior to convergence, there will be two sources of error: one due to the fact that convergence has not been achieved and that the solution is Tikhonov-Miller regularized, and one due to the amplified observation noise:

$$\begin{aligned} \|\hat{\boldsymbol{f}}_{k+1} - \boldsymbol{f}\| &\leq \sum_{u,v} \frac{\alpha|\mu_{uv}|^2 + |\lambda_{uv}|^2|1 - \beta(|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2)|^{k+1}}{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2}|(\boldsymbol{f}, \boldsymbol{z}_{uv})| \\ &+ \sum_{u,v} \frac{|\lambda_{uv}||1 - (1 - \beta(|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2))^{k+1}|}{|\lambda_{uv}|^2 + \alpha|\mu_{uv}|^2}|(\boldsymbol{w}, \boldsymbol{z}_{uv})| \\ &= \mathrm{E}_r(\alpha, k) + \mathrm{E}_n(\alpha, k). \end{aligned} \tag{4.25}$$

This expression reduces to several of the ones already derived if the number of iterations is increased to ∞, or if the regularization parameter α is set to zero. The behavior of the error terms will be illustrated by an example later on.

In [45,46] Katsaggelos *et al.* observe that the term $(\boldsymbol{I} - \alpha\beta\boldsymbol{C}^t\boldsymbol{C})$ in (4.21) behaves like a low-pass filter, suppressing the noise in the iterates. The characteristics of this term are obviously related to the properties of the original image because the regularizing operator $\boldsymbol{C}$ is closely related to the image model $\boldsymbol{A}$. It was proposed to compress this term into a single low-pass filter $\boldsymbol{C}_s$, which would reflect spectral *a priori* knowledge or knowledge about the image model of the original image. The iterations (4.21) then become:

$$\hat{\boldsymbol{f}}_{k+1} = \boldsymbol{C}_s\hat{\boldsymbol{f}}_k + \beta\boldsymbol{D}^t(\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}_k). \tag{4.26}$$

If we let ν_{uv} denote the eigenvalues of $\boldsymbol{C}_s$, and assume that $\boldsymbol{C}_s$ has the same set of eigenvectors as $\boldsymbol{D}$, then the convergence conditions for the iterations (4.26) are given by

$$|\nu_{uv} - \beta|\lambda_{uv}|^2| < 1. \tag{4.27}$$

Two particular choices for $\boldsymbol{C}_s$ were also suggested. In the first place it was proposed to use the image model coefficients $a(k,l)$ to define $\boldsymbol{C}_s = \boldsymbol{A}$. A

second possibility that was given, is to use a noise smoothing Wiener filter, which assumes the form

$$C_s = \boldsymbol{R}_{ff}(\boldsymbol{R}_{ff} + \boldsymbol{R}_{ww})^{-1}, \tag{4.28}$$

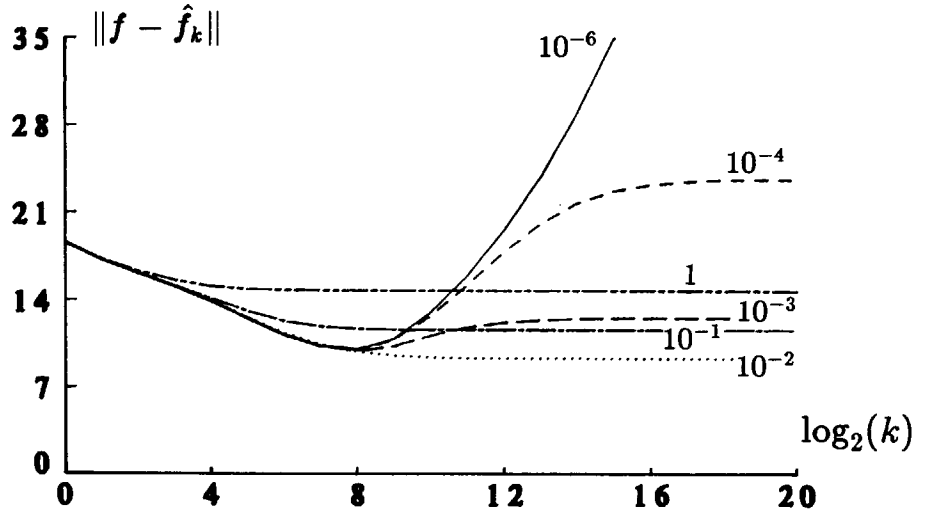
where $\boldsymbol{R}_{ff}$ and $\boldsymbol{R}_{ww}$ are the autocorrelation matrices of the original image and the noise, respectively. In this case the limiting solution of (4.26) is quite similar to the parametric Wiener restoration filter [3]. In practice, the construction of the Wiener smoothing filter or the choice of the image model coefficients may sometimes be easier than the selection of a regularizing operator and the related bound E.

Example of Tikhonov-Miller Regularized Iterations

This example shows the effect of the Tikhonov-Miller regularization when implemented via an iterative scheme, and the behavior of the regularization and noise magnification error. Consider again the restoration of the defocused cameraman image in Figure 3.3b. A number of regularized restoration results were formed iteratively. The 2-D Laplacian regularizing operator was used, and $\beta = 1.0$. The total restoration error, and the regularization and noise magnification errors are plotted as a function of both the regularization parameter α and the iteration index k in Figure 4.4. The restoration results associated with these plots are given in Figure 4.5 as a function of α and k. The two columns in this figure correspond to $k = 15$ and 4000, while the three rows correspond to $\alpha = 10^{-6}, 10^{-3}$ and 10^{-1}. Observe that the top-most row corresponds closely to the results of the (reblurred) VanCittert iterative algorithm, while the right column corresponds to the results of the Tikhonov-Miller regularized restoration filter.

4.4 Implementations with Faster Convergence

The iterative restoration procedures that we have discussed so far are based on the minimization of an objective function $\Phi(\boldsymbol{f})$ using the method of steepest descent. It is well known, however, that this method may converge very slowly. There has been a lot of work in optimization theory to develop iterative algorithms which converge in less iterations than required for the method of steepest descent [28,63]. In this section we will



(a)

Figure 4.4: (a) Total restoration error, (b) Regularization error, (c) Noise magnification error as a function of α and k.

first analyze the behavior of the steepest descent iteration (4.21), and next describe several more efficient implementations of this iterative restoration algorithm [23,51,65,68,69,79,83,90].

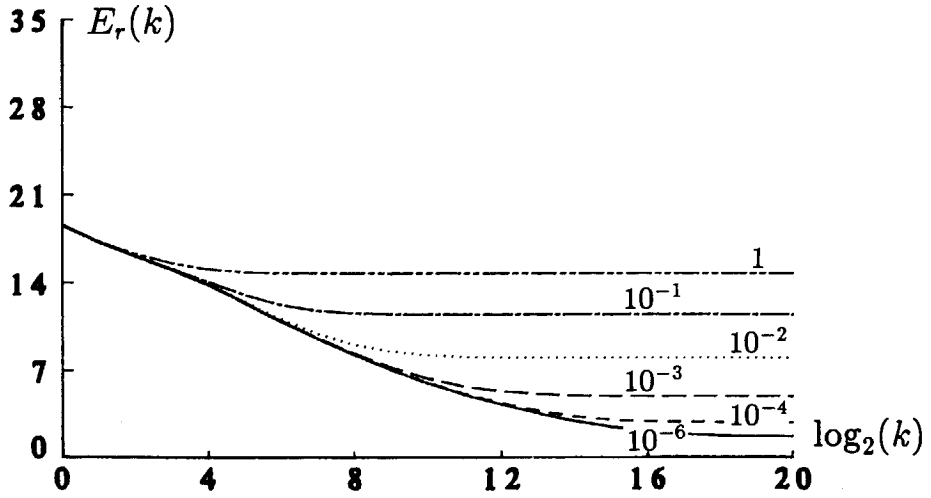
4.4.1 Analysis of Convergence Speed

The convergence speed of an iterative algorithm towards its limiting solution can be quantified by its convergence order and its convergence rate. An iterative scheme converges geometrically with order $\mathcal{R}$ if the error $\|\hat{f}_k - \hat{f}_\infty\|$ for k sufficiently large can be given as [63]:

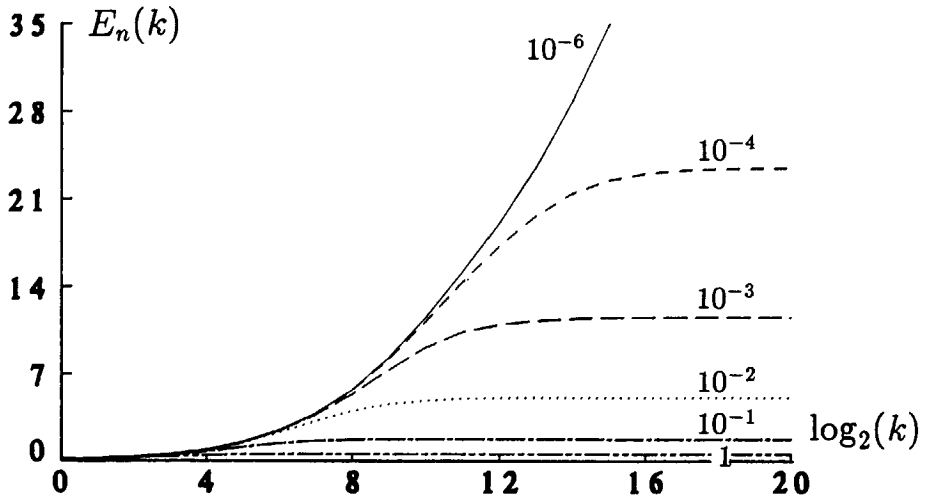
$$\|\hat{f}_{k+1} - \hat{f}_\infty\| \approx \kappa \|\hat{f}_k - \hat{f}_\infty\|^{\mathcal{R}}, \text{ for } \mathcal{R} \geq 1. \tag{4.29}$$

Here $\kappa \geq 0$ is the convergence rate. A large value of κ corresponds to a slow convergence, while a small value indicates fast convergence. Further, the larger the convergence order is, the faster the algorithm converges. For $\mathcal{R} = 1$ the process is said to converge linearly, in which case (4.29) is conveniently written as:

$$\|\hat{f}_k - \hat{f}_\infty\| \leq \kappa^k \|\hat{f}_0 - \hat{f}_\infty\|, \tag{4.30}$$



(b)



(c)

Figure 4.4: Continued.

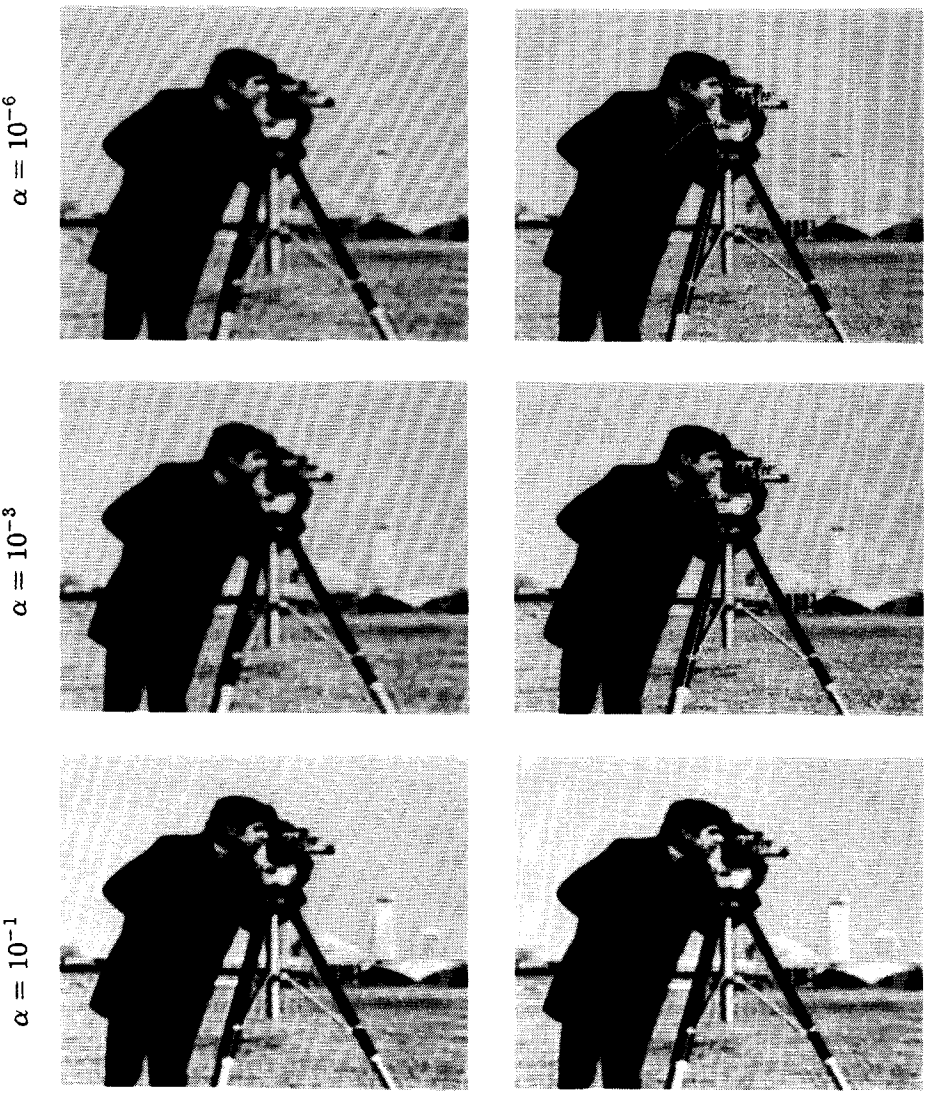


Figure 4.5: Restoration results as a function of α and k.

and $0 \leq \kappa \leq 1$. For $\kappa = 0$ the convergence is said to be superlinear (or in other words the iterative process terminates within a finite number of iterations), while for $\kappa = 1$ the convergence is sublinear.

Iterative schemes of the type (4.2), (4.15) and (4.21) can all be regarded as members of the following generic iteration, which minimizes a quadratic objective function $\Phi(\boldsymbol{f})$ by the method of steepest descent:

$$\hat{\boldsymbol{f}}_{k+1} = \hat{\boldsymbol{f}}_k + \beta(\boldsymbol{h} - \boldsymbol{B}\hat{\boldsymbol{f}}_k). \tag{4.31}$$

For a nonsingular matrix $\boldsymbol{B}$ we have $\hat{\boldsymbol{f}}_\infty = \boldsymbol{B}^{-1}\boldsymbol{h}$. The following relation can then be established:

$$\begin{aligned} \hat{\boldsymbol{f}}_{k+1} - \hat{\boldsymbol{f}}_\infty &= (\boldsymbol{I} - \beta\boldsymbol{B})\hat{\boldsymbol{f}}_k + \beta\boldsymbol{h} - \hat{\boldsymbol{f}}_\infty \\ &= (\boldsymbol{I} - \beta\boldsymbol{B})\hat{\boldsymbol{f}}_k + \beta\boldsymbol{B}\hat{\boldsymbol{f}}_\infty - \hat{\boldsymbol{f}}_\infty \\ &= (\boldsymbol{I} - \beta\boldsymbol{B})(\hat{\boldsymbol{f}}_k - \hat{\boldsymbol{f}}_\infty), \qquad (4.32) \\ \Rightarrow \quad \|\hat{\boldsymbol{f}}_{k+1} - \hat{\boldsymbol{f}}_\infty\| &\leq \|(\boldsymbol{I} - \beta\boldsymbol{B})\|\|(\hat{\boldsymbol{f}}_k - \hat{\boldsymbol{f}}_\infty)\|. \qquad (4.33) \end{aligned}$$

Here $\|\boldsymbol{I} - \beta\boldsymbol{B}\|$ denotes the norm of the matrix, which is given by its largest eigenvalue. Thus if we let ρ_{min} be the eigenvalue of $\boldsymbol{B}$ with smallest norm, (4.33) becomes

$$\|\hat{\boldsymbol{f}}_{k+1} - \hat{\boldsymbol{f}}_\infty\| \leq |1 - \beta\rho_{min}|\|\hat{\boldsymbol{f}}_k - \hat{\boldsymbol{f}}_\infty\|. \tag{4.34}$$

Hence, iterative schemes of the type (4.31) converge linearly with a convergence rate of $\kappa = |1 - \beta\rho_{min}|$.

For the iteration (4.21) we have $(\boldsymbol{I} - \beta\boldsymbol{B}) = (\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C}))$, and ρ_{min} is the smallest eigenvalue of $(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C})$. Since the smallest eigenvalue is nearly always close to zero in an image restoration application, the iterations (4.21) converge slowly. For instance, in the example in Section 4.3 the optimal value of $\alpha = 0.01$ yields a minimal eigenvalue of $|\rho_{min}| = 3.2\,10^{-4}$; the iterative scheme associated with this value of the regularization parameter has a convergence rate of $\kappa = 0.9997$ ($\beta = 1.0$). Observe that in general the convergence rate depends on the regularization parameter α, because ρ_{min} is a function of α. If α is increased, i.e. if the restoration problem is regularized more strongly, ρ_{min} attains a larger value, by which the speed of convergence increases as well.

In the iterations (4.21) β was assumed to have a fixed value satisfying the

condition in (4.24). But since β controls the convergence rate, as follows from (4.34), it is desirable to optimize its value at each iteration step.

The optimal value of β at the iteration k, denoted by β_k, can be derived by minimizing $\Phi(\hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{r}_k)$, i.e. choose that β_k along the path $\hat{\boldsymbol{f}}_{k+1} = \hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{r}_k$ which minimizes $\Phi(\hat{\boldsymbol{f}}_{k+1})$. Here $\Phi(\boldsymbol{f})$ is the objective function in (3.27), and $\boldsymbol{r}_k$ the steepest descent direction associated with this objective function. The solution to this minimization problem is given by

$$\beta_k = \frac{\|\boldsymbol{r}_k\|^2}{\|\boldsymbol{D}\boldsymbol{r}_k\|^2 + \alpha\|\boldsymbol{C}\boldsymbol{r}_k\|^2}. \tag{4.35}$$

For the optimized method of steepest descent the convergence is still linear, but its convergence rate is now given by [82]

$$\|\hat{\boldsymbol{f}}_k - \hat{\boldsymbol{f}}_\infty\| \leq \text{constant} \left(\frac{|\rho_{max}| - |\rho_{min}|}{|\rho_{max}| + |\rho_{min}|} \right)^k, \tag{4.36}$$

where ρ_{max} and ρ_{min} are the largest and smallest eigenvalues of $(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C})$, respectively. By using the relation (4.24), and by the fact that $|\rho_{min}| \ll |\rho_{max}|$, we find

$$\frac{|\rho_{max}| - |\rho_{min}|}{|\rho_{max}| + |\rho_{min}|} = \frac{1 - \left|\frac{\rho_{min}}{\rho_{max}}\right|}{1 + \left|\frac{\rho_{min}}{\rho_{max}}\right|} \approx 1 - 2\left|\frac{\rho_{min}}{\rho_{max}}\right| \leq |1 - \beta\rho_{min}|. \tag{4.37}$$

Hence, the above inequality shows that the optimized method of steepest descent will converge faster than the method of steepest descent with a fixed value for β. Although the optimization procedure for β increases the convergence speed of the algorithm, the improvements are usually moderate and may not justify the efforts involved. An experimental comparison between the two methods will be given later in this section.

4.4.2 Method of Conjugate Gradients

Motivated by the desire to achieve more rapid convergence, the method of conjugate gradients has been successfully used in optimization theory [63]. Conjugate direction methods, which were originally introduced for purely quadratic problems, can be viewed as a special orthogonal expansion of the solution of the minimization problem. This expansion is generated by making use of information from previous iteration steps. This section focuses on the use of the conjugate gradients method to minimize the objective

function (3.27). One of the advantages of this method is its convergence in a finite number of iteration steps when exact arithmetic (no rounding errors) is used, i.e. the convergence is superlinear [63]. When non-exact arithmetic is used or the problem is non-quadratic, the method will no longer converge in a finite number of iterations because the conjugacy condition will no longer hold. It has been experimentally shown, however, that the conjugate gradients method converges always faster than the method of steepest descent, while the computational complexity is not significantly increased.

The basic form of the conjugate gradients algorithm, which thus represents an alternative to the iteration (4.21), is given by

$$\begin{aligned} \boldsymbol{r}_k &= -\frac{1}{2}\nabla_f \Phi(f)|_{\hat{\boldsymbol{f}}_k} = -(\boldsymbol{D}^t\boldsymbol{D} + \alpha \boldsymbol{C}^t\boldsymbol{C})\hat{\boldsymbol{f}}_k + \boldsymbol{D}^t\boldsymbol{g}, \\ \boldsymbol{p}_k &= \boldsymbol{r}_k + \gamma_k \boldsymbol{p}_{k-1}, \\ \hat{\boldsymbol{f}}_{k+1} &= \hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{p}_k. \end{aligned} \tag{4.38}$$

In this scheme $\hat{\boldsymbol{f}}_k$ is modified in the direction of the vector $\boldsymbol{p}_k$ instead of the steepest descent direction $\boldsymbol{r}_k$. The parameter γ_k controls the conjugacy of the subsequent directions $\boldsymbol{p}_k$. For $\gamma_k \to 0$ the iterations (4.38) reduce to (4.21). It can be shown that the optimal value for γ_k and β_k are given by [63]:

$$\gamma_k = \frac{\|\boldsymbol{r}_k\|^2}{\|\boldsymbol{r}_{k-1}\|^2}, \tag{4.39}$$

$$\beta_k = \frac{\boldsymbol{r}_k^t \boldsymbol{p}_k}{\|\boldsymbol{D}\boldsymbol{p}_k\|^2 + \alpha\|\boldsymbol{C}\boldsymbol{p}_k\|^2}. \tag{4.40}$$

The conjugate gradients method converges linearly, with a convergence rate given by [82]

$$\|\hat{\boldsymbol{f}}_k - \hat{\boldsymbol{f}}_\infty\| \leq \text{constant} \left(\frac{\sqrt{|\rho_{max}|} - \sqrt{|\rho_{min}|}}{\sqrt{|\rho_{max}|} + \sqrt{|\rho_{min}|}} \right)^k. \tag{4.41}$$

Although (4.41) is a rather loose upper bound to the convergence rate of the conjugate gradients iteration (for example, it does not show that the method of conjugate gradients converges (theoretically) within a finite number of iterations), it is still useful for a comparison with the optimized

method of steepest descent. Since $|\rho_{max}| > |\rho_{min}|$, it is straightforward to prove that

$$\frac{\sqrt{|\rho_{max}|} - \sqrt{|\rho_{min}|}}{\sqrt{|\rho_{max}|} + \sqrt{|\rho_{min}|}} < \frac{|\rho_{max}| - |\rho_{min}|}{|\rho_{max}| + |\rho_{min}|}, \tag{4.42}$$

which shows that the conjugate gradients method has a higher convergence speed than the method of steepest descent. In addition to this, experimental results exhibit a convergence speed much larger than indicated by the bound in (4.41), as we will see in the next section. Observe that the difference in computational complexity between the method of steepest descent and the conjugate gradients algorithm merely consists of the computation of $\boldsymbol{p}_k$ and γ_k, which is insignificant compared with the other computations required within a single iteration step.

4.4.3 Iteration Method with Higher Convergence Order

Although the method of the conjugate gradients has a considerably higher convergence speed than the method of steepest descent, both techniques essentially converge linearly. Recently a number of papers have described the use of an iterative restoration method, which has a convergence order larger than one [23,51,68,69,90]. As is shown in [51,69], these algorithms are based on a Taylor expansion of the matrix to be inverted. The basic form of the iterations, which require a "double iteration", is given by

$$\begin{aligned} \hat{\boldsymbol{f}}_0 &= \beta \boldsymbol{D}^t \boldsymbol{g}, \\ \boldsymbol{B}_0 &= \boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{D} + \alpha \boldsymbol{C}^t\boldsymbol{C}), \\ \hat{\boldsymbol{f}}_{k+1} &= \sum_{j=0}^{\mathcal{R}_k - 1} \boldsymbol{B}_k^j \hat{\boldsymbol{f}}_k \qquad (\mathcal{R}_k \geq 2), \\ \boldsymbol{B}_{k+1} &= \boldsymbol{B}_k \boldsymbol{B}_k \cdots \boldsymbol{B}_k = \boldsymbol{B}_k^{\mathcal{R}_k}. \end{aligned} \tag{4.43}$$

Here $\mathcal{R}_k$ determines the convergence order in each iteration step. Sufficient convergence conditions are again given by (4.24).

The convergence behavior of the above iterations can be analyzed by comparing the explicit expressions for the $(k+1)^{\text{st}}$ iterate of (4.43) with the one obtained from the standard steepest descent iteration (4.21) with fixed β. If we use a constant value of $\mathcal{R}_k$ throughout the iterations ($\mathcal{R}_k = \mathcal{R}$),

we can write for the $(k+1)^{\text{st}}$ iterate of (4.43):

$$\hat{\boldsymbol{f}}_{k+1} = \sum_{r=0}^{\mathcal{R}^{k+1}-1} (\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C}))^r \beta \boldsymbol{D}^t\boldsymbol{g}. \tag{4.44}$$

For the iterations (4.21) we get after $k+1$ iterations:

$$\hat{\boldsymbol{f}}_{k+1} = \sum_{r=0}^{k} (\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{C}))^r \beta \boldsymbol{D}^t\boldsymbol{g}. \tag{4.45}$$

By comparing (4.44) and (4.45) it can be seen that the two procedures compute exactly the same solutions. However, the steepest descent algorithm requires $\mathcal{R}^{k+1} - 1$ iterations to obtain the same solution that the iterations (4.43) reach after only $k+1$ iterations. The convergence order for (4.43) is therefore equal to $\mathcal{R}$ [69,90]. The extra expense for the increased convergence order is more computations in a single iteration step, and extra storage capacity for the (block-circulant) matrix $\boldsymbol{B}_k$. The optimal value for $\mathcal{R}$ depends therefore strongly on the actual implementation of the iterations. However, if a very large convergence order is used, only a few iterations are required to achieve convergence, but the evaluation of the matrix $\boldsymbol{B}_k$ may become extremely computationally involving. Practical implementations typically use a value of $\mathcal{R} = 2$ or $\mathcal{R} = 3$.

Convergence Comparison between the Algorithms

In this example we consider the convergence behavior of the steepest descent algorithm with fixed and optimized value of β, the conjugate gradients algorithm, and the iterative algorithm with $\mathcal{R}-$th order convergence. The noisy defocused image in Figure 3.3b was restored using the iterative Tikhonov-Miller restoration algorithm with a 2-D Laplacian regularizing operator, and $\alpha = 0.01$. The smallest and largest eigenvalue have a value of $3.2\,10^{-4}$ and 1.00, respectively. In Figure 4.6 the convergence behavior is shown by plotting $\|\hat{\boldsymbol{f}}_k - \hat{\boldsymbol{f}}_\infty\|$ as a function of the iteration index k. In Table 4.1 the theoretical convergence rates are listed for the method of steepest descent and conjugate gradients, which are computed from the bounds given in this section. Also listed in this table are the experimental convergence rates, computed from the first 200 iterations. It is observed that the practical convergence rates are smaller than the theoretical upper bounds, i.e. the algorithms converge faster than theoretically predicted. For an increasing number of iterations, however, the practically computed

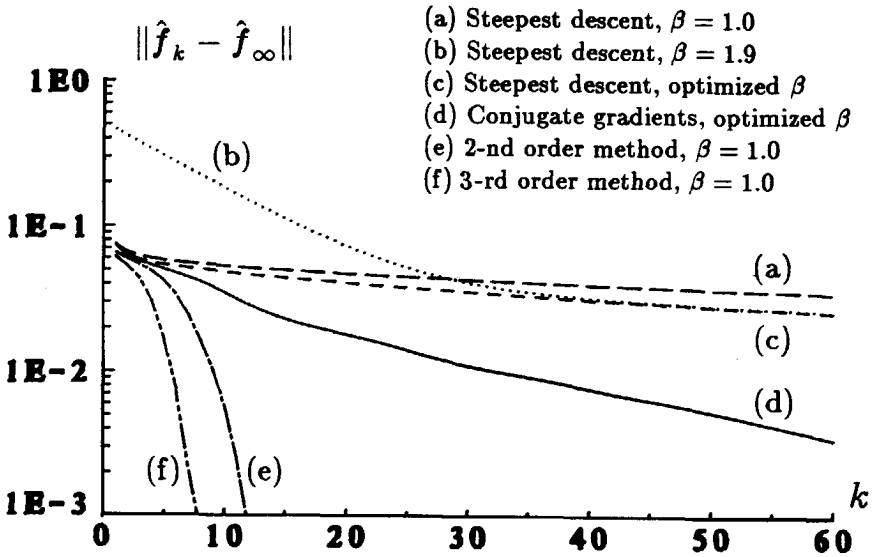


Figure 4.6: Convergence behavior of the method of steepest descent, conjugate gradients method, and higher order method.

convergence rates approach the theoretical bounds more closely. The measured convergence rate for the higher order methods are averaged values over the first 8 iterations.

Another way to evaluate the convergence performance of an iterative scheme is shown in Table 4.2. This table lists the theoretical and practical number of iterations that are required to reduce $\|\hat{f}_k - \hat{f}_\infty\|$ by a factor of 10. Clearly the Tables 4.1 and 4.2 and Figure 4.6 illustrate that the higher order methods are most efficient, followed by the conjugate gradients method, and the steepest descent methods. It is also observed that optimizing β in the method of steepest descent does not significantly improve the convergence rate when compared with using a fixed value of β.

Table 4.1: Experimental and theoretical convergence rate.

Method	β	$\mathcal{R}$	Experimental Rate	Theoretical Rate
Steepest descent	1.0	1	0.99658	0.99967
Steepest descent	1.9	1	0.99602	0.99938
Steepest descent	optimized	1	0.99590	0.99934
Conjugate gradient	optimized	1	0.94631	0.96453
2-nd order	1.0	2	≈ 50	-
3-rd order	1.0	3	≈ 1000	-

Table 4.2: Number of iterations required to reduce the error by a factor of 10.

Method	β	$\mathcal{R}$	Experimental	Theoretical
Steepest descent	1.0	1	670	6970
Steepest descent	1.9	1	570	3710
Steepest descent	optimized	1	560	3490
Conjugate gradient	optimized	1	41	64
2-nd order	1.0	2	10	13
3-rd order	1.0	3	6	8

Chapter 5

Image Restoration with Ringing Reduction

In the preceding chapters we have seen that the Tikhonov-Miller regularization, the method of truncated iterations, or the combination of these two, could be used to stabilize the inversion of the ill-conditioned blurring matrix $\boldsymbol{D}$ and to suppress the noise magnification in this way. As a result another type of error occurred in the restored images, namely the regularization error. Restoration algorithms are often criticized because of these artifacts (called "ringing", superwhites and superblacks, or overshoots and undershoots). The reason for this is twofold: in the first place most original scenes are known to be without ringing effects, i.e. the restoration result is not consistent with our *a priori* knowledge, and in the second place ringing artifacts reduce both the visual and measurable quality of the restored images. In Section 5.1 we will first have a closer look at the origin of these artifacts in linear spatially invariant (LSI) restoration schemes [55]. Section 5.2 continues with the description of an adaptive iterative restoration algorithm which incorporates two methods to reduce the ringing artifacts [55], namely (i) the use of *a priori* constraints, and (ii) the local regulation of the balance between the noise magnification and regularization error.

The basic form of the algorithm presented in Section 5.2 originates from a nonlinear optimization based on the method of steepest descent. The use of the more rapidly converging method of conjugate gradients will be discussed in Section 5.3. This chapter is concluded with various experiments which demonstrate the significance of ringing reduction in image restoration.

5.1 Analysis of Ringing Artifacts

5.1.1 The Error Spectrum

Let us consider a general linear spatially invariant (LSI) regularized restoration filter $\boldsymbol{H}(\alpha)$ that is characterized by its discrete Fourier coefficients $H(u,v;\alpha)$ [1], and is indexed by a regularization parameter α. Algorithms falling in this category are for instance the Tikhonov-Miller regularized filters and the truncated iterative process. Because of the regularization $H(u,v;\alpha)$ deviates from the inverse filter $D^{-1}(u,v)$. This deviation can be measured by the error spectrum $E(u,v;\alpha)$, defined as

$$E(u,v;\alpha) = 1 - H(u,v;\alpha)D(u,v). \tag{5.1}$$

Through some straightforward manipulations we arrive at the following expression for the restored image:

$$\begin{aligned}\hat{F}(u,v) &= H(u,v;\alpha)G(u,v) = H(u,v;\alpha)\left[D(u,v)F(u,v) + W(u,v)\right] \\ &= F(u,v) - E(u,v;\alpha)F(u,v) + \frac{1 - E(u,v;\alpha)}{D(u,v)}W(u,v). \end{aligned} \tag{5.2}$$

In agreement with the error analyses in Chapters 3 and 4 we find that the restoration error consists of two terms:

$$\hat{F}(u,v) - F(u,v) = -E(u,v;\alpha)F(u,v) + \frac{1 - E(u,v;\alpha)}{D(u,v)}W(u,v), \tag{5.3}$$

or

$$\begin{aligned}|\hat{F}(u,v) - F(u,v)| &\leq |E(u,v;\alpha)F(u,v)| + \left|\frac{1 - E(u,v;\alpha)}{D(u,v)}\right| |W(u,v)| \\ &= \mathrm{E}_r(\alpha) + \mathrm{E}_n(\alpha). \end{aligned} \tag{5.4}$$

In the spatial domain these relations correspond to

$$\begin{aligned}\hat{f}(i,j) - f(i,j) &= -e(i,j;\alpha) * f(i,j) \\ &\quad + (1 - e(i,j;\alpha)) * d^{-1}(i,j) * w(i,j), \end{aligned} \tag{5.5}$$

[1] In this section we will conveniently make use of discrete Fourier transforms when dealing with spatially invariant restoration procedures.

and

$$|\hat{f}(i,j) - f(i,j)| \leq |e(i,j;\alpha) * f(i,j)| + |(1 - e(i,j;\alpha)) * d^{-1}(i,j) * w(i,j)|, \quad (5.6)$$

where the error sequence $e(i,j;\alpha)$ has been defined as the inverse DFT of $E(u,v;\alpha)$. As in the previous chapters $\mathrm{E}_r(\alpha)$ is called the regularization error, and $\mathrm{E}_n(\alpha)$ is the noise magnification error. In this section we will study the relations (5.3)-(5.6) and the qualitative properties of the two error terms.

General properties of $\mathrm{E}_r(\alpha)$ and $\mathrm{E}_n(\alpha)$ that follow from the concept of regularization, are [73,99]:

$$\lim_{\alpha \to 0} \mathrm{E}_r(\alpha) = 0, \quad (5.7)$$

$$\lim_{\alpha \to 0} \mathrm{E}_n(\alpha) \to \infty, \quad (5.8)$$

$$\lim_{\sigma_w \to 0} \mathrm{E}_n(\alpha) = 0, \quad \text{for any } \alpha. \quad (5.9)$$

The relations (5.7) and (5.8) have been illustrated by the examples in Section 3.3, 4.2 and 4.3. For a given amount of noise in the observed blurred image, quantified by the variance σ_w^2, the optimal solution $\hat{f}(i,j)$ is theoretically determined by the value of α which minimizes the total restoration error $\mathrm{E}_r(\alpha) + \mathrm{E}_n(\alpha)$. For many linear spatially invariant restoration filters there exists at least one optimal value of α, since $\mathrm{E}_r(\alpha)$ is a monotonically increasing function of α and $\mathrm{E}_n(\alpha)$ is a monotonically decreasing function. The optimal value of the regularization parameter is SNR dependent, because the noise magnification error $\mathrm{E}_n(\alpha)$ is proportional to σ_w^2. Unfortunately, the optimal value of α is never known in practice, because $\mathrm{E}_r(\alpha)$ in (5.4) cannot be evaluated without the availability of the original image.

The noise magnification error has a global degrading effect because it results from a globally random phenomenon — the observation noise. On the other hand, the regularization error is a function of $f(i,j)$, and its effects will therefore be related strongly to the local structures encountered within the image. Ringing artifacts, a structure dependent phenomenon, are therefore attributable to the regularization error.

5.1.2 Relation between the Error Spectrum and Ringing Artifacts

The relationship between the regularization error and ringing can be made more clear by studying the properties of the error spectrum and the error sequence. Unfortunately, general expressions that define these properties can hardly be found because they depend on the specific restoration filter considered, the type of blur and the severity of this blur. However, for a wide range of blurs and restoration filters of interest, at least one and often two of the following general characteristics will hold:

(i) since $D(u,v)$ is generally some type of low-pass filter, $E(u,v;\alpha)$ has the shape of a high-pass filter, and

(ii) $E(u,v;\alpha)$ has the shape of a band-pass filter near zero crossings of $D(u,v)$, with $E(u,v;\alpha) = 1$ for true spectral zeros. Since practical blurs of interest have multiple zero crossings, which are often approximately equally spaced (such as motion and defocusing blurs), $E(u,v;\alpha)$ takes the form of a multiple band-pass filter.

Both properties follow from the fact that balancing the regularization and noise magnification error requires $(1 - E(u,v;\alpha))$ to cut out the inverse filtered noise $D^{-1}(u,v)W(u,v)$ particularly in those spectral regions where $|D(u,v)| \to 0$. It is well known, that filtering an image with a high-pass filter with characteristic frequency ω_c, will give rise to ringing artifacts with a period of $2\pi/\omega_c$ samples (or Gibb's oscillations as they are usually called in this context). These artifacts, however, become far more pronounced when $E(u,v;\alpha)$ contains band-pass sections which are approximately equally spaced with distance ω_c (such as happens with motion blur). In conclusion, the combination of the above mentioned characteristics of $E(u,v;\alpha)$ in particular will give rise to severe and widely distributed ringing artifacts in restored images.

In the sequel we study the error spectra and error sequences of the two types of LSI restoration filters described in the previous chapters. We select linear motion blur over L pixels along the horizontal axis as the image blur, because this type of blur is known to represent a worst-case situation with respect to the severity and distribution of the ringing effects. In Figure 5.1a the cameraman image is shown, which has been degraded by linear motion blur with $L = 8$ and by additive white Gaussian noise with SNR=30 dB. The restoration result using the Tikhonov-Miller regularized

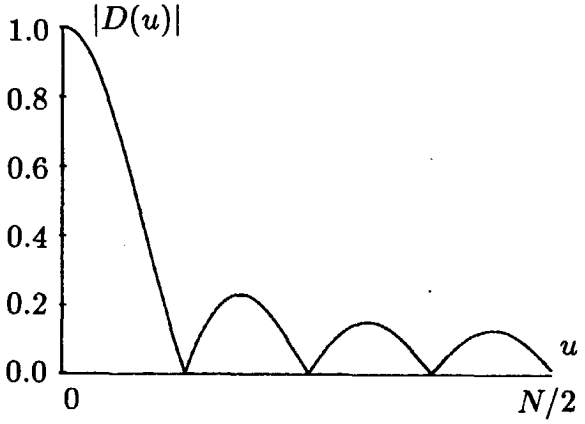
(a) (b)

Figure 5.1: (a) Cameraman image degraded by horizontal linear motion blur ($L = 8$) and noise (SNR=30 dB); (b) Tikhonov-Miller restoration result with severe ringing artifacts.

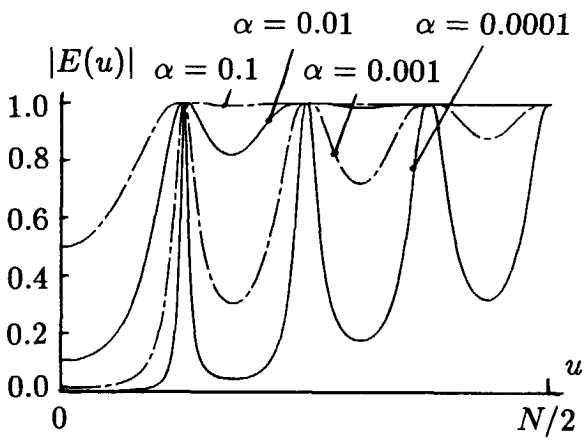
filter (with $C(u,v) =$ 2-D Laplacian filter, and $\alpha = 5.10^{-2}$) is given in Figure 5.1b, in which the ringing is clearly present. In Figure 5.2a the modulus of the sinc-shaped transfer function of the blur has been plotted. From Eq. (3.31) the error spectrum of the Tikhonov-Miller regularized filter can be obtained straightforwardly:

$$E(u,v;\alpha) = \left[1 + \frac{|D(u,v)|^2}{\alpha |C(u,v)|^2}\right]^{-1}. \tag{5.10}$$

The error spectrum for various values of the regularization parameter α has been plotted in Figure 5.2b, while the corresponding error sequences are shown in Figure 5.2c. Observe that particularly when the error spectrum tends to have peaks, the error sequence is dominated by positive impulses at $j = \pm kL$ $(k = 0, 1, 2, \cdots)$ (Figure 5.2d). From (5.5) we see that in the spatial domain the regularization error is a convolution of $-e(i,j;\alpha)$ with $f(i,j)$. In flat regions of the image this convolutional action has hardly any effect, but in the vicinity of steep intensity transitions it will cause negative versions of the intensity transitions to appear at distances $j = \pm kL$ $(k = 1, 2, \cdots)$. This explains why spurious oscillations were observed at regular distances from the edges in the regularization error and associated restored images in the experiments in Chapters 3 and 4.

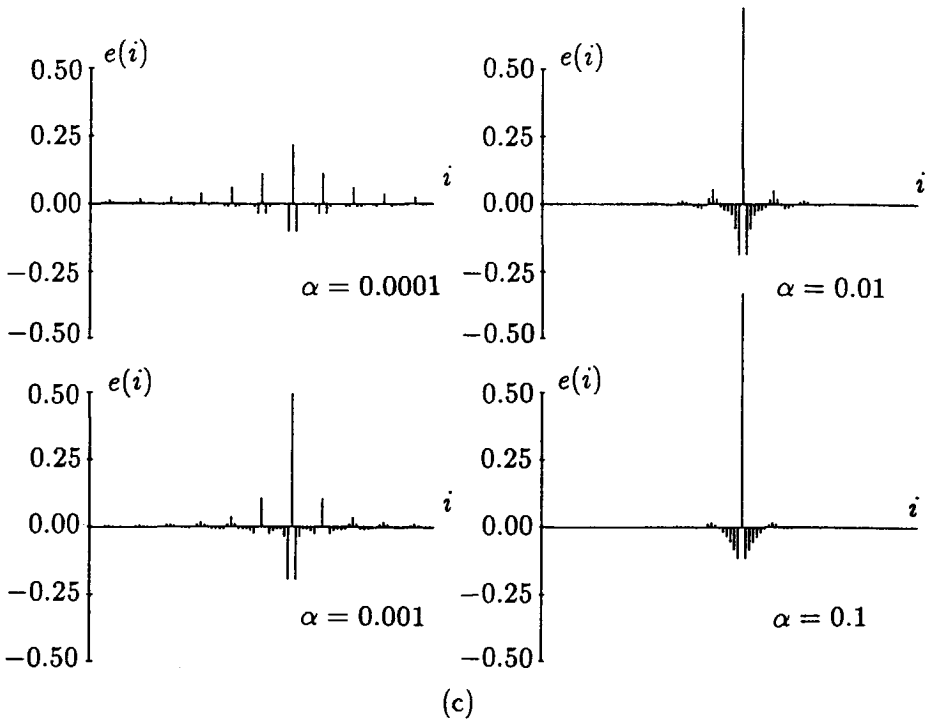


(a)

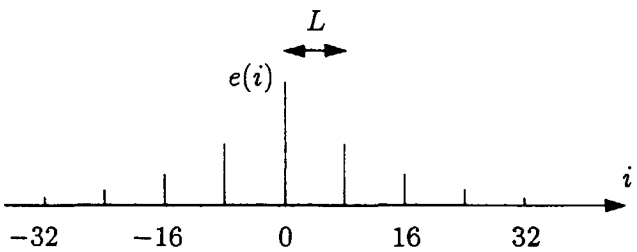


(b)

Figure 5.2: Properties of the regularization error of the Tikhonov-Miller regularized filter for linear motion blur over 8 pixels: (a) Modulus of the transfer function; (b) Modulus of the error spectra; (c) Error sequences; (d) Typical behavior of dominant impulses in an error sequence.



(c)



(d)

Figure 5.2: Continued.

The severity of the ringing artifacts (determined by the distribution and the size of the positive spikes in $e(i,j;\alpha)$) depends on the value of α. Since the optimal value of α is SNR-dependent, the severity of ringing artifacts is a function of the SNR as well. An interesting detail of this relation is that the ringing is negligible for both very small values of α (high SNR) as well as for large values of α (low SNR), although $E_r(\alpha)$ itself is a monotonically increasing function of α. Clearly for low SNRs the restoration filter has degenerated to a low-pass noise smoothing filter, in which the regularization error manifests itself no longer as ringing, but as severe blurring of edges.

The error spectrum for the iterative inverse restoration filter as formulated in (4.15) and (4.16) is given by:

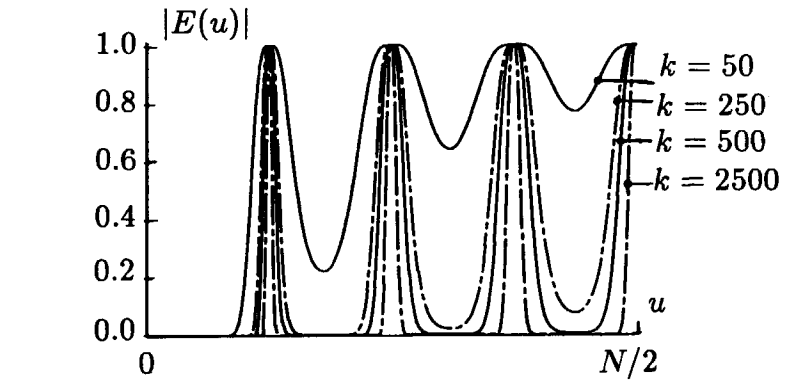
$$E(u,v;\frac{1}{k}) = (1-\beta|D(u,v)|^2)^k. \tag{5.11}$$

The error spectrum and the error sequence for different values of the iteration index k are shown in Figure 5.3. Observe that particularly for a small number of iterations (corresponding to a large value of the associated regularization parameter) the ringing effects will be relatively severe.

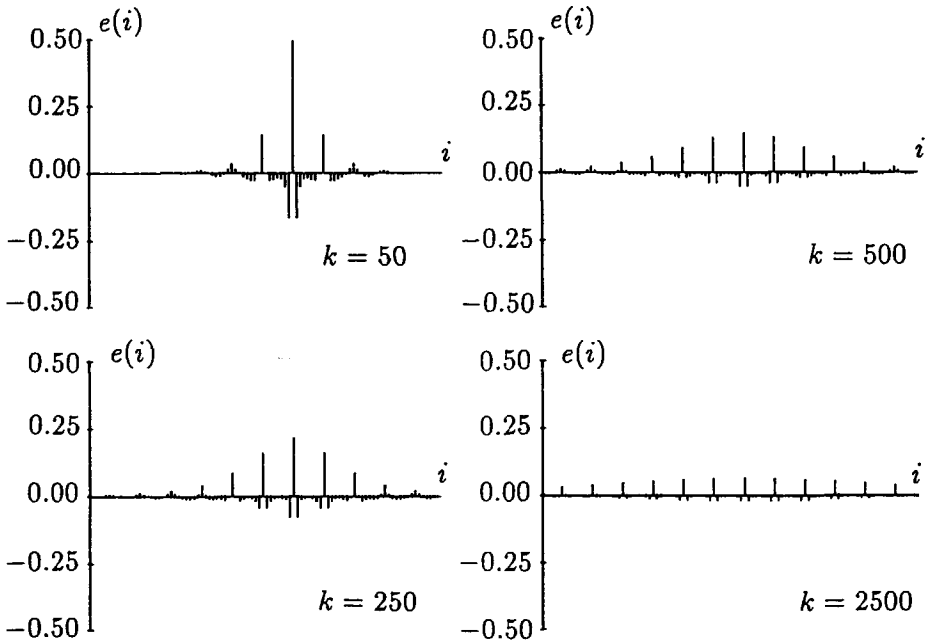
In conclusion, the above analysis shows that ringing effects near sharp intensity transitions in the restored images are attributable to the regularization. Since in LSI restoration filters the behavior of the regularization and noise magnification errors is globally controlled by (5.3) or (5.5), the appearance of excessively amplified noise, ringing effects or a combination of these two is inevitable. We therefore have to consider restoration methods which are not based on LSI filters.

5.1.3 Ringing Reduction Methods

In image restoration it is usually known that most original scenes are without ringing. One way to reduce ringing artifacts is, therefore, through the use of *a priori* knowledge of the original image. This can be done by enforcing the restored image to satisfy certain (possibly nonlinear) constraints. If the image data, for example, consists of blurred bright point sources against a black background (such as occurs in astronomical imaging), then the ringing manifests itself as negative values. A positivity constraint on the restoration can then be used to replace all negative values by zero, thus preventing the ringing from happening.



(a)



(b)

Figure 5.3: Properties of the regularization error of the iterative scheme (4.15) for linear motion over 8 pixels: (a) Modulus of the error spectra; (b) Error sequences.

The method of projections onto convex sets, described in Section 3.4, is particularly suitable for ringing reduction when the constraints can be made tight, i.e. when there is a lot of *a priori* knowledge describing the original image very accurately. Unfortunately, such powerful constraints can often not be assumed when dealing with the restoration of more complicated signals, such as images of natural scenes. Then the use of constraints alone is usually insufficient to significantly reduce ringing.

Another technique for reducing ringing locally regulates the balance between the regularization and noise magnification error. This adaptation depends on the local image properties, such as its edge content. In this way the restoration filter becomes spatially variant, and may be specified as

$$h(i,j;k,l;\alpha) = (1 - e(i,j;k,l;\alpha)) * d^{-1}(i,j). \tag{5.12}$$

Adaptive restoration focuses on designing $h(i,j;k,l;\alpha)$ in such a way that the total restoration error resulting from (5.12) is smaller than the one resulting from the nonadaptive restoration filter. As a result the regularization error is reduced as well. In designing the adaptive restoration filter, it should be kept in mind that the complexity of $h(i,j;k,l;\alpha)$ must be acceptable from a computational point of view. It is, for example, not feasible to explicitly define a different restoration filter for every position in the image because of the computation and storage capacity such an approach would require. In the next section an adaptive iterative method will be described that defines spatially variant restoration through the use of weighting matrices.

We next consider the required behavior of $e(i,j;k,l;\alpha)$ in edgy regions and in relatively smooth areas. Near sharp intensity transitions the regularization error dominates the noise magnification error, while in the more gradual areas the convolution of $f(i,j)$ with $-e(i,j;k,l;\alpha)$ has hardly any effect, so that $\mathrm{E}_n(\alpha)$ dominates $\mathrm{E}_r(\alpha)$. Consequently, by regularizing the edgy regions less strongly, the local regularization error, and hence the severity of the ringing, is reduced. At the same time resolution enhancement is achieved. When the regularization is reduced, the noise magnification is increased because these two errors are counterproductive. Fortunately, however, it is known from psychophysical experiments that the visibility of noise is greatly masked by sharp intensity transitions, whereas blurring generally appears to be unacceptable in this context [2]. This noise masking effect validates decreasing the regularization error and increasing the noise magnification error near edges. On the other hand blurring is

acceptable in nearly constant regions of the image, but noise magnification is not. Hence, in these regions of the image the restoration process has to concentrate on noise smoothing instead of removing the blur.

5.2 Constrained Adaptive Iterative Restoration

5.2.1 Introduction

Restoration filters which are implemented in the Fourier domain, such as the Tikhonov-Miller regularized filter, are unsuitable for incorporating *a priori* constraints and adaptive restoration. Therefore much research has recently been focused towards the modification of existing recursive and iterative restoration procedures.

In [98] Tekalp *et al.* describe a multiple model edge-adaptive Kalman filter in two dimensions, in which a number of image models is used to filter the image in agreement with the local edge orientation. Maximum *a posteriori* decision logic was used to choose between various models at each position in the image. In [4] Angwin follows a related approach, but this method is based on assigning an image model to an entire image region. In this way different local image properties lead to different local image models. In [41] Jeng *et al.* proposed to remove the local mean from the image and to normalize the residual. In this way the image is homogenized (i.e. the sharp intensity transitions are taken out of the image). A Kalman filter was again used to restore the resulting normalized residual image. Any of these implementations outperform the non-adaptive Kalman restoration filters, but none of them can incorporate additional *a priori* constraints.

Adaptive iterative restoration was first suggested by Ichioka *et al.* in [38]. In their approach the convergence speed of the iterative approximation of the inverse filter in (4.13) was locally regulated on the basis of the image content. In smooth regions the convergence speed was made low, while in edgy regions the restoration process converged fast. This approach can be interpreted as running more iterations in edgy regions than in smooth regions. As a result the associated regularization parameter, which is proportional to $1/k$, takes large values in smooth regions (strong regularization) and small values near edges (hardly any regularization).

In [46,47] Katsaggelos *et al.* considered the spatially variant implemen-

tation of the iterative Tikhonov-Miller regularized filter in (4.21). They replaced the term $(\boldsymbol{I} - \alpha\beta\boldsymbol{C}^t\boldsymbol{C})$ in (4.21) by a spatially varying noise smoothing filter. The tuning of this filter was based on the noise masking principle. This adaptive restoration procedure is closely related to the one that will be described in this section.

In [89] Sezan *et al.* proposed an adaptive restoration algorithm that uses the projection onto convex sets (POCS) algorithm. In their approach the restoration result is constrained to be "close" to the blurred image in smooth image regions, in this way preventing ringing artifacts. They also considered using the adaptive approach to be presented in this section, in the setting of POCS. It was concluded that, although it is conceptually possible to achieve adaptive restoration in this way, the practical implementation is extremely computationally demanding.

In the following we will describe an iterative image restoration algorithm, in which both adaptivity and *a priori* constraints are used to suppress ringing artifacts [55]. It can be considered an extension of the iterative restoration algorithms that were described in the previous chapters. In the remainder of this chapter we will make again use of the compact matrix-vector notations for the purpose of notational convenience.

5.2.2 *A Priori* Knowledge

In image restoration we usually have the disposal of an estimate of the norm of the noise present in the blurred image. A set of feasible solutions can then be defined by restricting the norm of the residual $\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}$ to this noise norm, which we denote by ϵ. To define the adaptive restoration algorithm the norm of the residual is taken in a weighted space:

$$\|\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}\|_R = \left[(\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}})^t\boldsymbol{R}(\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}})\right]^{\frac{1}{2}} \leq \epsilon. \tag{5.13}$$

Here $\boldsymbol{R}$ is a $MN \times MN$ diagonal matrix containing weight coefficients $r_{ij} \in [0, 1]$, where r_{ij} is associated with the pixel at the position (i, j). The computation of the values for r_{ij} will be discussed later in this section.

The Tikhonov-Miller regularization, as described in Chapter 3, provides the second kind of *a priori* knowledge. To regulate locally the trade-off between the noise magnification error and the regularization error, we propose imposing an upper bound onto the norm of the filtered image $\boldsymbol{C}\hat{\boldsymbol{f}}$ in a weighted space:

$$\|\boldsymbol{C}\hat{\boldsymbol{f}}\|_S = \left[(\boldsymbol{C}\hat{\boldsymbol{f}})^t\boldsymbol{S}(\boldsymbol{C}\hat{\boldsymbol{f}})\right]^{\frac{1}{2}} \leq E. \tag{5.14}$$

Here S is again a $MN \times MN$ diagonal weighting matrix containing positive coefficients $s_{ij} \in [0,1]$. The bound E is assumed to be known *a priori*.

Finally, the third kind of *a priori* knowledge is a (combination of) deterministic constraint(s). The set of solutions C described by this constraint has to be convex and closed. Associated with this set is the nonexpansive projection P. The solution to the restoration problem is required to be an element of this set C, hence it has to be a fixed point of the projection P (Appendix B).

The weight coefficients r_{ij} and s_{ij} in (5.13) and (5.14), respectively, have to be chosen on the basis of prior knowledge about the original image in order to minimize the restoration error. Such information is, however, hardly ever available, which makes the computation of the optimal weight coefficients impossible for practical situations of interest. For this reason we have to adopt a more practical point of view. For example, we can base the coefficients on a local variance measure computed from the blurred image itself [46,47,55], or compute them from a preliminary non-adaptive restoration result [55]. Qualitatively, we observe that in order to reduce ringing artifacts by decreasing the regularization near sharp intensity transitions we have $s_{ij} \ll 1$, and $r_{ij} \approx 1$ (enforcing inverse filtering). On the other hand, to smooth the noise in the more gradual regions we have $s_{ij} \approx 1$ and $r_{ij} \ll 1$.

It is worthwhile noticing that the weighting matrix $\boldsymbol{R}$ may also account for the nonstationarity of the noise. If the noise has a local variance $\sigma_w^2(i,j)$ at the position (i,j), the weight coefficients r_{ij} are assigned the reciprocal value of $\sigma_w^2(i,j)$, i.e. $\boldsymbol{R} = \boldsymbol{R}_{ww}^{-1}$, where $\boldsymbol{R}_{ww}$ is the (diagonal) autocorrelation matrix of the noise. In a worst-case situation some image elements may have been corrupted in such a serious way (for example due to shot noise or recording errors) that these outliers must be excluded from the restoration process. This is achieved by assigning the value zero to the corresponding weight coefficients r_{ij}.

5.2.3 Formulation of the Algorithm

Following the same line of reasoning as in Section 3.3, we can compute a solution which satisfies (5.13) and (5.14) by Miller's approach. The two inequalities are combined into a single quadrature formula:

$$\Phi(\hat{\boldsymbol{f}}) = \|\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}\|_R^2 + \alpha\|\boldsymbol{C}\hat{\boldsymbol{f}}\|_S^2 \leq 2\epsilon^2, \tag{5.15}$$

where the regularization parameter has the fixed value $\alpha = (\epsilon/E)^2$. The adaptive Tikhonov-Miller regularized solution minimizes $\Phi(\hat{f})$, and is given by the normal equations:

$$(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\hat{\boldsymbol{f}} = \boldsymbol{D}^t\boldsymbol{R}\boldsymbol{g}. \tag{5.16}$$

We require $(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$ to be a nonsingular matrix yielding a unique solution to (5.16). This is not a severe restriction since exact spectral zeros occur rarely in practical situations of interest. Further, the extension to singular matrices is straightforward by considering pseudo-inverses [72] instead of true inverse matrices.

It is now observed that the computation of $\hat{\boldsymbol{f}}$ requires the matrix $(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$ of size $MN \times MN$ to be inverted. Because this matrix represents a space-variant operator, we cannot reduce the computational complexity by applying the standard diagonalization procedure for block-circulant matrices, i.e. Fourier domain filtering. Furthermore, we cannot guarantee that the solution $\hat{\boldsymbol{f}}$ will satisfy the deterministic constraint C, nor can (5.16) be modified so that the constraint will always be met. For these reasons the solution $\hat{\boldsymbol{f}}$ is computed using an iterative method, which simultaneously offers the possibility of imposing the constraint C onto the solution.

Following VanCittert's procedure, we rewrite (5.16) as

$$\begin{aligned} \hat{\boldsymbol{f}} &= \hat{\boldsymbol{f}} + \beta\{\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{g} - (\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\hat{\boldsymbol{f}}\} \\ &= \mathcal{G}(\hat{\boldsymbol{f}}). \end{aligned} \tag{5.17}$$

A solution to (5.16) can be computed by applying the method of successive approximations to (5.17):

$$\begin{aligned} \hat{\boldsymbol{f}}_{k+1} &= \mathcal{G}(\hat{\boldsymbol{f}}_k) \\ &= (\boldsymbol{I} - \alpha\beta\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\hat{\boldsymbol{f}}_k + \beta\boldsymbol{D}^t\boldsymbol{R}(\boldsymbol{g} - \boldsymbol{D}\hat{\boldsymbol{f}}_k). \end{aligned} \tag{5.18}$$

This iteration is the adaptive counterpart of the iteration (4.21). Note that (5.18) is in fact an iterative optimization algorithm based on the method of steepest descent for minimizing (5.15). In order to ensure convergence of the iterations (5.18), β must satisfy:

$$0 < \beta < \frac{2}{|\rho_{max}|}, \tag{5.19}$$

where ρ_{max} is the largest eigenvalue of the matrix $(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$. The above result was obtained by a trivial extension of the procedure which led to the convergence requirements for the iteration (4.21).

The constraint C is introduced in the algorithm by applying the projection operator P in every iteration step:

$$\hat{f}_{k+1} = P\,\mathcal{G}(\hat{f}_k). \tag{5.20}$$

As a result each iterate $\hat{f}_{k+1}$ will satisfy the *a priori* constraint C. The convergence of (5.20) cannot be shown by using an eigenvalue analysis approach, because the projection P is a possibly nonlinear operation. The more general contraction mapping theorem [100], however, provides a means to establish the convergence of nonlinear algorithms such as (5.20). It states that any iterative scheme $x_{k+1} = I(x_k)$ converges to the unique fixed point x^* of the operator I, provided that I is a contraction mapping. Since the projection operator P is nonexpansive by definition, the concatenated mapping $P\mathcal{G}$ will be contractive if $\mathcal{G}$ is a contraction mapping. In Appendix B it is shown that $\mathcal{G}$ is contractive when β satisfies (5.19). Hence, the iteration (5.20) converges if (5.18) converges.

Substituting the definition of the mapping $\mathcal{G}$ into (5.20) yields the constrained adaptive iterative image restoration algorithm:

$$\begin{aligned} \hat{f}_0 &= 0 \\ \hat{f}_{k+1} &= P\left[(\boldsymbol{I} - \alpha\beta\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\hat{f}_k + \beta\boldsymbol{D}^t\boldsymbol{R}(\boldsymbol{g} - \boldsymbol{D}\hat{f}_k)\right]. \end{aligned} \tag{5.21}$$

It is pointed out that these iterations can partially be implemented in the discrete Fourier domain, for example to evaluate the (circular) convolutions associated with $\boldsymbol{D}$ and $\boldsymbol{C}$.

In Appendix B it is shown that the limiting solution $\hat{f}_\infty$ of the iterations (5.21) minimizes the functional $\Phi(\hat{f})$ in (5.15) subject to the constraint C. Hence, if there exists a solution which satisfies both (5.15) and the constraint C, it will be obtained by the above iterative algorithm. Further, within the set of solutions satisfying both (5.15) and the constraint C, the iterations in (5.21) converge to the solution $\hat{f}_\infty$ which minimizes $\Phi(\hat{f})$. It is therefore guaranteed that the Tikhonov-Miller solution, obtained from the unconstrained minimization of (5.15), is optimally approximated within the set C. Clearly, $\hat{f}_\infty$ will coincide with the Tikhonov-Miller solution if the latter is an element of the set C (consistency of the iterative scheme). The substitution of the limiting solution $\hat{f}_\infty$ in (5.15) gives an *a posteriori* check on the claimed noise norm and the bound E.

5.3 Conjugate Gradients-based Implementation

The iterative restoration algorithm proposed in the previous section is essentially a constrained optimization algorithm based on the method of steepest descent, and can alternatively be written as:

$$\begin{aligned} \boldsymbol{r}_k &= -\frac{1}{2}\nabla_f \Phi(\boldsymbol{f})|_{\hat{\boldsymbol{f}}_k} = -(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha \boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\hat{\boldsymbol{f}}_k + \boldsymbol{D}^t\boldsymbol{R}\boldsymbol{g}, \\ \hat{\boldsymbol{f}}_{k+1} &= \mathcal{P}\,\mathcal{G}(\hat{\boldsymbol{f}}_k) = \mathcal{P}\left[\hat{\boldsymbol{f}}_k + \beta \boldsymbol{r}_k\right]. \end{aligned} \tag{5.22}$$

It minimizes the objective function (5.15) on the convex set $\mathcal{C}$. In Chapter 4 it was already shown that the method of steepest descent may converge slowly. Therefore, this section will consider more efficient ways to minimize (5.15) subject to the constraint $\mathcal{C}$.

A first improvement on the convergence speed can be achieved by optimizing the value of the convergence parameter β at each iteration. This optimal value, denoted by β_k, minimizes $\Phi(\mathcal{P}[\hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{r}_k])$. If $\mathcal{P}$ is not present in the iterations, the following analytic result can be derived:

$$\beta_k = \frac{\|\boldsymbol{r}_k\|^2}{\|\boldsymbol{D}\boldsymbol{r}_k\|_R^2 + \alpha\|\boldsymbol{C}\boldsymbol{r}_k\|_S^2}. \tag{5.23}$$

In the more general case, which includes the (possibly nonlinear) projection $\mathcal{P}$, we cannot obtain an explicit expression for the optimal value of β_k. Therefore, line search methods need to be employed such as the repeated quadratic interpolation or the golden section rule [63]. These methods essentially evaluate the objective function $\Phi(\mathcal{P}[\hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{r}_k])$ for a number of suitably chosen values of β_k, and select the one which yields the minimal value for the objective function. Although optimizing β_k at each iteration slightly increases the convergence speed, as will be illustrated by an example later on, the additional computation required to repeatedly evaluate the objective function in (5.15) makes this approach rather unattractive.

In Section 4.4 it was shown that employing a conjugate gradients based optimization or a higher order convergence method can lead to very efficient iterative restoration algorithms. Unfortunately, the methods that have a converge order larger than one cannot be applied in conjunction with

the nonlinear projection operator, nor can they handle spatially variant restoration because of memory requirements [23,51,68,69]. We will therefore consider the application of the conjugate gradients algorithm (4.38) in adaptive and nonlinear restoration.

Although the use of a nonlinear projection $\mathcal{P}$ is not consistent with conjugate direction methods, they have nonetheless been used in conjunction with them. The most simple and computationally efficient approach is to project the iterates themselves after each iteration step [51,65]. More complicated methods to incorporate (nonlinear) constraints, such as the gradient projection method [63], will not be considered here.

The conjugate gradients method for constrained adaptive restoration can thus be formulated as:

$$\begin{aligned} \boldsymbol{r}_k &= -\frac{1}{2}\nabla_f \Phi(f)\big|_{\hat{f}_k} = -(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha \boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\hat{\boldsymbol{f}}_k + \boldsymbol{D}^t\boldsymbol{R}\boldsymbol{g}, \\ \boldsymbol{p}_k &= \boldsymbol{r}_k + \gamma_k \boldsymbol{p}_{k-1}, \\ \hat{\boldsymbol{f}}_{k+1} &= \mathcal{P}\left[\hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{p}_k\right]. \end{aligned} \tag{5.24}$$

Eq. (5.24) represents a true conjugate gradients algorithm, only if the projection $\mathcal{P}$ is omitted. In that case the iterations (5.24) minimize the objective function $\Phi(\hat{f})$ in (5.15), and the optimal values for γ_k and β_k are given by [28]:

$$\gamma_k = \frac{\|\boldsymbol{r}_k\|^2}{\|\boldsymbol{r}_{k-1}\|^2}, \tag{5.25}$$

$$\beta_k = \frac{\boldsymbol{r}_k^t \boldsymbol{p}_k}{\|\boldsymbol{D}\boldsymbol{p}_k\|_R^2 + \alpha\|\boldsymbol{C}\boldsymbol{p}_k\|_S^2}. \tag{5.26}$$

The above expressions and the iterations (5.24) can again be implemented partially in the discrete Fourier domain.

Clearly, by incorporating the projection $\mathcal{P}$ into the algorithm, the concept of an orthogonal solution decomposition, on which the conjugate gradients method is based, can no longer hold. However, for many restoration problems of interest the vector $\boldsymbol{p}_{k-1}$ is still useful in determining the current direction of modifications (i.e. $\boldsymbol{p}_k$). This is particularly true when the effect of the projection operator is relatively small at each iteration, though its effect on the entire optimization can be very significant [51]. The choice for the values of β_k and γ_k becomes more difficult as well if a projection operator is used. Although analytical expressions can no longer be obtained, usually a suitable choice for γ_k is still given by (5.25). In

order to ensure convergence the optimal value for β_k must be found by a line search procedure, i.e. an algorithm that minimizes the objective function $\Phi(\mathcal{P}[\hat{\boldsymbol{f}}_k + \beta_k \boldsymbol{p}_k])$ as a function of β_k. Since the method of conjugate gradients has a considerably higher convergence speed than the method of steepest descent, evaluating the objective function $\Phi(\hat{\boldsymbol{f}})$ a number of times for each iteration step is now worth the additional effort.

Since β_k is optimized at each iteration step, $\Phi(\hat{\boldsymbol{f}}_{k+1})$ decreases monotonically, i.e. the iterations are ensured to converge. As the iterations (5.24) progress, they usually behave more and more like the steepest descent iterations of (5.21). An explanation for this effect is that as $k \to \infty$, the iterates $\hat{\boldsymbol{f}}_k$ and $\hat{\boldsymbol{f}}_{k+1}$ lie very close to each other within the convex set C. Hence, the gradients $\boldsymbol{r}_k$ and $\boldsymbol{r}_{k+1}$ are approximately the same vectors. As a result $\boldsymbol{p}_k \to \boldsymbol{r}_k$, i.e. (5.24) becomes a steepest descent iteration as $k \to \infty$. An immediate consequence of this is that the limiting solution of the iterations (5.24) is identical to the one obtained by the steepest descent based iteration (5.21).

5.4 Experimental Restoration Results

The performance of the constrained iterative image restoration method in (5.21) is illustrated by three examples on artificially blurred images in this section. The results will be compared with an important representative of the nonadaptive restoration methods, namely the Tikhonov-Miller regularized restoration filter of (3.31). The 2-D Laplacian filter was used as the regularizing operator $\boldsymbol{C}$.

In the first experiment the effects of the weighting matrices $\boldsymbol{R}$ and $\boldsymbol{S}$ in combination with an intensity constraint will be studied. Further, the converge behavior of the iterations (5.21), (5.22) and (5.24) will be compared. Where the first experiment deals with the restoration of a natural and therefore "complex" scene, the second experiment addresses the restoration of a simple text image, in which the application of an intensity constraint is very effective. Again, the difference in convergence behavior between the various algorithms will be illustrated. The last experiment deals with the worst-case situation with respect to ringing, namely the restoration of a motion blurred image. In this case in particular the use of the weighting matrix $\boldsymbol{S}$ will be shown to be very effective. It will also be illustrated that the matrix $\boldsymbol{R}$ can be used to handle the presence of outliers in the blurred image data.

In the experiments the weight coefficients r_{ij} and s_{ij} were tuned as follows. A local variance measure $\sigma_g^2(i,j)$ was first computed from the blurred image $\boldsymbol{g}$ [46]:

$$\sigma_g^2(i,j) = \frac{1}{(2P+1)(2Q+1)} \sum_{k=i-P}^{i+P} \sum_{l=j-Q}^{j+Q} [g(k,l) - m_g(i,j)]^2, \qquad (5.27)$$

where $(2P+1) \times (2Q+1)$ is the size of the analysis window in which the local variance is computed, and where $m_g(i,j)$ is the local mean, given by

$$m_g(i,j) = \frac{1}{(2P+1)(2Q+1)} \sum_{k=i-P}^{i+P} \sum_{l=j-Q}^{j+Q} g(k,l). \qquad (5.28)$$

In regions with slow intensity transitions the local variance takes approximately the value of the noise variance, whereas the larger local variances correspond to "edgy" image regions. Using this local variance measure, the coefficients in the weighting matrix S were computed as

$$s_{ij} = \frac{1}{1 + \mu \max[0, \sigma_g^2(i,j) - \sigma_w^2]}, \qquad (5.29)$$

where the noise variance σ_w^2 was estimated from a smooth image region, and where μ is a tuning parameter. In a similar way the coefficients r_{ij} were tuned according to

$$r_{ij} = \frac{1}{1 + (\mu \max[0, \sigma_g^2(i,j) - \sigma_w^2])^{-1}}, \qquad (5.30)$$

Experiment 1

The first example considers the constrained adaptive restoration of the defocused cameraman image in Figure 5.4a (R=3, SNR=40 dB). To control the regularization locally, the use of the weighting matrix $\boldsymbol{R}$ was considered first ($P = Q = 2, \mu = 10^{-3}, \alpha = 10^{-3}$, and $S = I$). In addition to the adaptive approach the intensities in the restored image were constrained to the intensity interval $[240, 10]$. Associated with this convex set is the following projection operator:

$$\mathcal{P}[\hat{f}(i,j)] = \min\{240, \max[\hat{f}(i,j), 10]\}. \qquad (5.31)$$

The restoration result obtained by (5.24) is shown in Figure 5.4c, and has an SNR improvement of $\eta_{\text{SNR}} = 6.9$ dB. This result is to be compared with

(a)

(b)

(c)

(d)

Figure 5.4: Constrained adaptive restoration: (a) Defocused cameraman image ($R = 3$) with noise added (SNR=40 dB); (b) Nonadaptive Tikhonov-Miller restoration result; (c,d) Constrained adaptive iterative restoration result.

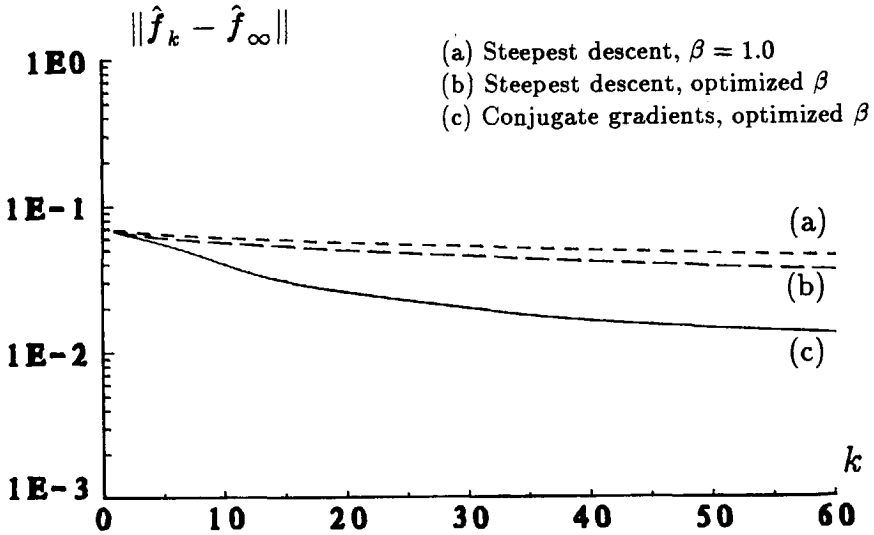


Figure 5.5: Convergence behavior for three different implementations in restoring the defocused cameraman image.

the nonadaptive Tikhonov-Miller regularized restoration result in Figure 5.4b, which has an SNR improvement of 6.1 dB.

Next, the effect of the weighting matrix S was considered ($P = Q = 2, \mu = 0.5, \alpha = 0.3$, and $R = I$). Again the projection operator in (5.31) was used. The resulting restoration is shown in Figure 5.4d, and has an SNR improvement of 8.2 dB. Obviously the adaptivity introduced through the weighting matrix S works best in this case. Further, the combination of S and R in the restoration process does not lead to significant additional improvements compared with using S alone. Since similar effects are observed in other restoration experiments, we conclude that it is often sufficient to use only the weighting matrix S in adaptively controlling the local regularization and noise magnification error. The experiment 3 will, however, show a restoration example in which both the weighting matrices R and S can advantageously be used.

In Figure 5.5 the convergence behavior of the steepest descent algorithm with fixed β, the steepest descent algorithm with optimized β, and the conjugate gradients based implementation is shown. It is observed that, even though a nonlinear projection operator is used within the con-

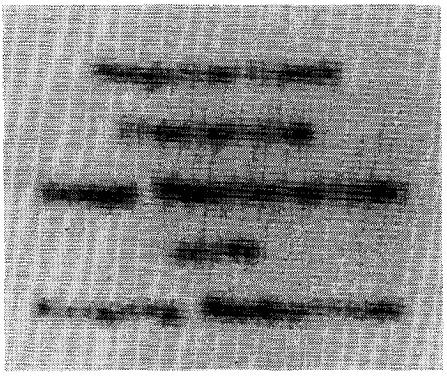
Table 5.1: SNR improvement (dB) for the restoration of the defocused text image.

SNR (dB)	Unconstrained restoration	Constrained restoration
20	2.7	4.8
30	5.5	10.8
40	9.8	21.8
50	14.5	35.2
60	19.5	44.2

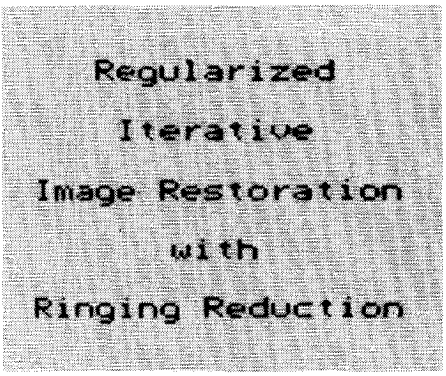
jugate gradients algorithm, its convergence speed outperforms the steepest descent algorithms significantly.

Experiment 2

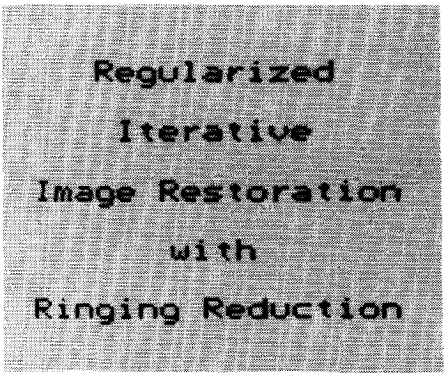
The next experiment considers the restoration of a "text image" (size is 256×256 pixels), that was defocused with $R = 7$. White Gaussian noise was added at $20, 30, 40, 50$ and 60 dB, respectively. In Figure 5.6a the blurred image with SNR=30 dB is shown. The restoration obtained by Tikhonov-Miller's method of regularization is given in Figure 5.6b ($\alpha = 3.10^{-3}$). Although the numerical improvement of 5.5 dB is acceptable and the text becomes readible, the restored image shows annoying ringing effects. To prevent ringing from happening, the intensities were constrained to the interval $[210, 25]$. Since this constraint represents very accurate knowledge about the original image intensities, the restoration result obtained by the constrained iterative algorithm is very good. It has an SNR improvement of 10.8 dB, and is shown in Figure 5.6c. In Table 5.1 the SNR improvements for various SNRs are listed for the Tikhonov-Miller regularized filter and the constrained iterative restoration algorithm. Figure 5.7 gives a comparison between the convergence behavior of the steepest descent implementation with fixed β and with optimized β, and the conjugate gradients implementation. Even though the deterministic constraint plays an important role in the restoration process, and hence the concept of an orthogonal solution decomposition breaks down theoretically, the conjugate gradients implementation (5.24) turns out to be still very useful.



(a)



(b)



(c)

Figure 5.6: Restoration of a text image: (a) Defocused image; (b) Unconstrained restoration result; (c) Constrained restoration result.

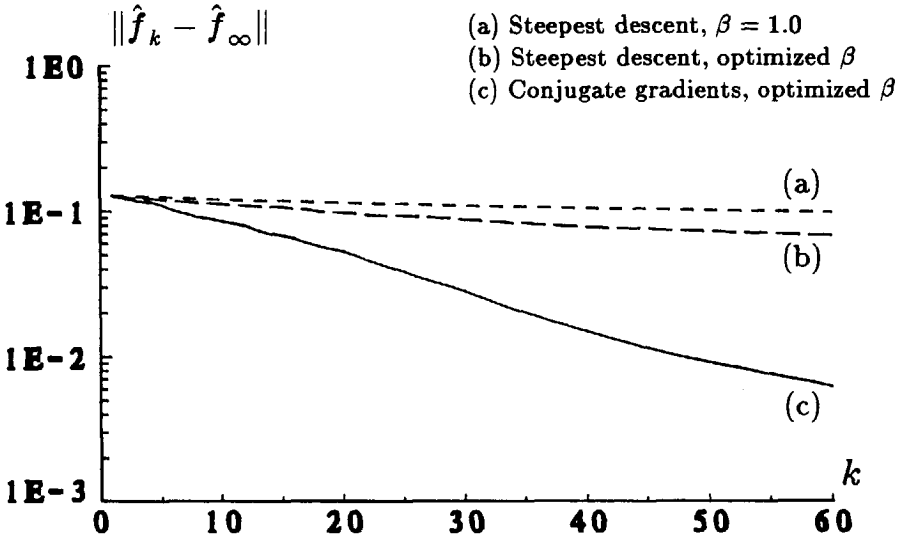


Figure 5.7: Convergence behavior for three different implementations in restoring the defocused text image.

Experiment 3

The last experiment deals with the worst-case situation (with respect to ringing effects) of linear motion blur. The cameraman image was blurred by horizontal motion over 9 pixels and noise was added with SNR = 30 dB (Figure 5.8a). The unconstrained nonadaptive restoration result is shown in Figure 5.8b ($\alpha = 0.05$), and has an SNR improvement of 5.5 dB. The restoration result obtained by the constrained adaptive algorithm using the weighting matrix S and by using a projection operator to constrain the intensities to the range $[240, 10]$, is shown in Figure 5.8c ($\alpha = 3.0, \mu = 0.5, P = Q = 2$). The SNR improvement is 7.4 dB. Particularly in this situation the restoration result has considerably fewer ringing artifacts. As was discussed in this chapter, this can only be achieved by reducing the regularization near edges in the image, yielding more noise magnification. This effect is illustrated in Figure 5.8d, which shows a magnified portion of the restoration result in Figure 5.8c. Table 5.2 lists the performance of the restoration algorithm for various SNRs.

(a)

(b)

(c)

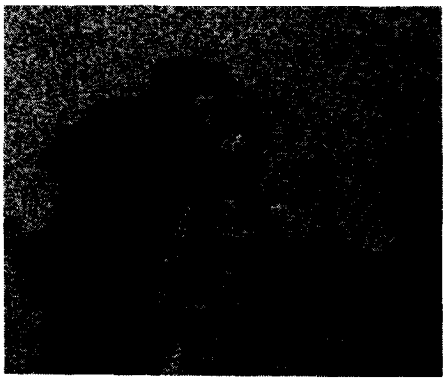
(d)

Figure 5.8: Restoration of the cameraman image: (a) Motion blurred image; (b) Unconstrained nonadaptive restoration result; (c) Constrained adaptive restoration result; (d) Enlargement of (c).

Table 5.2: SNR improvement (dB) for the restoration of the motion blurred cameraman image.

SNR (dB)	Unconstrained nonadaptive restoration	Unconstrained adaptive restoration	Constrained adaptive restoration
20	2.8	3.9	4.3
30	5.5	7.1	7.4
40	9.7	11.6	11.8
50	14.6	16.8	16.9
60	19.8	21.3	21.3

In experiment 1 it was concluded that the regularization error can sufficiently be controlled by the matrix $\boldsymbol{S}$. If, however, the observed image is corrupted severely due to for instance shot noise, the matrix $\boldsymbol{R}$ can be used to exclude the outliers from the restoration process. In order to illustrate the effects of such an algorithm 50 percent of the pixels in the noisy blurred image in Figure 5.8b were randomly discarded. The resulting image is shown in Figure 5.9a, where all the discarded pixels are given an intensity of 0 (black). In the restoration process the corresponding coefficients r_{ij} were set to zero to exclude the erroneous data. The restoration result from this noisy blurred image with incomplete data is presented in Figure 5.9b, showing the robustness of the iterations (5.24).



(a)

(b)

Figure 5.9: (a) 50 percent of the data in Figure 5.8a discarded; (b) Restoration result.

Chapter 6

Maximum Likelihood Image Identification

Sofar we have assumed that the PSF of the image formation system, the variance of the observation noise, and a model for the original image were known prior to the restoration process. In many practical situations of interest these parameters are, however, not available. Image identification (sometimes referred to as blur identification, image-blur identification, or *a posteriori* restoration) focuses on developing estimation procedures, which identify all the information that is required to restore an image from the noisy blurred image itself.

Again we will assume that the structure of the models (2.8) and (2.13) is an appropriate abstraction of reality, but now the coefficients $d(m,n)$ and $a(k,l)$, and the variances of the noise processes $v(i,j)$ and $w(i,j)$ are unknown. Then the image identification problem can be specified as the estimation of the parameter vector θ, defined as

$$\theta = (\theta_1, \theta_2, \cdots, \theta_M)^t = (d(m,n), a(k,l), \sigma_w^2, \sigma_v^2)^t, \tag{6.1}$$

from the observed noisy blurred image $g(i,j)$. The major focus will be on developing procedures which provide the maximum likelihood (ML) estimator of the unknown parameters.

In Section 6.1 we will start with a brief overview of conventional image identification methods, including spectral and cepstral techniques for estimating the PSF. In Section 6.2 the image identification problem will be formulated as a maximum likelihood problem. The likelihood function will be established, and some of its properties will be investigated. From this formulation it follows that ML image estimation requires the optimization

of a complicated nonlinear function. Since analytic solutions cannot in general be obtained, the remainder of this chapter (as well as Chapters 7 and 8) will be concerned with recently proposed procedures for efficient optimization of the likelihood function.

This chapter aims at reviewing existing methods, which seem to be unrelated, under a unifying ML framework [59]. In this way it is shown that all identification methods known sofar in the literature are merely different implementations of the same estimator, resulting from different modeling assumptions and/or considerations about the computational complexity or computer resources available.

In Section 6.3 we will first consider the class of image identification methods which is based on the assumption of noiseless observed data. It will be shown that in this case the ML image identification problem can be reformulated as a least-squares problem. Next, in Section 6.4 two identification methods will be discussed which do not assume noisefree image data.

A common feature of the ML methods discussed in this chapter is that they nearly all employ gradient-based optimization procedures, either in an implicit or explicit manner. In Chapter 7 a more elegant iterative algorithm will be introduced which does not make use of gradients to optimize the likelihood function.

6.1 Conventional Identification Methods

Probably the most straightforward way to obtain the point-spread function of an image formation system is to make use of an analytic description of the system and to substitute all the parameters required, such as camera misadjustment, object distances, object motion, and camera motion. If such an analytic model is not available, or if its parameters are not exactly known, one could analyze the image formation process by recording an original image with known pattern $k(i,j)$ (e.g. an impulse or an edge) a number of times. If we denote the recorded images by $g_r(i,j)$, with $r = 1, 2, \cdots, R$, and we take R large enough, and if we assume that each image $g_r(i,j)$ has been blurred by the same PSF, then the PSF can be solved from the following relation:

$$\frac{1}{R}\sum_{r=1}^{R} g_r(i,j) \quad = \quad \frac{1}{R}\sum_{r=1}^{R} [d(i,j) * k(i,j) + w_r(i,j)]$$

$$= d(i,j) * k(i,j) + \sum_{r=1}^{R} w_r(i,j)$$
$$\approx d(i,j) * k(i,j). \tag{6.2}$$

In fact this method can be regarded as solving the classical system identification problem, where the response of a (linear) system is identified given its input and output. Unfortunately, such an approach is rarely applicable in image identification.

A related but more practically oriented approach assumes that certain patterns in the observed blurred image are *a priori* known. For example, if a blurred edge is observed, an idealized (sharp) edge in the original image could be assumed to be associated with this edge. Again, in practice this approach does not work very well, because small deviations between the assumed idealized pattern and the actual (unknown) pattern in the original image give rise to large deviations in the PSF identified in this way. Furthermore, this method does not account for the noise in the observed image.

A more successful approach towards image identification is due to Stockham *et al.* [93] and Cannon [18]. Their spectral and cepstral methods concentrate on PSFs whose Fourier transform has a regular pattern of zero-crossings (such as uniform linear motion blur). Since these zeros can usually also be located in the Fourier transform of the blurred image, they can be used to identify this class of PSFs.

If we neglect the noise contribution in the observed blurred image, then the power density spectra of the original and blurred image are related by

$$|G(u,v)|^2 = |D(u,v)|^2 \, |F(u,v)|^2. \tag{6.3}$$

If the images $\boldsymbol{g}$ and $\boldsymbol{f}$ are divided in non-overlapping subimages, denoted by $g_s(i,j)$ and $f_s(i,j)$, the power density spectra of these subimages will approximately satisfy a relation similar to (6.3) (neglecting boundary effects):

$$|G_s(u,v)|^2 \approx |D(u,v)|^2 \, |F_s(u,v)|^2. \tag{6.4}$$

If the subimages are large compared to the extent of the PSF, then this is a good approximation. Taking the logarithm of (6.4), and adding the results for each of the S subimages gives:

$$\frac{1}{S} \sum_{s=1}^{S} \log |G_s(u,v)|^2 \approx \frac{1}{S} \sum_{s=1}^{S} \log |F_s(u,v)|^2 + \log |D(u,v)|^2. \tag{6.5}$$

The first sum on the right can be approximated by an average power spectrum over an ensemble of images, and next be subtracted from the expression on the left hand side. Resulting from this will be an approximation to the magnitude of the response of the blurring function. However, this procedure is very sensitive to deviations in the average power spectrum, it does not account for noise, and the PSF obtained is only a rough approximation of the true PSF.

If $D(u,v)$ has zero-crossings, then (6.5) will show large negative peaks at these locations. In this way, a zero-crossing pattern can easily be extracted from $|G(u,v)|$, from which it is usually possible to obtain a fair indication of the type, extent, and orientation of the blur (for example, the length and angle of linear motion blur, or the radius of defocusing blur).

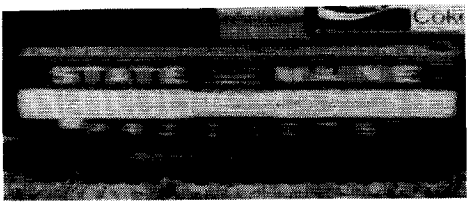
An alternative to the above for linear motion blur involves the computation of the cepstrum. The (power) cepstrum, denoted by $\hat{g}(i,j)$, is the inverse discrete Fourier transform of (6.5), and has the important property that if two signals are convolved, their cepstra add. Thus, if the noise is again neglected, we have

$$\hat{g}(i,j) = \mathcal{F}^{-1}\{\frac{1}{S}\sum_{s=1}^{S}\log|G_s(u,v)|^2\} = \frac{1}{S}\sum_{s=1}^{S}\hat{f}_s(i,j) + \hat{d}(i,j). \qquad (6.6)$$

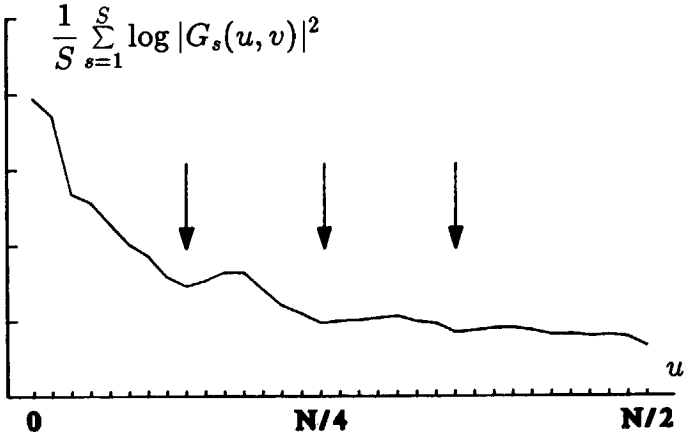
For horizontal motion blur over L pixels, $|D(u,v)|$ has zeros in horizontal direction at multiples of $\frac{N}{L}$. As a result $\hat{d}(i,j)$ has a large negative spike at the horizontal axis at a distance L. As an example consider the photographically blurred train image in Figure 6.1a, showing horizontal motion blur. The row-wise summed log-spectrum, formed of 32 rows in the center of the image, is shown in Figure 6.1b, and the associated cepstrum in Figure 6.1c. The cepstrum displays prominent spikes at $i = 8$ and $i = 9$, indicating that the length of motion is in the range of 8 to 9 pixels.

Shortcomings of the above approaches are that PSFs which do not have spectral zeros (such as Gaussian blurs), cannot be identified in this way, and that the presence of noise in the observed image is not taken into account. For this reason the methods described in this section should be considered as a means for obtaining a first impression of the type and severity of the blur present in the image. After such an initial investigation of the blurred image more accurate methods need to be employed, such as the maximum likelihood approach described next.

(a)



(b)



(c)

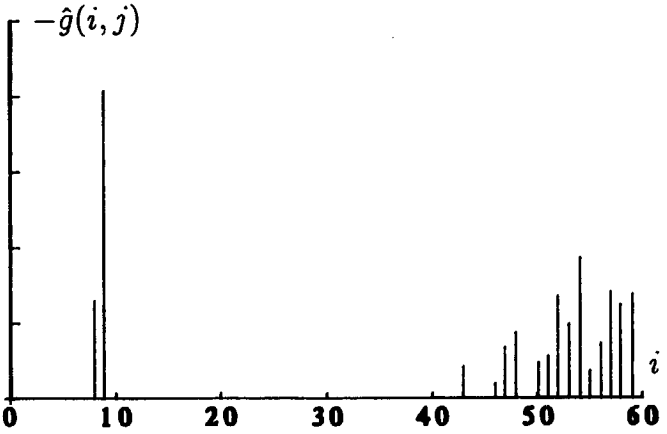


Figure 6.1: Identification of motion blur: (a) Blurred image; (b) 1-D log-spectrum; (c) Cepstrum from (b).

6.2 Maximum Likelihood Estimator

6.2.1 Introduction

If we assume that the original image can be modeled by the 2-D AR model in (2.8), and that the image formation system can be modeled by (2.13), the image identification problem can be considered a parameter estimation problem, where the unknown parameters are given by θ defined in (6.1). In this thesis we will focus on computing maximum likelihood (ML) estimates of θ, because ML procedures have shown to be very powerful in many applications [6,22,26,62,85].

Maximum likelihood estimation procedures are based on optimizing the probability density function (PDF) of the observed image $\boldsymbol{g}$ with respect to the unknown parameters. In this way the ML estimator finds the various values in θ which most likely resulted in the blurred image observed. Since the PDF of $\boldsymbol{g}$ is required in the ML estimation of θ, we need to assign PDFs to the stochastic processes involved in the image and observation model. In the light of the practical experience of researchers in image identification and restoration we assume that the driving process (or modeling error) $\boldsymbol{v}$ in the image model

$$\boldsymbol{f} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{v} \tag{6.7}$$

is a homogeneous Gaussian distributed white noise process with zero mean and covariance matrix $\boldsymbol{Q}_v = \sigma_v^2 \boldsymbol{I}$ ($\sigma_v^2 > 0$), and is uncorrelated with $\boldsymbol{f}$. The PDF of $\boldsymbol{f}$, given the PDF of $\boldsymbol{v}$ and the model (6.7), is then given by:

$$p(\boldsymbol{f}; \boldsymbol{A}, \boldsymbol{Q}_v) = \sqrt{\frac{\det |\boldsymbol{I} - \boldsymbol{A}|^2}{2\pi^{MN} \det |\boldsymbol{Q}_v|}} \exp \left\{ -\frac{1}{2} \boldsymbol{f}^t (\boldsymbol{I} - \boldsymbol{A})^t \boldsymbol{Q}_v^{-1} (\boldsymbol{I} - \boldsymbol{A}) \boldsymbol{f} \right\}. \tag{6.8}$$

It is recalled that $(\boldsymbol{I} - \boldsymbol{A})$ is nonsingular, because the image model $a(k,l)$ is assumed to be stable.

The observation noise $\boldsymbol{w}$ in the image formation model

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{f} + \boldsymbol{w} \tag{6.9}$$

is also assumed to be a homogeneous Gaussian distributed white noise process with zero mean and covariance matrix $\boldsymbol{Q}_w = \sigma_w^2 \boldsymbol{I}$ ($\sigma_w^2 > 0$), and is uncorrelated with $\boldsymbol{v}$. The PDF of $\boldsymbol{g}$, given the PDF of the observation

noise, the model (6.9), and the original image $\boldsymbol{f}$, can be expressed as:

$$p(\boldsymbol{g}/\boldsymbol{f};\boldsymbol{D},\boldsymbol{Q}_w) = \frac{1}{\sqrt{2\pi^{MN}\det|\boldsymbol{Q}_w|}} \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})^t\boldsymbol{Q}_w^{-1}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})\right\}. \tag{6.10}$$

Here the PDF of the blurred image data is conditioned on both the original image (denoted by $\boldsymbol{g}/\boldsymbol{f}$) and on the deterministic parameters $\boldsymbol{D}$ and $\boldsymbol{Q}_w$ (denoted by $\boldsymbol{g};\boldsymbol{D},\boldsymbol{Q}_w$).

6.2.2 Definition of the Likelihood Function

The maximum likelihood estimator of the parameter vector θ is defined by [62,85]

$$\hat{\theta}_{\mathrm{ml}} = \arg\left\{\max_{\theta\in\Theta}\mathcal{L}^*(\theta)\right\} = \arg\left\{\max_{\theta\in\Theta}\log p(\boldsymbol{g};\theta)\right\}. \tag{6.11}$$

Here $\mathcal{L}^*(\theta)$ denotes the (log-)likelihood function of θ, $p(\boldsymbol{g};\theta)$ denotes the *a priori* probability density function of the observed image given θ, and Θ specifies the range of the parameters θ. In order to compute $p(\boldsymbol{g};\theta)$ we combine (6.7) and (6.9) as follows:

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{f} + \boldsymbol{w} = \boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{v} + \boldsymbol{w}. \tag{6.12}$$

Since $\boldsymbol{v}$ and $\boldsymbol{w}$ are uncorrelated Gaussian processes, $p(\boldsymbol{g};\theta)$ is Gaussian as well, with zero mean and covariance matrix $\boldsymbol{P}$ given by

$$\begin{aligned}\boldsymbol{P} = \mathrm{Cov}\,(\boldsymbol{g};\theta) &= \mathrm{E}\left\{(\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{v}+\boldsymbol{w})(\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{v}+\boldsymbol{w})^t\right\} \\ &= \sigma_v^2\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})^{-t}\boldsymbol{D}^t + \sigma_w^2\boldsymbol{I}. \end{aligned} \tag{6.13}$$

By substituting this result for $p(\boldsymbol{g};\theta)$ into (6.11), dropping all terms which are independent of θ, and premultiplying the result by -2, the ML image identification problem can be expressed as follows:

$$\hat{\theta}_{\mathrm{ml}} = \arg\left\{\min_{\theta\in\Theta}\mathcal{L}(\theta)\right\} = \arg\min_{\theta\in\Theta}\left\{\log(\det|\boldsymbol{P}|) + \boldsymbol{g}^t\boldsymbol{P}^{-1}\boldsymbol{g}\right\}. \tag{6.14}$$

If the convolutions appearing in the image model and observation equation are interpreted as circular convolutions, the matrices $\boldsymbol{D}$ and $\boldsymbol{A}$ have a block-circulant structure, and $\boldsymbol{P}$ has a block-circulant structure as well. Associated with $\boldsymbol{P}$ is the following 2-D convolution kernel $p(i,j)$:

$$p(i,j) = \sigma_v^2 d(i,j) * (1-a(i,j))^{-1} * (1-a(-i,-j))^{-1} * d(-i,-j) + \sigma_w^2. \tag{6.15}$$

We observe that all matrices in (6.13) commute because they are related to convolutions. Further, the covariance matrix $\boldsymbol{P}$ is positive definite provided that $\sigma_w^2 > 0$. Since there is always noise present in a blurred image, this condition is satisfied. As a consequence the inverse of $\boldsymbol{P}$ and the logarithm of $\det|\boldsymbol{P}|$ always exist.

We should note that the dimension of $\boldsymbol{P}$ is $MN \times MN$, where typical values of M, N are 128, 256 or even larger. Hence, the direct evaluation of (6.13) and (6.14) is hardly ever feasible in practice. If, however, we assume that the matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ are block-circulant, the likelihood function $\mathcal{L}(\theta)$ can be evaluated very efficiently in the frequency domain:

$$\begin{aligned}\mathcal{L}(\theta) &= \log(\det|\boldsymbol{P}|) + \boldsymbol{g}^t\boldsymbol{P}^{-1}\boldsymbol{g} \\ &= \log\left\{\prod_{u,v} P(u,v)\right\} + \sum_{u,v}\frac{|G(u,v)|^2}{P(u,v)} \\ &= \sum_{u,v}\left\{\log P(u,v) + \frac{|G(u,v)|^2}{P(u,v)}\right\}, \end{aligned} \tag{6.16}$$

where

$$P(u,v) = \sigma_v^2\frac{|D(u,v)|^2}{|1-A(u,v)|^2} + \sigma_w^2, \tag{6.17}$$

and $D(u,v)$, $A(u,v)$, $P(u,v)$, $G(u,v)$ are the DFT of $d(m,n)$, $a(k,l)$, $p(i,j)$ and $g(i,j)$, respectively. Alternatively, the likelihood function may be evaluated recursively in the spatial domain by using the concept of prediction error. This approach will be discussed in Section 6.4.2.

6.2.3 Properties of the Estimator

By making use of a maximum likelihood estimation procedure, the estimated value $\hat{\theta}$ of θ is guaranteed to be consistent and asymptotically efficient [85]. This means that in the limit for an infinite number of data points, $\hat{\theta}$ is unbiased and the variance of the estimation error is equal to the Cramer-Rao lower bound. However, since images consist only of a finite number of pixels, the maximum likelihood estimator for θ may not be unbiased or efficient.

It might therefore be interesting to evaluate the theoretical and practical (i.e. measured) bias and estimation error for the identification problem under consideration. Unfortunately, analytic expressions for the bias of this estimator cannot be obtained. As a consequence the Cramer-Rao lower bound (which can be analytically obtained) cannot be checked against the

measured variances of the estimation error, because the (unknown) bias is always included in such a measurement. In other words, it is not possible to discriminate between the bias and the variance of the estimation error. In addition to this, computed results may also be subject to numerical inaccuracies, which may prevent finding the exact optimum of the likelihood function. In the next chapter we will illustrate the various types of errors by a numerical example. With hindsight we remark that for low signal-to-noise ratios, the variations in the estimated parameters cause the largest difference between the actual and estimated parameters, while for moderate to high SNRs biases tend to dominate the estimated parameters, though these effects depend on the type and severity of the blur considered.

If we have a closer look at (6.16) and (6.17), we observe that $\mathcal{L}(\theta)$ is independent of the phase information of $1 - A(u,v)$ and $D(u,v)$. This fact does not constitute a problem for the uniqueness of the estimates of $a(k,l)$, since we have assumed that the image model is stable and its support is causal. Then the zeros of the image model always lie inside the unit bicircle. As a result $1 - A(u,v)$ always has minimum phase. This eliminates the phase ambiguity for $1 - A(u,v)$ and results in unique estimates for $a(k,l)$.

On the other hand, the estimates of $d(m,n)$ are not unique in the absence of any additional constraints, mainly because blurs may have any kind of phase. Hence, the ML procedure can only correctly identify the magnitude-squared of $D(u,v)$, i.e. $|D(u,v)|^2$. Non-uniqueness of the estimates of $d(m,n)$ can in general be avoided by enforcing the solution to satisfy certain constraints. In image identification it is common to assume that the PSF is symmetric, i.e. $d(m,n) = d(-m,-n)$. In this way the phase of the PSF is enforced to be either zero or $\pm\pi$. Unfortunately, even when we exploit the symmetry of the PSF, the phase of the solution is not defined uniquely because of the presence of $\pm\pi$, which translates into an ambiguity of the sign of $D(u,v)$. In practice, however, the sign-ambiguity is not a serious problem, because two additional constraints can be imposed on the solution to alleviate this problem. These are: (i) the PSF coefficients are nonnegative, and (ii) the support of the PSF, $\mathcal{S}_d$, is finite.

The above constraints on the PSFs are nearly always restrictive enough to obtain a unique estimate for the PSF. In developing the algorithms for the optimization of $\mathcal{L}(\theta)$ we will assume that the actual PSF is symmetric, positive, and has finite support, thereby restricting the class of PSFs that can be correctly identified. It should be noted that other forms of sym-

metries, such as $d(m,n) = d(-m,n)$ or $d(m,n) = d(m,-n)$ could also be used as constraints.

6.2.4 Analytic Solutions

In a first attempt to optimize $\mathcal{L}(\theta)$ we may consider an explicit analytic solution. Although the partial derivates of the likelihood function with respect to the elements of θ can be computed (see Section 6.4.1), equating these to zero does not lead to any useful analytic expression for the parameters to be estimated, mainly because of the non-quadratic behavior of $\log(\det|\boldsymbol{P}|)$.

There is, however, one exception: let us introduce the regularization parameter α according to Miller:

$$\alpha = \frac{\sigma_w^2}{\sigma_v^2}. \tag{6.18}$$

Substituting α into the likelihood function (6.14) results in the following (equivalent) ML problem:

$$\hat{\theta'}_{\text{ml}} = \arg\min_{\theta' \in \Theta'} \left\{ MN \log(\sigma_v^2) + \log(\det|\tilde{\boldsymbol{P}}|) + \frac{1}{\sigma_v^2} \boldsymbol{g}^t \tilde{\boldsymbol{P}}^{-1} \boldsymbol{g} \right\}, \tag{6.19}$$

where

$$\tilde{\boldsymbol{P}} = \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{A})^{-1}(\boldsymbol{I} - \boldsymbol{A})^{-t}\boldsymbol{D}^t + \alpha \boldsymbol{I}, \tag{6.20}$$

and $\theta' = (d(m,n), a(k,l), \alpha, \sigma_v^2)^t$. This formulation allows us to derive an analytic expression for either σ_v^2 or σ_w^2. Confining ourselves to the analytic expression for σ_v^2 (the other case is similar), we find:

$$\hat{\sigma}_v^2 = \frac{1}{MN} \boldsymbol{g}^t \tilde{\boldsymbol{P}}^{-1} \boldsymbol{g}. \tag{6.21}$$

The interpretation of (6.19) and (6.20) is that for given $\boldsymbol{D}, \boldsymbol{A}$ and α, the optimal value of σ_v^2 can be determined analytically via (6.20) and (6.21). Contrary to this, the minimization in (6.19) requires the optimal value of σ_v^2. Therefore, the above procedure does not make the actual optimization of the likelihood function much easier. More importantly, however, (6.18) reflects the well-known fact that in image restoration only the ratio of σ_w^2 to σ_v^2 is relevant, such as can be seen from the class of Tikhonov-Miller regularized restoration filters. In this sense the elimination of the actual

noise variances from the identification procedure is very natural.

In Section 6.3.1 it will be shown that for the hypothetical case $\alpha = 0$ (i.e. the observed blurred image is noiseless) the likelihood function can be optimized independently of σ_v^2. Image identification algorithms based on this assumption will be discussed in Section 6.3. For values of α which are much smaller than the diagonal elements of $\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})^{-t}\boldsymbol{D}^t$, we may very well assume that the effect of α in (6.20) is negligible. This means that we are allowed to apply identification methods which are based on noiseless data to noisy data [13,52,96,97], and account for the presence of noise later on. On the other hand, for very noisy data, $\alpha\boldsymbol{I}$ dominates $\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{I}-\boldsymbol{A})^{-t}\boldsymbol{D}^t$. In this case $\tilde{\boldsymbol{P}}$ becomes independent of $\boldsymbol{D}$ and $\boldsymbol{A}$, the entries of which can therefore no longer be identified correctly. The solution to (6.19) is then given by

$$\hat{\sigma}_w^2 = \frac{1}{MN}\boldsymbol{g}^t\boldsymbol{g}, \tag{6.22}$$

i.e. the observed image is considered to be composed of observation noise only.

Since an analytic solution for $\hat{\theta}$ cannot be found in general, numerical solution strategies have to be considered in order to optimize $\mathcal{L}(\theta)$. An exhaustive search method or any other more intelligent search method could theoretically be employed to find a solution to (6.19). The dimension of the solution space is, however, very large due to the total number of unknowns. For example, a moderate blur with an extent of 5×5 pixels and a first-order NSHP image model support would already lead to a 17-dimensional solution space. Unless extremely tight constraints can be enforced onto the PSF and image model, global search strategies are not feasible for optimizing the likelihood function. In the following sections we will consider recursive and non-recursive algorithms, in which gradient-based optimization procedures play a prominent role.

6.3 Implementations for Noiseless Data

In this section we will discuss two different implementations of the likelihood estimator which are based on the assumption of noiseless observed image data, i.e. $\sigma_w^2 = 0$. It will be shown that under this assumption the ML problem can be reformulated as a least-squares problem.

6.3.1 Least-Squares Solution

In the pioneering work of Tekalp *et al.* [96] the identification of the image model and PSF coefficients was considered for the case of noiseless data. They showed that by combining the image model (2.13) and the observation equation (2.8), the image identification problem can be considered equivalent to the identification of a 2-D autoregressive moving-average process of the following form:

$$g(i,j) = \sum_{k,l \in S_a} a(k,l) g(i-k, j-l) + \sum_{m,n \in S_d} d(m,n) v(i-m, j-n). \quad (6.23)$$

The AR-coefficients are determined by the image model coefficients, while the MA-coefficients are determined by the (normalized) PSF coefficients. In [96,97], Tekalp *et al.* propose recursive methods to estimate $d(m,n)$ and $a(k,l)$ from $g(i,j)$. We will show here that these estimation procedures implement the optimization of the likelihood function (6.14) under the assumption of noiseless data.

In the case $\sigma_w^2 = 0$ the covariance matrix $\boldsymbol{P}$ in (6.13) becomes:

$$\boldsymbol{P} = \sigma_v^2 \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{A})^{-1} (\boldsymbol{I} - \boldsymbol{A})^{-t} \boldsymbol{D}^t. \quad (6.24)$$

If we now assume, without loss of generality, that the PSF of the degrading system is first-quadrant causal (this can be achieved by introducing a fixed delay), that we have a causal image model, and that $\boldsymbol{D}$ and $\boldsymbol{A}$ are associated with linear convolutions, then the determinant of the covariance matrix $\boldsymbol{P}$ becomes equal to $[\sigma_v^2 d(0,0)^2]^{MN}$. (Observe that the difference between a linear and circular convolution is apparent only at the boundaries of the image; hence if we neglect boundary effects, the two approaches are identical.) The likelihood function $\mathcal{L}(\theta)$ in (6.14) can now be rewritten as follows:

$$\mathcal{L}(\theta') = MN \log(d(0,0)^2 \sigma_v^2) + \frac{1}{d(0,0)^2 \sigma_v^2} \boldsymbol{g}^t \boldsymbol{D'}^{-t} (\boldsymbol{I} - \boldsymbol{A})^t (\boldsymbol{I} - \boldsymbol{A}) \boldsymbol{D'}^{-1} \boldsymbol{g}. \quad (6.25)$$

where $\boldsymbol{D'} = \boldsymbol{D}/d(0,0)$, or $d'(m,n) = d(m,n)/d(0,0)$, indicating a normalization of the PSF in both matrix and scalar notation. The parameter vector θ' is composed of the image model parameters $a(k,l)$, the normalized PSF coefficients $d'(m,n)$, and the normalized variance $\mu = d(0,0)^2 \sigma_v^2$. Equation (6.25) shows that for a given $\boldsymbol{D'}$ and $\boldsymbol{A}$, $\mathcal{L}(\theta')$ can be optimized

analytically with respect to μ, yielding

$$\mu = d(0,0)^2\sigma_v^2 = \frac{1}{MN}\boldsymbol{g}^t\boldsymbol{D'}^{-t}(\boldsymbol{I}-\boldsymbol{A})^t(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{D'}^{-1}\boldsymbol{g}. \tag{6.26}$$

Taking into account that the sum of the PSF coefficients equals 1, and that $d(0,0) \geq 0$, both $d(0,0)$ and σ_v^2 can be obtained from (6.26). Hence, under the assumption of noiseless data, analytic expressions for the estimates of $d(0,0)$ and σ_v^2 can be found. Further it is observed from (6.25) that $\mathcal{L}(\theta')$ can be optimized independently of σ_v^2, as was already mentioned before.

The optimization of the likelihood function $\mathcal{L}(\theta')$ is equivalent to solving a least-squares problem. This can be shown by introducing a new variable ϵ, called the (prediction) error signal:

$$\epsilon = (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{D'}^{-1}\boldsymbol{g}. \tag{6.27}$$

An alternative formulation of ϵ shows that $\epsilon(i,j)$ can be computed recursively as:

$$\begin{aligned}\epsilon(i,j) &= g(i,j) - \sum_{k,l\in S_a} a(k,l)g(i-k,j-l) \\ &\quad - \sum_{m,n\in S_d\setminus\{(0,0)\}} d'(m,n)\epsilon(i-m,j-n). \end{aligned} \tag{6.28}$$

Minimizing the likelihood function $\mathcal{L}(\theta')$ in (6.25) then reduces to minimizing the variance of $\epsilon(i,j)$:

$$\begin{aligned}\mathcal{L}(\theta') &= MN\log(d(0,0)^2\sigma_v^2) + \frac{1}{d(0,0)^2\sigma_v^2}\boldsymbol{g}^t\boldsymbol{D'}^{-t}(\boldsymbol{I}-\boldsymbol{A})^t(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{D'}^{-1}\boldsymbol{g} \\ &= c_1 + \frac{1}{c_2}\boldsymbol{\epsilon}^t\boldsymbol{\epsilon} = c_1 + \frac{1}{c_2}\sum_{i,j}\epsilon(i,j)^2. \end{aligned} \tag{6.29}$$

Here c_1 and c_2 are constants which are independent of $a(k,l)$ and $d'(m,n)$. We conclude that for noiseless blurred images the original ML identification problem can be regarded as a 2-D least-squares problem. Further, (6.28) is identical to the 2-D ARMA model in (6.23) (with $\epsilon(i,j) = v(i,j)$). Therefore, optimizing the likelihood function by minimizing the variance of $\epsilon(i,j)$ can be interpreted as the problem of identifying the coefficients of a 2-D ARMA process, as was proposed by Tekalp *et al.* [96].

Standard iterative gradient-based techniques can be employed for minimizing (6.29). The gradients of $\mathcal{L}(\theta')$ can be computed analytically with respect to each of the unknowns θ'_k, and can be expressed as follows:

$$\frac{\partial \mathcal{L}(\theta')}{\partial \theta'_k} = 2 \sum_{i,j} \epsilon(i,j) \frac{\partial \epsilon(i,j)}{\partial \theta'_k}, \tag{6.30}$$

where the partial derivatives of $\epsilon(i,j)$ are recursively given by

$$\begin{aligned} \frac{\partial \epsilon(i,j)}{\partial a(k,l)} &= -g(i-k,j-l) - \sum_{p,q \in S_d \setminus \{(0,0)\}} d'(p,q) \frac{\partial \epsilon(i-p,j-q)}{\partial a(k,l)}, \\ \frac{\partial \epsilon(i,j)}{\partial d'(m,n)} &= -\epsilon(i-m,j-n) - \sum_{p,q \in S_d \setminus \{(0,0)\}} d'(p,q) \frac{\partial \epsilon(i-p,j-q)}{\partial d(m,n)}. \end{aligned} \tag{6.31}$$

In order to guarantee the stability of these recursive equations the normalized PSF $d'(m,n)$ must be a minimum phase sequence, i.e. the zeros of $d'(m,n)$ must lie within the unit bi-circle. This assumption is, however, in general not satisfied, because many PSFs (such as symmetric ones) correspond to non-minimum phase systems. Therefore Tekalp *et al.* proposed the following two computational methods: (i) decompose the unknown PSF into 4 quarter-plane convolutional factors, each of which is stable in its direction of recursion, and next identify one of these factors (using the symmetry of the factors) by the above identification method. The final PSF is formed by cascading the individual convolutional factors [96], or (ii) decompose the magnitude squared of the PSF into 2 nonsymmetric half-plane (NSHP) factors in order to identify the magnitude response of the PSF [97].

6.3.2 Parallel 1-D Least-Squares Solution

In the previous section it was shown that under the assumption of noiseless data, the ML image identication problem becomes a 2-D least-squares problem, which could also be interpreted as the identification of a 2-D ARMA model. Biemond *et al.* [13] showed that under certain conditions it is possible to decompose the 2-D ARMA model identification problem into a set of parallel 1-D (complex) ARMA model identification problems. Well-known 1-D identification methods were subsequently applied to each of the parallel channels independently to obtain the least-squares estimates

of the parameters [13,16,44]. In this section we will show that identifying these parallel 1-D ARMA models is another way to implement the optimization of the likelihood function.

We first rewrite $\mathcal{L}(\theta)$ in (6.16) as the sum of N partial likelihood functions each of which is based on a column of the image. This is done by simply splitting the summation over both the columns and rows in (6.16) into a separate summation over the rows (denoted by the summation index u), followed by a separate summation over the columns (denoted by v):

$$\begin{aligned}
\mathcal{L}(\theta) &= \sum_{u,v=(0,0)}^{(M-1)(N-1)} \left\{ \log P(u,v) + \frac{|G(u,v)|^2}{P(u,v)} \right\} \\
&= \sum_{v=0}^{N-1} \left[\sum_{u=0}^{M-1} \left\{ \log P_v(u) + \frac{|G_v(u)|^2}{P_v(u)} \right\} \right] \\
&= \sum_{v=0}^{N-1} \left[\log \prod_{u=0}^{M-1} P_v(u) + \sum_{u=0}^{M-1} \frac{|G_v(u)|^2}{P_v(u)} \right] \\
&= \sum_{v=0}^{N-1} \left[\log \det |\boldsymbol{P}_v| + \boldsymbol{g}_v^t \boldsymbol{P}_v^{-1} \boldsymbol{g}_v \right] \\
&= \sum_{v=0}^{N-1} \mathcal{L}_v(\theta). \qquad (6.32)
\end{aligned}$$

Here $G_v(u)$ denotes the v-th column of $G(u,v)$, and the mixed spatial-frequency domain vector $\boldsymbol{g}_v = (g_v(0), g_v(1), \ldots, g_v(M-1))^t$ denotes the inverse 1-D DFT of $G_v(u)$. Similarly, $P_v(u)$ denotes the v-th column of $P(u,v)$, and $\boldsymbol{P}_v$ is the circulant matrix associated with the convolution kernel $p_v(i)$, which is the inverse 1-D DFT of $P_v(u)$. Referring to (6.13) and (6.17) we can write

$$\boldsymbol{P}_v = \sigma_v^2 \boldsymbol{D}_v (\boldsymbol{I} - \boldsymbol{A}_v)^{-1} (\boldsymbol{I} - \boldsymbol{A}_v)^{-t} \boldsymbol{D}_v^t + \sigma_w^2 \boldsymbol{I}, \qquad (6.33)$$

or

$$P_v(u) = \sigma_v^2 \frac{|D_v(u)|^2}{|1 - A_v(u)|^2} + \sigma_w^2, \qquad v \in [0, N-1]. \qquad (6.34)$$

Associated with these relations is the following convolution kernel for each of the columns:

$$p_v(i) = \sigma_v^2 d_v(i) * (1 - a_v(i))^{-1} * (1 - a_v(-i))^{-1} * d_v(-i) + \sigma_w^2. \qquad (6.35)$$

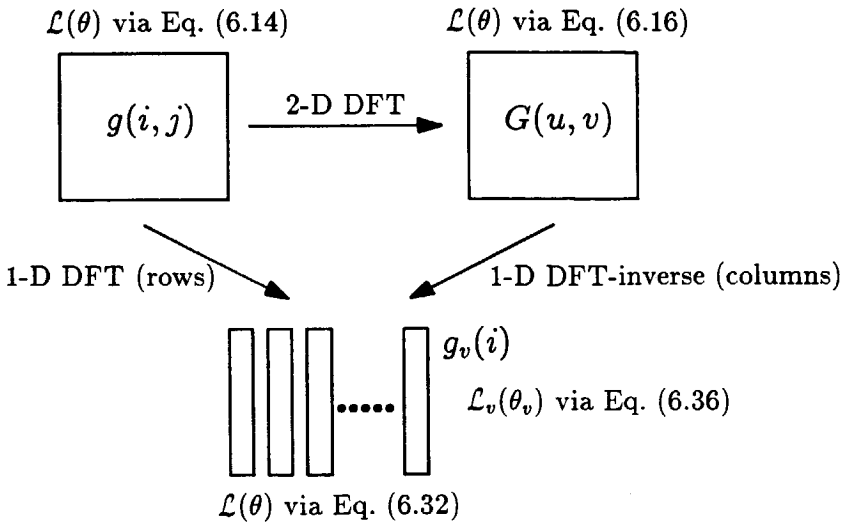


Figure 6.2: Relations between the various (equivalent) definitions of the likelihood function.

Here $D_v(u)$ and $A_v(u)$ are the v-th column of $D(u,v)$ and $A(u,v)$, respectively, and $d_v(i)$ and $a_v(i)$ are their 1-D column-wise inverse DFT. Alternatively, we may also regard $d_v(i)$ as the v-th DFT coefficient resulting from a 1-D forward DFT of the i-th row of $d(i,j)$. Similar results hold for $a_v(i), p_v(i)$ and $g_v(i)$. Figure 6.2 shows the relations between the various definitions of the likelihood function and the use of the 1-D and 2-D DFTs.

The interpretation of (6.32) is that if we apply a 1-D DFT to each of the rows of $g(i,j)$, we can compute the value of the likelihood function $\mathcal{L}(\theta)$ as the sum of the individual likelihood functions $\mathcal{L}_v(\theta)$, each of which is related to a column (or "channel") of $g(i,j)$. We would now like to optimize $\mathcal{L}(\theta)$ by optimizing each of the individual likelihood functions $\mathcal{L}_v(\theta)$ independently (possibly in parallel). This decoupled optimization cannot be performed directly by equating the gradients of $\mathcal{L}_v(\theta)$ with respect to θ to zero, because the channels are not decoupled with respect to the elements of θ. (In other words, the optimal PSF computed for instance from the data in the channel v_1 may not be optimal for the channel v_2.) However, it

can be observed from (6.35) that the channels are decoupled with regard to the coefficients $a_v(i)$ and $d_v(i)$. Therefore each $\mathcal{L}_v(\theta)$ can be optimized independently with respect to these transformed coefficients.

Unfortunately, such an independent optimization is not possible for σ_w^2 and σ_v^2 because they appear in the expressions for each $\mathcal{L}_v(\theta)$. However, if we (initially) assume that the observation noise is negligible ($\sigma_w^2 \approx 0$), it is sufficient to optimize $\mathcal{L}(\theta)$ only with respect to the image model and PSF coefficients, as was shown in Section 6.3.1. The remainder of this section therefore applies to blurred images in which the amount of noise is small.

We define $\theta_v = (d_v(i), a_v(i))^t$ $(\forall v)$, and write $\mathcal{L}_v(\theta) = \mathcal{L}_v(\theta_v)$ to stress that the likelihood function of channel v needs to be optimized only with respect to the parameter vector θ_v. If we assume, without loss of generality, that the image model and PSF have a (semi-) causal support, the 1-D convolution kernels $d_v(i)$ and $a_v(i)$ have a causal support as well. If we also assume that $d_v(i)$ and $a_v(i)$ denote linear convolutions, we can rewrite $\mathcal{L}_v(\theta_v)$ in (6.32) as follows:

$$\begin{aligned}\mathcal{L}_v(\theta_v) &= M\log(d_v(0)^2\sigma_v^2) + \frac{1}{d_v(0)^2\sigma_v^2}\boldsymbol{g}_v^t\boldsymbol{D}_v'^{-t}(\boldsymbol{I}-\boldsymbol{A}_v)^t(\boldsymbol{I}-\boldsymbol{A}_v)\boldsymbol{D}_v'^{-1}\boldsymbol{g}_v \\ &= c_1 + \frac{1}{c_2}\boldsymbol{\epsilon}_v^t\boldsymbol{\epsilon}_v = c_1 + \frac{1}{c_2}\sum_{i=0}^{M-1}\epsilon_v(i)^2. \qquad (6.36)\end{aligned}$$

Here $\boldsymbol{D}_v' = \boldsymbol{D}_v/d_v(0)$ indicates a normalization of the PSF, and $\boldsymbol{\epsilon}_v = (\boldsymbol{I}-\boldsymbol{A}_v)\boldsymbol{D}_v'^{-1}\boldsymbol{g}_v$ is the error signal in the v-th channel. Observe the correspondence between the above 1-D case and the 2-D case in (6.29).

Finally, the optimization of $\mathcal{L}(\theta)$ with respect to θ can be computed as follows:

$$\nabla_\theta\mathcal{L}(\theta) = \nabla_\theta\sum_v\mathcal{L}_v(\theta_v) = \sum_v\frac{\partial\theta}{\partial\theta_v}\nabla_{\theta_v}\mathcal{L}_v(\theta_v) = 0, \qquad (6.37)$$

where $\partial\theta/\partial\theta_v$ denotes the Jacobian matrix of the transform from θ to θ_v (i.e. the 1-D row-wise DFT). Since the parameter vectors θ_v are mutually uncoupled, an immediate solution to (6.37) is given by

$$\nabla_{\theta_v}\mathcal{L}_v(\theta_v) = 0, \qquad \forall v, \qquad (6.38)$$

or

$$\frac{\partial}{\partial a_v(k)}\mathcal{L}_v(\theta_v) = \frac{\partial}{\partial a_v(k)}\sum_{i=0}^{M-1}\epsilon_v(i)^2 = 0, \qquad \forall v, \qquad (6.39)$$

$$\frac{\partial}{\partial d_v(m)} \mathcal{L}_v(\theta_v) = \frac{\partial}{\partial d_v(m)} \sum_{i=0}^{M-1} \epsilon_v(i)^2 = 0, \qquad \forall v. \tag{6.40}$$

Hence, for noiseless data we find that the 2-D ML image identification problem can be reformulated as a set of independent parallel 1-D least-squares identification problems. After identifying the parameters $d_v(m)$ and $a_v(k)$ in each of the channels, the original parameters $d(m,n)$ and $a(k,l)$ can be found by a 1-D row-wise inverse DFT. Again additional procedures are required in order to assign a unique phase to the PSF [13]. Finally σ_v^2 can be obtained by using (6.21).

Using a discussion similar to the one in Section 6.3.1, it is straightforward to show that solving (6.39) and (6.40) is equivalent to the identification of a 1-D (complex) ARMA model using a least-squares method. In this respect the algorithm resulting from the above reformulation of the ML problem is identical to the approach originally proposed by Biemond *et al.* [13]. A variety of "standard" 1-D identification procedures can now be employed to solve $a_v(k)$ and $d_v(m)$ from (6.39) and (6.40), respectively, for each of the parallel channels. One dimensional recursive identification methods, such as the RML or RELS methods [62], have been used by Katayama *et al.* [44]; the nonrecursive Graupe, Moore and Krause technique was used by Biemond *et al.* [13], while Blanc-Fèraud *et al.* [16] considered the use of gradient-based optimization procedures.

6.4 Implementations for Noisy Data

The image identification methods for noiseless blurred images as described in the previous section, lead to acceptable results when the signal-to-noise ratio of the blurred image is relatively high. As can be expected, for lower SNRs these methods do not perform very well. Therefore, more recent reseach in image identification has been focused towards identifying noisy images. In this section we will describe two methods to optimize $\mathcal{L}(\theta)$ for noisy data. In the first method (Section 6.4.1) the gradients can be formulated analytically in the spatial domain, but need to be evaluated in the frequency domain in order to avoid operations on large matrices. In the second approach (Section 6.4.2) the likelihood function is formulated in the spatial domain through the use of the Kalman filter in two dimensions. In this case, however, analytic expressions for the gradients cannot be obtained.

6.4.1 Gradient-based Iterative Optimization

In this section we consider the use of the following standard gradient-based iterative optimization algorithm to maximize $\mathcal{L}(\theta)$ [52]:

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \beta \nabla_\theta \mathcal{L}(\hat{\theta}^{(k)}), \tag{6.41}$$

where the gradient of $\mathcal{L}(\theta)$ is defined by

$$\begin{aligned} \nabla_\theta \mathcal{L}(\theta) &= \left[\frac{\partial \mathcal{L}(\theta)}{\partial \theta_1}, \frac{\partial \mathcal{L}(\theta)}{\partial \theta_2}, \ldots, \frac{\partial \mathcal{L}(\theta)}{\partial \theta_M} \right]^t \\ &= \left[\frac{\partial \mathcal{L}(\theta)}{\partial d(m,n)}, \frac{\partial \mathcal{L}(\theta)}{\partial a(k,l)}, \frac{\partial \mathcal{L}(\theta)}{\partial \sigma_v}, \frac{\partial \mathcal{L}(\theta)}{\partial \sigma_w} \right]^t, \end{aligned} \tag{6.42}$$

and β controls the convergence of the iterations. For the image identification problem the partial derivative of $\mathcal{L}(\theta)$ with respect to one of the elements of θ (say θ_i) is given by:

$$\begin{aligned} \frac{\partial \mathcal{L}(\theta)}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i}\{\log(\det |\boldsymbol{P}|)\} + \frac{\partial}{\partial \theta_i}\{\boldsymbol{g}^t \boldsymbol{P}^{-1} \boldsymbol{g}\} \\ &= \text{tr}\left\{ \frac{\partial \boldsymbol{P}}{\partial \theta_i} \boldsymbol{P}^{-1} \right\} - \text{tr}\left\{ \boldsymbol{g} \boldsymbol{P}^{-1} \frac{\partial \boldsymbol{P}}{\partial \theta_i} \boldsymbol{P}^{-1} \boldsymbol{g}^t \right\}, \end{aligned} \tag{6.43}$$

where $\text{tr}\{\boldsymbol{X}\}$ denotes the trace of the matrix $\boldsymbol{X}$. Thus, the iterative optimization algorithm (6.41) can be realized, if we can evaluate the partial derivatives of $\boldsymbol{P}$ with respect to θ_i.

Since $\boldsymbol{P}$ itself is quadratic in $d(m,n)$, $a(k,l)$, σ_v and σ_w, the evaluation of $\partial \boldsymbol{P}/\partial \theta_i$ is relatively simple, yielding:

$$\frac{\partial \boldsymbol{P}}{\partial d(m,n)} = \sigma_v^2 (\boldsymbol{I} - \boldsymbol{A})^{-1} (\boldsymbol{I} - \boldsymbol{A})^{-t} \{ \boldsymbol{S}_{m,n} \boldsymbol{D}^t + \boldsymbol{S}_{-m,-n} \boldsymbol{D} \}, \tag{6.44}$$

$$\begin{aligned} \frac{\partial \boldsymbol{P}}{\partial a(k,l)} = \sigma_v^2 \boldsymbol{D} (\boldsymbol{I} - \boldsymbol{A})^{-1} (\boldsymbol{I} - \boldsymbol{A})^{-t} \boldsymbol{D}^t \{ \boldsymbol{S}_{k,l} (\boldsymbol{I} - \boldsymbol{A})^{-1} \\ + \boldsymbol{S}_{-k,-l} (\boldsymbol{I} - \boldsymbol{A})^{-t} \}, \end{aligned} \tag{6.45}$$

$$\frac{\partial \boldsymbol{P}}{\partial \sigma_v} = 2\sigma_v \boldsymbol{D} (\boldsymbol{I} - \boldsymbol{A})^{-1} (\boldsymbol{I} - \boldsymbol{A})^{-t} \boldsymbol{D}^t, \tag{6.46}$$

$$\frac{\partial \boldsymbol{P}}{\partial \sigma_w} = 2\sigma_w \boldsymbol{I}. \tag{6.47}$$

Here $S_{m,n}$ is a "shift operator", which is a block-circulant matrix defined through the convolution kernel $\delta(i-m, j-n)$, i.e.

$$S_{m,n} \sim \delta(i-m, j-n) = \begin{cases} 1, & i = m \text{ and } j = n, \\ 0, & \text{elsewhere.} \end{cases} \tag{6.48}$$

The evaluation of the above equations is very impractical, because they require operations on large matrices. As an alternative (6.43) can be evaluated in the DFT domain as:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_i} = \sum_{u,v} \left\{ \left[\frac{1}{P(u,v)} - \frac{|G(u,v)|^2}{P(u,v)^2} \right] \frac{\partial P(u,v)}{\partial \theta_i} \right\}, \tag{6.49}$$

provided that we assume that the matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ have a block-circulant structure. The counterparts of (6.44)–(6.47) in the Fourier domain become:

$$\begin{aligned} \frac{\partial P(u,v)}{\partial d(m,n)} &= \sigma_v^2 \frac{1}{|1-A(u,v)|^2} \left\{ D(-u,-v) \exp\{-2\pi j(\frac{um}{M} + \frac{vn}{N})\} \right. \\ &\quad \left. + D(u,v) \exp\{2\pi j(\frac{um}{M} + \frac{vn}{N})\} \right\}, \end{aligned} \tag{6.50}$$

$$\begin{aligned} \frac{\partial P(u,v)}{\partial a(k,l)} &= \sigma_v^2 \frac{|D(u,v)|^2}{|1-A(u,v)|^2} \left\{ \frac{1}{1-A(u,v)} \exp\{-2\pi j(\frac{uk}{M} + \frac{vl}{N})\} \right. \\ &\quad \left. + \frac{1}{1-A(-u,-v)} \exp\{2\pi j(\frac{uk}{M} + \frac{vl}{N})\} \right\}, \end{aligned} \tag{6.51}$$

$$\frac{\partial P(u,v)}{\partial \sigma_v} = 2\sigma_v \frac{|D(u,v)|^2}{|1-A(u,v)|^2}, \tag{6.52}$$

$$\frac{\partial P(u,v)}{\partial \sigma_w} = 2\sigma_w. \tag{6.53}$$

In [52] Lagendijk *et al.* showed that it is possible to explicitly incorporate linear relations among the PSF and image model coefficients (such as symmetry properties of the PSF) into the computation of $\partial \boldsymbol{P}/\partial \theta_i$. In this way the uniqueness of the solution obtained by the above iterations could be guaranteed.

It is also observed that in deriving the above iterative scheme, no restrictions were imposed on the causality of the PSF or image model. Therefore these iterations can also be used to identify, for example, non-causal image models and non-causal symmetric (i.e. non-minimum phase) PSFs.

Unfortunately, a serious drawback of this method is that (6.41) represents a basic steepest descent iteration, the convergence of which is known to be slow in general. In addition to this the evaluation of the various components of the gradient of $\mathcal{L}(\theta)$ is computationally involving. In Chapter 7 we will describe another iterative optimization algorithm which is computationally more attractive than the standard iteration presented in this section.

6.4.2 Prediction Error Based Solution

The gradient-based solution of the preceding section has been implemented in the DFT domain in order to avoid huge matrix operations in the spatial domain. Alternatively it is possible to compute the likelihood function in the spatial domain by expressing the likelihood function in a recursive form.

Following Angwin [4,53], we first assume that the image is raster scanned, and that the elements of the observed noisy blurred image $\boldsymbol{g}$ become available one by one. (Note that in the derivation of the gradient-based implementations we used the matrix-vector notation, which assumes that all data $\boldsymbol{g}$ is available at once.) We can then represent the data at pixel position (i,j) in a recursive prediction error equation form [6]:

$$g(i,j) = \hat{g}(i,j;\theta) + \epsilon(i,j;\theta), \tag{6.54}$$

where $\hat{g}(i,j;\theta)$ is the predicted value of $g(i,j)$ using all observed image data up to the pixel (i,j) given the parameter vector θ, and $\epsilon(i,j;\theta)$ denotes the resulting error, called the prediction error.

A suitable prediction $\hat{g}(i,j;\theta)$ is obtained by minimizing the variance of the prediction error $\sigma^2_{\epsilon(i,j;\theta)} = \mathrm{E}[(\hat{g}(i,j;\theta) - g(i,j;\theta))^2]$. The solution to this linear minimum variance prediction problem can be computed by employing the Kalman filter in two dimensions [112]. To this end, we first represent the image model and observation equation by the following state-space equations (c.f. Chapter 3):

$$\boldsymbol{s}^{(i,j)} = \boldsymbol{T}\boldsymbol{s}^{(i,j-1)} + [1,0,0,\cdots,0]^t v(i,j), \tag{6.55}$$

$$g(i,j) = \boldsymbol{d}^t \boldsymbol{s}^{(i,j)} + w(i,j), \tag{6.56}$$

where the state at the coordinates (i,j) is denoted by $\boldsymbol{s}^{(i,j)}$. For noncausal PSFs the observed image data $g(i,j)$ needs to be delayed over a suitable

fixed distance. The estimated state $\hat{\boldsymbol{s}}^{(i,j)}$ is then given by

$$\hat{\boldsymbol{s}}_b^{(i,j)} = \boldsymbol{T}\hat{\boldsymbol{s}}_a^{(i,j-1)}, \tag{6.57}$$

$$\hat{\boldsymbol{s}}_a^{(i,j)} = \hat{\boldsymbol{s}}_b^{(i,j)} + \boldsymbol{k}^{(i,j)}[g(i,j) - \boldsymbol{d}^t\hat{\boldsymbol{s}}_b^{(i,j)}]. \tag{6.58}$$

On the basis of the above Kalman filter, we can write an expression for both the recursive prediction $\hat{g}(i,j;\theta)$ and the variance of the prediction error $\epsilon(i,j;\theta)$ as:

$$\hat{g}(i,j;\theta) = \boldsymbol{d}^t\hat{\boldsymbol{s}}_b^{(i,j)}, \tag{6.59}$$

$$\sigma^2_{\epsilon(i,j;\theta)} = \mathrm{E}[\epsilon(i,j;\theta)\epsilon(i,j;\theta)^t] = \boldsymbol{d}^t\boldsymbol{P}_b^{(i,j)}\boldsymbol{d} + \sigma_w^2. \tag{6.60}$$

Here $\boldsymbol{P}_b^{(i,j)} = \mathrm{E}[(\boldsymbol{s}_b^{(i,j)} - \hat{\boldsymbol{s}}_b^{(i,j)})(\boldsymbol{s}_b^{(i,j)} - \hat{\boldsymbol{s}}_b^{(i,j)})^t]$ denotes the error covariance matrix of the state at the pixel (i,j) after the prediction step, and can be obtained immediately from the Kalman filter.

Given that $\hat{g}(i,j;\theta)$ is the linear minimum variance prediction of $g(i,j)$, and the fact that $g(i,j)$ has a Gaussian distribution, we can conclude that $\epsilon(i,j;\theta)$ is a white process uncorrelated with $\hat{g}(i,j;\theta)$, and that its PDF $p(\epsilon(i,j);\theta)$ is Gaussian with zero mean and variance $\sigma^2_{\epsilon(i,j;\theta)}$ [112]:

$$p(\epsilon(i,j);\theta) = \left[2\pi\sigma^2_{\epsilon(i,j;\theta)}\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{\epsilon(i,j;\theta)^2}{\sigma^2_{\epsilon(i,j;\theta)}}\right\}. \tag{6.61}$$

Using the above recursive formulation of $g(i,j)$, we can express the likelihood function in a recursive form as well. To this end we let $\boldsymbol{g}^{(m,n)} = [g(0,0), g(0,1), \ldots, g(m,n)]^t$, i.e. the image vector containing all the observed data up to and including $g(m,n)$, and $\mathcal{L}^{(m,n)}(\theta) = \log\{p(\boldsymbol{g}^{(m,n)};\theta)\}$ denotes the log-likelihood function of the observed data up to and including $g(m,n)$. By using the following general relation for evaluating a joint PDF:

$$p(\boldsymbol{g}^{(m,n)};\theta) = p(g(m,n)/\boldsymbol{g}^{(m,n-1)};\theta)\, p(\boldsymbol{g}^{(m,n-1)};\theta), \tag{6.62}$$

we can rewrite $\mathcal{L}^{(m,n)}(\theta)$ as the product of the log-likelihood functions of $\epsilon(i,j;\theta)$ [6]:

$$\begin{aligned}\mathcal{L}^{(m,n)}(\theta) &= \log\{p(\boldsymbol{g}^{(m,n)};\theta)\} \\ &= \log\{p(g(m,n)/\boldsymbol{g}^{(m,n-1)};\theta)\, p(\boldsymbol{g}^{(m,n-1)};\theta)\} \\ &= \log\{p(g(m,n) - \hat{g}(m,n);\theta)\, p(\boldsymbol{g}^{(m,n-1)};\theta)\} = \cdots =\end{aligned}$$

$$= \log \prod_{(i,j)=(0,0)}^{(m,n)} p(g(i,j) - \hat{g}(i,j);\theta)$$

$$= \sum_{(i,j)=(0,0)}^{(m,n)} \log\{p(g(i,j) - \hat{g}(i,j);\theta)\}$$

$$= \sum_{(i,j)=(0,0)}^{(m,n)} \log\{p(\epsilon(i,j);\theta)\}. \tag{6.63}$$

Substituting the PDF $p(\epsilon(i,j);\theta)$ into (6.63), dropping all constant terms, and premultiplying the result by -2 leads to [4,6,53]:

$$\mathcal{L}^{(m,n)}(\theta) = \sum_{(i,j)=(0,0)}^{(m,n)} \left\{ \log \sigma^2_{\epsilon(i,j;\theta)} + \frac{(g(i,j) - \hat{g}(i,j;\theta))^2}{\sigma^2_{\epsilon(i,j;\theta)}} \right\}. \tag{6.64}$$

Equation (6.64) shows that the likelihood function can be evaluated as a sum of scalar functions, which is considerably simpler than the direct evaluation of (6.13) and (6.14). Substituting (6.59) and (6.60) into (6.64) yields a computable recursive expression for the likelihood function, in which the Kalman filter in two dimensions is utilized to recursively compute the prediction errors and the prediction error variances. Optimizing $\mathcal{L}(\theta)$ in (6.14), which uses all the available data $g(i,j)$, $i \in [0, M-1], j \in [0, N-1]$, is now equivalent to optimizing $\mathcal{L}^{(M-1,N-1)}(\theta)$ given by (6.64).

It should be noted, however, that the recursive computation of the likelihood function does not lead to an identification procedure which is recursive in the parameters to be estimated. For example, suppose that the likelihood function $\mathcal{L}^{(m,n)}(\theta)$ has been optimized with respect to θ. Then the addition of a single new observed data point $g(m, n+1)$ to (6.64) requires a new likelihood function $\mathcal{L}^{(m,n+1)}(\theta)$ to be optimized, the result of which cannot be expressed analytically in terms of the parameter estimates at (m,n). Thus, although a finite memory (Kalman) filter is employed to compute the likelihood function, this solution strategy does not (easily) lead to a finite memory parameter identification scheme.

In [4] (6.64) is optimized by an iterative gradient-based procedure. This, however, shows another problem with the above formulation, namely the partial derivatives of the likelihood function (6.64) cannot be obtained analytically because of the nonlinearities introduced by the Ricatti equation. Therefore, numerical procedures have to be utilized in order to compute the partial derivatives of the likelihood function with respect to each

of the unknown parameters. This, however, involves filtering the blurred image a number of times. Therefore even if efficient implementations are employed [4,112], the use of the Kalman filter usually requires more computation than the frequency domain approach, especially for large PSFs. Note that the technique described in this section is essentially a mixed iterative-recursive identification method, where numerical gradients are used to iteratively optimize the likelihood function which is computed recursively.

Chapter 7

Image Identification Using the EM-Algorithm

In the preceding chapter we have seen that the image identification problem can be formulated as a maximum likelihood parameter estimation problem. The optimization of the associated likelihood function turned out to be not a trivial problem, because of its highly nonlinear character, in a relatively large number of unknowns. For noiseless blurred images the ML estimation problem was shown to be identical to a least-squares estimation problem, for which elegant optimization algorithms were proposed in the literature. If, however, the likelihood function for noisy data is considered, these algorithms are no longer applicable, and gradient-based optimization strategies need to be employed, using either a frequency domain expression or a recursive spatial domain relation for the likelihood function. Disadvantages of these methods are that they are computationally involving, and that convergence of the gradient-based optimization is often difficult to control due to the nonlinear properties of the likelihood function.

In this chapter we present an image identification algorithm that is based on the expectation-maximization (EM) algorithm to compute ML parameter estimates [22]. The presence of noise in the observed data is directly taken into account. The advantage of this iterative algorithm is that it avoids operating directly on the nonlinear likelihood function in (6.14). Essentially, it alternates between the solving of a relatively simple identification problem (yielding a new set of parameters θ), and the restoration of the blurred image (yielding new input for the identification stage). Each of the two stages involved requires the solving of linear equations only [53,54,56,57]. In addition the identification algorithm is very

flexible in incorporating a number of modifications, as will be discussed in the next chapter.

In Section 7.1 we will start with a brief review of the EM-algorithm. Next, this algorithm will be applied to the ML image identification problem in Section 7.2. The two stages involved in the EM-algorithm, the E-step and the M-step, will be discussed in detail in Sections 7.3 and 7.4, respectively. Finally, this chapter is concluded with an experimental evaluation of the proposed identification algorithm. The EM-algorithm will also be compared experimentally with the least-squares implementation of the ML estimator for noiseless data described in Section 6.3.1.

7.1 Review of the EM-Algorithm

It has long since been recognized that computing maximum likelihood (ML) parameter estimates can be a highly complicated task in many relevant estimation problems. The EM-algorithm, presented by Dempster *et al.* in [22], is a general iterative method to compute ML estimates if the observed data can be regarded "incomplete". Ever since similar or related algorithms have appeared in the literature, such as in [26,71,86]. It has been shown in various signal processing applications that the use of the EM-like algorithms leads to computationally efficient estimation algorithms, such as in 1-D ARMA process identification [20,21,71], parameter estimation of superimposed signals [27], maximum likelihood noise cancellation [26], and tomographic image reconstruction [80].

Let $\mathcal{Y}$ denote the observed "incomplete" data which possesses the PDF $p(\mathcal{Y};\theta)$, where θ again is the vector of parameters to be estimated. The ML estimator of θ, based on the available incomplete data, is given by

$$\hat{\theta}_{\mathrm{ml},\mathcal{Y}} = \arg\left\{\max_{\theta\in\Theta} \mathcal{L}_{\mathcal{Y}}(\theta)\right\} = \arg\left\{\max_{\theta\in\Theta} \log p(\mathcal{Y};\theta)\right\}, \tag{7.1}$$

which is a complicated problem in general. The incomplete data is related to some complete data, denoted by $\mathcal{X}$, through a non-invertible many-to-one transformation $\mathcal{T}$:

$$\mathcal{Y} = \mathcal{T}(\mathcal{X}). \tag{7.2}$$

It is noted that the choice of the complete data is not unique, though some choices may be more succesful than others. We restrict ourselves here to cases where $\mathcal{T}$ represents a linear projection; for each given $\mathcal{Y}$ there exists

a hyperplane $\mathcal{X}(\mathcal{Y})$ whose elements satisfy (7.2). The PDF of $\mathcal{X}$, which is also indexed by θ, is related to $\mathcal{Y}$ as follows:

$$p(\mathcal{Y};\theta) = \int_{\mathcal{X}(\mathcal{Y})} p(\mathcal{X};\theta) d\mathcal{X}. \tag{7.3}$$

At this point it is assumed that the complete data has been chosen in such a way that computing the ML estimator of θ from the complete data, i.e. solving

$$\hat{\theta}_{\mathrm{ml},\mathcal{X}} = \arg\left\{\max_{\theta\in\Theta} \mathcal{L}_{\mathcal{X}}(\theta)\right\} = \arg\left\{\max_{\theta\in\Theta} \log p(\mathcal{X};\theta)\right\}, \tag{7.4}$$

is significantly simpler than solving (7.1). However, the complete data is not available, but only observed via the non-invertible relation (7.2). Starting out with an estimate of the parameter vector, called $\hat{\theta}^{(k)}$, the EM-algorithm finds the conditional expectation of the log-likelihood of complete data, denoted by $\mathcal{L}(\theta;\hat{\theta}^{(k)})$, given the observed incomplete data and $\hat{\theta}^{(k)}$:

$$\begin{aligned}\mathcal{L}(\theta;\hat{\theta}^{(k)}) &= \mathrm{E}\left\{\log p(\mathcal{X};\theta)/\mathcal{Y};\hat{\theta}^{(k)}\right\} \\ &= \int_{\mathcal{X}(\mathcal{Y})} \log\{p(\mathcal{X};\theta)\}\, p(\mathcal{X}/\mathcal{Y};\hat{\theta}^{(k)})\, d\mathcal{X},\end{aligned} \tag{7.5}$$

where $p(\mathcal{X}/\mathcal{Y};\hat{\theta}^{(k)})$ is the conditional PDF of the complete data, given the incomplete data and the estimate $\hat{\theta}^{(k)}$. Equation (7.5) is called the E-step of the EM-algorithm. In the M-step $\mathcal{L}(\theta;\hat{\theta}^{(k)})$ is maximized with respect to θ. This leads to a new parameter estimate $\hat{\theta}^{(k+1)}$:

$$\hat{\theta}^{(k+1)} = \arg\left\{\max_{\theta\in\Theta} \mathcal{L}(\theta;\hat{\theta}^{(k)})\right\}. \tag{7.6}$$

By alternating (7.5) and (7.6) the iterative EM-algorithm is obtained, which converges to a stationary point of $\mathcal{L}_{\mathcal{Y}}(\theta)$ [22]. Sufficient convergence conditions are that $\mathcal{L}(\theta;\hat{\theta}^{(k)})$, which forms the basis of the EM-algorithm, is continuous in both θ and $\hat{\theta}^{(k)}$ [22,26,114]. At each iteration cycle the maximization in (7.6) ensures that $\mathcal{L}_{\mathcal{Y}}(\theta)$ increases. This maximization requirement can be relaxed somewhat. Namely, the EM-algorithm converges as long as $\hat{\theta}^{(k+1)}$ is chosen in such a way that $\mathcal{L}_{\mathcal{Y}}(\theta)$ is increased [22]. This variant is called a Generalized EM (GEM) algorithm. It is noted here that like many other hill-climbing optimization techniques, the EM-algorithm does not necessarily converge to the global optimum of $\mathcal{L}_{\mathcal{Y}}(\theta)$, but instead

it may stabilize at a local optimum. In such situations various starting points $\hat{\theta}^{(0)}$ may be needed. Since the EM-based iterative scheme is not controlled by a convergence parameter β, such as in the gradient-based optimization of (6.41), its convergence is very robust.

It is observed that $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ and $\mathcal{L}_{\mathcal{X}}(\theta)$ have the same dependence on θ [22,26], and that where $\mathcal{L}_{\mathcal{X}}(\theta)$ is defined on the true complete data, $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ uses the conditional expectation of the complete data. The maximization of $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ with respect to θ is therefore of the same complexity as the maximization of $\mathcal{L}_{\mathcal{X}}(\theta)$. Because of this, the EM-algorithm is an attractive alternative to the direct evaluation of (7.1) only if the solution to (7.4) can be computed relatively easily. Since the complexity of (7.4) depends immediately on the choice of $\mathcal{X}$, it now becomes apparent that the choice of the complete data is of crucial importance in the use of the EM-algorithm.

7.2 EM-Algorithm Applied to Image Identification

In image restoration and identification problems the noisy blurred image is the only data available, which therefore establishes the incomplete data. In [53,57] Lagendijk *et al.* specify the complete data as the stacked lexicographically ordered images $\boldsymbol{f}$ and $\boldsymbol{g}$:

$$\mathcal{X} = \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{bmatrix}. \tag{7.7}$$

The complete and incomplete data are now related via

$$\mathcal{Y} = \boldsymbol{g} = \begin{bmatrix} \boldsymbol{O} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{g} \end{bmatrix} = \mathcal{T}\mathcal{X}, \tag{7.8}$$

where $\boldsymbol{O}$ and $\boldsymbol{I}$ are an $MN \times MN$ identity and zero matrix, respectively. The particular choice of this complete data satisfies the requirement that the solving of the ML problem (7.4) should be easy, namely:

- The image model coefficients and the related modeling error variance σ_v^2 can be obtained immediately from the (linear) 2-D Yule-Walker equations, because the autocorrelation coefficients can be computed directly from the original image $\boldsymbol{f}$.

- The point-spread function and the related observation noise variance σ_w^2 follow directly from a classical system identification problem, namely determining the impulse response of a linear system with known input (the original image) and noisy output (the observed image).

It is pointed out that it is not necessary to include the noise processes $\boldsymbol{v}$ and $\boldsymbol{w}$ into the definition of the complete data, because these do not provide any additional useful information for the ML identification. Therefore the use of other choices for $\mathcal{X}$ and $\mathcal{Y}$, such as considered in [48], turns out to be less successful.

Substituting the above choices for $\mathcal{X}$ and $\mathcal{Y}$ into (7.5) yields the following E-step:

$$\begin{aligned} \mathcal{L}(\theta;\hat{\theta}^{(k)}) &= \mathrm{E}\left\{\log p(\boldsymbol{f},\boldsymbol{g};\theta)/\boldsymbol{g};\hat{\theta}^{(k)}\right\} \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\log p(\boldsymbol{f},\boldsymbol{g};\theta)\; p(\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)})\, df(1)\cdots df(MN). \end{aligned} \tag{7.9}$$

We now need to evaluate $\mathcal{L}(\theta;\hat{\theta}^{(k)})$ for the case at hand. The joint probability density function $p(\boldsymbol{f},\boldsymbol{g};\theta)$ is readily found from (6.8) and (6.10):

$$\begin{aligned} p(\boldsymbol{f},\boldsymbol{g};\theta) = p(\boldsymbol{g}/\boldsymbol{f};\theta)\,p(\boldsymbol{f};\theta) = \sqrt{\frac{\det|\boldsymbol{I}-\boldsymbol{A}|^2}{2\pi^{2MN}\det|\boldsymbol{Q}_v\boldsymbol{Q}_w|}} \\ \exp\left\{-\frac{1}{2}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})^t\boldsymbol{Q}_w^{-1}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f}) - \frac{1}{2}\boldsymbol{f}^t(\boldsymbol{I}-\boldsymbol{A})^t\boldsymbol{Q}_v^{-1}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{f}\right\}. \end{aligned} \tag{7.10}$$

By combining $p(\boldsymbol{f},\boldsymbol{g};\theta)$ and $p(\boldsymbol{g};\theta)$, which is Gaussian with covariance $\boldsymbol{P}$ given by (6.13), we get the following expression for the conditional PDF $p(\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)})$:

$$\begin{aligned} p(\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)}) &= \frac{p(\boldsymbol{f},\boldsymbol{g};\hat{\theta}^{(k)})}{p(\boldsymbol{g};\hat{\theta}^{(k)})} \\ &= \frac{1}{\sqrt{2\pi^{MN}\det|\hat{\boldsymbol{V}}^{(k)}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{f}-\hat{\boldsymbol{f}}^{(k)})^t\hat{\boldsymbol{V}}^{(k)^{-1}}(\boldsymbol{f}-\hat{\boldsymbol{f}}^{(k)})\right\}. \end{aligned} \tag{7.11}$$

Here $\hat{\boldsymbol{f}}^{(k)}$ and $\hat{\boldsymbol{V}}^{(k)}$ denote the conditional mean and covariance matrix of $\boldsymbol{f}$, respectively, at the k-th iteration (that is, they are based on $\hat{\theta}^{(k)}$):

$$\hat{\boldsymbol{f}}^{(k)} = \mathrm{E}(\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)}) = \hat{\boldsymbol{V}}^{(k)}\boldsymbol{D}^t\boldsymbol{Q}_w^{-1}\boldsymbol{g}, \tag{7.12}$$

$$\hat{\boldsymbol{V}}^{(k)} = \mathrm{Cov}(\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)}) = \left[(\boldsymbol{I}-\boldsymbol{A})^t\boldsymbol{Q}_v^{-1}(\boldsymbol{I}-\boldsymbol{A}) + \boldsymbol{D}^t\boldsymbol{Q}_w^{-1}\boldsymbol{D}\right]^{-1}. \tag{7.13}$$

(We have dropped the index (k) of $\boldsymbol{D}$, $\boldsymbol{A}$, $\boldsymbol{Q}_w$ and $\boldsymbol{Q}_v$ in order to keep these equations readable.) Finally, substitution of $p(\boldsymbol{f},\boldsymbol{g};\theta)$ and $p(\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)})$ into (7.9) yields (see Appendix C):

$$\begin{aligned}\mathcal{L}(\theta;\hat{\theta}^{(k)}) = \\ c_1 - MN\log(\sigma_v^2\sigma_w^2) + \log\det|\boldsymbol{I}-\boldsymbol{A}|^2 - \frac{1}{\sigma_w^2}\boldsymbol{g}^t\boldsymbol{g} + \frac{2}{\sigma_w^2}\mathrm{tr}\{\boldsymbol{D}\hat{\boldsymbol{R}}_{fg}^{(k)}\} + \\ - \frac{1}{\sigma_w^2}\mathrm{tr}\{\boldsymbol{D}\hat{\boldsymbol{R}}_{ff}^{(k)}\boldsymbol{D}^t\} - \frac{1}{\sigma_v^2}\mathrm{tr}\{(\boldsymbol{I}-\boldsymbol{A})\hat{\boldsymbol{R}}_{ff}^{(k)}(\boldsymbol{I}-\boldsymbol{A})^t\},\end{aligned} \tag{7.14}$$

where c_1 is an additive constant term, and where the conditional autocorrelation matrix $\hat{\boldsymbol{R}}_{ff}^{(k)}$ and crosscorrelation matrix $\hat{\boldsymbol{R}}_{fg}^{(k)}$ are defined by

$$\hat{\boldsymbol{R}}_{ff}^{(k)} = \mathrm{E}(\boldsymbol{f}\boldsymbol{f}^t/\boldsymbol{g};\hat{\theta}^{(k)}) = \hat{\boldsymbol{V}}^{(k)} + \hat{\boldsymbol{f}}^{(k)}\hat{\boldsymbol{f}}^{(k)t}, \tag{7.15}$$

$$\hat{\boldsymbol{R}}_{fg}^{(k)} = \mathrm{E}(\boldsymbol{f}\boldsymbol{g}^t/\boldsymbol{g};\hat{\theta}^{(k)}) = \hat{\boldsymbol{f}}^{(k)}\boldsymbol{g}^t. \tag{7.16}$$

Under the conditions that $\sigma_w^2 > 0$, $\sigma_v^2 > 0$, and that $(\boldsymbol{I}-\boldsymbol{A})$ is nonsingular, $\mathcal{L}(\theta;\hat{\theta}^{(k)})$ is continuous in θ and $\hat{\theta}^{(k)}$, and the EM-algorithm (7.9), (7.6) will converge. Note that these conditions were already assumed in the development of (6.8) and (6.10), and thus do not impose additional constraints.

7.3 The E-step of the Algorithm

Equation (7.14) depends on $\hat{\theta}^{(k)}$ only through the conditional autocorrelation matrix $\hat{\boldsymbol{R}}_{ff}^{(k)}$ and crosscorrelation matrix $\hat{\boldsymbol{R}}_{fg}^{(k)}$. For this reason the E-step of the algorithm consists of the mere evaluation of (7.12) and (7.13), and the computation of the conditional correlation matrices. Observe that this requires the solving of linear equations only.

If we assume that $\boldsymbol{A}$ and $\boldsymbol{D}$ have block-circulant structures, (7.12) and (7.13) can be implemented efficiently by using 2-D DFTs. As a consequence

$\hat{\boldsymbol{R}}_{ff}^{(k)}$ and $\hat{\boldsymbol{R}}_{fg}^{(k)}$ represent correlation matrices of weakly jointly stationary processes, and have block-circulant structures as well. We may therefore replace (7.15) and (7.16) by

$$\hat{r}_{ff}^{(k)}(p,q) = \hat{V}^{(k)}(p,q) + \frac{1}{MN} \sum_{(i,j)=(0,0)}^{(M-1,N-1)} \hat{f}^{(k)}(i,j)\, \hat{f}^{(k)}(i-p, j-q), \tag{7.17}$$

$$\hat{r}_{fg}^{(k)}(p,q) = \frac{1}{MN} \sum_{(i,j)=(0,0)}^{(M-1,N-1)} \hat{f}^{(k)}(i,j)\, g(i-p, j-q), \tag{7.18}$$

where $\hat{r}_{ff}^{(k)}(p,q)$ and $\hat{r}_{fg}^{(k)}(p,q)$ are the defining sequences of $\hat{\boldsymbol{R}}_{ff}^{(k)}$ and $\hat{\boldsymbol{R}}_{fg}^{(k)}$, respectively, and where $\hat{V}^{(k)}(p,q)$ is the defining sequence of $\hat{\boldsymbol{V}}^{(k)}$.

Observe that (7.12) is essentially an image restoration filter belonging to the Tikhonov-Miller class, where the image model is used as a regularizing operator (c.f. Section 3.3). Hence, in the process of identifying the image and blur parameters θ, a restoration result of the blurred image is obtained simultaneously in each E-step of the iterations. This can be considered one of the advantages of the EM-algorithm based approach towards image identification, since the quality of a restoration result using the current parameter estimates can be evaluated at each iteration step, and the iterative process can be truncated when a (visually) stationary or acceptable solution has been reached.

The equations (7.12), (7.13), (7.15) and (7.16) which form the E-step of the algorithm, have been efficiently implemented in the DFT domain by making use of the assumed block-circulant structure of the matrices that are involved in these equations. Alternatively the recursive counterpart of the Tikhonov-Miller restoration filters, i.e. the Kalman filter in two dimensions, may be employed to implement (7.12) and (7.13) in the spatial domain, depending on the available computational resources.

7.4 The M-step of the Algorithm

In the M-step $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ is maximized with respect to the parameters θ, yielding a new estimate $\hat{\theta}^{(k+1)}$. As can be seen directly from (7.14), the optimizations with respect to the image model coefficients and PSF coefficients are independent problems, and can be considered separately.

7.4.1 Image Model Identification

By substituting (7.14) into (7.6), and by dropping all constant terms that are image model independent, we arrive at the following optimization problem:

$$\{\hat{a}(k,l), \hat{\sigma}_v^2\} \leftarrow \max_{a(k,l),\sigma_v^2} J(\boldsymbol{A}, \sigma_v^2) = \max_{a(k,l),\sigma_v^2} \Big\{ \log\det |\boldsymbol{I} - \boldsymbol{A}|^2 + \\ - MN \log \sigma_v^2 - \frac{1}{\sigma_v^2} \mathrm{tr}\left\{ (\boldsymbol{I} - \boldsymbol{A}) \hat{\boldsymbol{R}}_{ff}^{(k)} (\boldsymbol{I} - \boldsymbol{A})^t \right\} \Big\}. \tag{7.19}$$

If we now restrict ourselves to causal image model supports S_a, and neglect the image model boundary effects, the matrix $(\boldsymbol{I} - \boldsymbol{A})$ is a lower triangular matrix and $\det |\boldsymbol{I} - \boldsymbol{A}| = 1$. Then (7.19) becomes quadratic in the image model coefficients $a(k,l)$, and its solution is given by the 2-D Yule-Walker equations:

$$\hat{r}_{ff}^{(k)}(p,q) = \sum_{k,l \in S_a} \hat{a}(k,l)\, \hat{r}_{ff}^{(k)}(p-k, q-l), \quad \forall (p,q) \in S_a, \tag{7.20}$$

$$\hat{\sigma}_v^2 = \hat{r}_{ff}^{(k)}(0,0) - \sum_{k,l \in S_a} \hat{a}(k,l)\, \hat{r}_{ff}^{(k)}(k,l). \tag{7.21}$$

Observe that (7.20) and (7.21) solve essentially the same model fitting problem as described in Section 2.2.3. In this case, however, the image model coefficients are fitted onto the second-order statistics of a restored image instead of the original image.

7.4.2 Blur Model Identification

The PSF coefficients and the observation noise variance are computed by maximizing (7.14) with respect to $d(m,n)$ and σ_w^2. Substituting (7.14) into (7.6), and dropping all blur model independent terms, yields

$$\{\hat{d}(m,n), \hat{\sigma}_w^2\} \leftarrow \max_{d(m,n),\sigma_w^2} J(\boldsymbol{D}, \sigma_w^2) = \max_{d(m,n),\sigma_w^2} \Big\{ -MN \log \sigma_w^2 + \\ - \frac{1}{\sigma_w^2} \boldsymbol{g}^t \boldsymbol{g} + \frac{2}{\sigma_w^2} \mathrm{tr}\left\{ \boldsymbol{D} \hat{\boldsymbol{R}}_{fg}^{(k)} \right\} - \frac{1}{\sigma_w^2} \mathrm{tr}\left\{ \boldsymbol{D} \hat{\boldsymbol{R}}_{ff}^{(k)} \boldsymbol{D}^t \right\} \Big\}. \tag{7.22}$$

In order to obtain a unique solution to (7.22) the PSF is enforced to be symmetric (c.f. Section 6.2.3):

$$\hat{d}(m,n) = \hat{d}(-m,-n), \qquad \forall (m,n) \in S_{\hat{d}}\,. \tag{7.23}$$

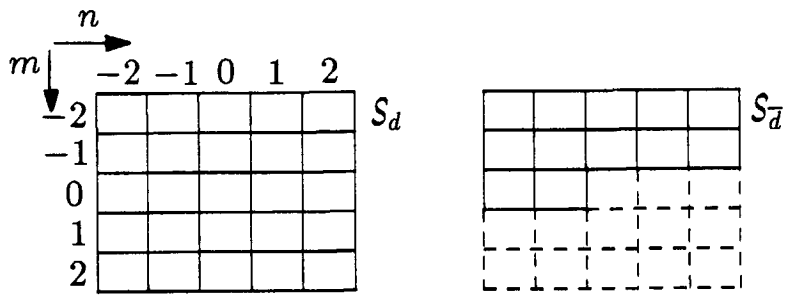


Figure 7.1: NSHP subset $S_{\overline{d}}$ of S_d containing the unique PSF coefficients.

Further, to satisfy the condition that the sum of the PSF coefficients equals 1.0, we choose $\hat{d}(0,0)$ as follows:

$$\hat{d}(0,0) = 1.0 - 2 \sum_{m,n \in S_{\overline{d}}} \hat{d}(m,n). \tag{7.24}$$

Here $S_{\overline{d}}$ is that subset of S_d which contains the unique PSF-defining coefficients $d(m,n)$ (e.g. $S_{\overline{d}}$ is a non-symmetric half plane, see Figure 7.1). All other PSF coefficients can be derived from these defining coefficients via (7.23) and (7.24). Even after substituting the above (linear) constraints into (7.22), $J(\boldsymbol{D}, \sigma_w^2)$ is quadratic in $d(m,n)$. Solving (7.22) is tedious but straightforward, and leads to the following set of linear equations:

$$\begin{aligned} &\hat{r}_{fg}^{(k)}(p,q) + \hat{r}_{fg}^{(k)}(-p,-q) - 2\hat{r}_{fg}^{(k)}(0,0) - 2\hat{r}_{ff}^{(k)}(p,q) + 2\hat{r}_{ff}^{(k)}(0,0) = \\ &\quad 2 \sum_{m,n \in S_{\overline{d}}} \hat{d}(m,n) \left\{ \hat{r}_{ff}^{(k)}(p-m,q-n) + \hat{r}_{ff}^{(k)}(p+m,q+n) + \right. \\ &\quad \left. - 2\hat{r}_{ff}^{(k)}(m,n) - 2\hat{r}_{ff}^{(k)}(p,q) + 2\hat{r}_{ff}^{(k)}(0,0) \right\}, \qquad \forall (p,q) \in S_{\overline{d}}, \end{aligned} \tag{7.25}$$

$$\hat{\sigma}_w^2 = \frac{1}{MN} \sum_{(i,j)=(0,0)}^{(M-1,N-1)} g(i,j)^2 - \sum_{m,n \in S_d} \hat{d}(m,n)\, \hat{r}_{fg}^{(k)}(-m,-n). \tag{7.26}$$

We note here that other relevant linear constraints, such as additional symmetry properties of the PSF, can be incorporated into (7.22) as well

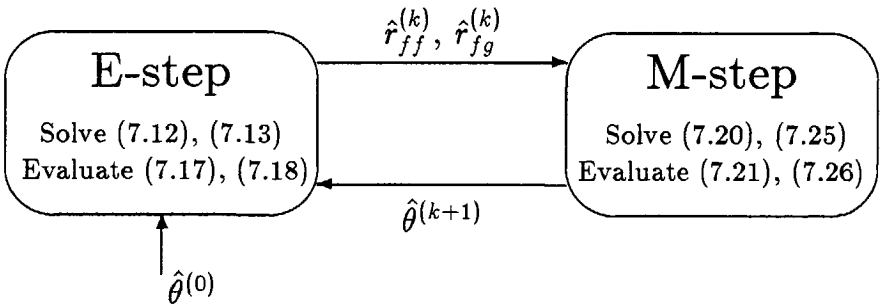


Figure 7.2: Structure of the entirely linear expectation-maximization image identification algorithm.

without significantly increasing the computational complexity of the PSF identification. It is also straightforward to incorporate the decomposition of PSFs into 4 quarter-plane or 2 NSHP convolutional factors as proposed in [96,97], or the dyadic factoring of PSFs as in [107].

In the preceding sections we have established an expectation-maximization image identification algorithm. Whereas the original ML image identification formulation required the solving of a highly complex nonlinear optimization problem, the proposed iterative procedure involves the solving of linear equations only (see Figure 7.2). In addition a restoration result of the noisy blurred image is obtained simultaneously in each cycle of the algorithm.

7.5 Experimental Results

In this section we will evaluate the performance of ML image identification using the EM-algorithm. Three aspects which are encountered in the practical use of the EM-algorithm and ML image identification in general, are (i) the initialization of the PSF coefficients, (ii) the existence of local optima in the likelihood function, and (iii) the estimation of the size of the PSF support (i.e. the number of unknown PSF coefficients). These items will be addressed experimentally in this section, and it will be shown that

heuristic arguments may provide answers to these questions. In the next chapter, however, two strategies will be presented which are superior to the heuristics employed here. Therefore, the main purpose of this experimental section is to emphasize the importance of the practical aspects mentioned above, and to evaluate the behavior of the EM-algorithm close to the global optimum of the likelihood function.

In all the experiments the artificially blurred cameraman image will be used in order to be able to check the identified parameter values against their true values, and to compute the improvement in SNR using the ideal and estimated parameters. The NSHP image model coefficients computed from the original image are given by:

$$a(k,l) = \begin{bmatrix} a(1,1) & a(1,0) & a(1,-1) \\ a(0,1) & & \end{bmatrix} = \begin{bmatrix} -0.372 & 0.644 & 0.143 \\ 0.578 & & \end{bmatrix},$$

and the variance of the modeling error $v(i,j)$ is equal to $\sigma_v^2 = 185.3$. It is pointed out, however, that image restoration results are not very sensitive to the actual image model (or regularization operator) used, as was discussed in Chapter 3. Hence, deviations between the actual and identified image model coefficients are of minor importance, at least when compared to the effects of deviations between the actual and estimated PSF coefficients.

In the experiments the noisy blurred image was preprocessed at its boundaries in order to (approximately) satisfy the circulant assumption. More specifically, at the borders of the blurred image a strip of 10 pixels was first discarded, and next replaced by values resulting from a linear interpolation of the intensity difference between facing borders. Although such a procedure replaces approximately 15 percent of the blurred data by artificial data, this does not significantly degrade the identification results.

7.5.1 Linear Motion Blur

The cameraman image was artificially blurred by horizontal linear motion blur over 8 pixels. The discrete PSF associated with this blur is computed via (2.24), and has a support of size 1×9. The PSF coefficients are given by:

$$\begin{aligned} d(m,n) &= [\cdots, d(0,0), d(0,1), d(0,2), d(0,3), d(0,4)] \\ &= [\cdots, 0.1250, 0.1250, 0.1250, 0.1250, 0.0625]. \end{aligned}$$

(a) (b)

Figure 7.3: (a) Motion blurred image with an SNR of 40 dB; (b) Restoration using the ideal parameters.

(Since the PSF is symmetric, we listed only $d(0,0)$ to $d(0,4)$.) Noise was added with SNRs of 60 dB ($\sigma_w^2 = 0.004$), 40 dB ($\sigma_w^2 = 0.35$), and 20 dB ($\sigma_w^2 = 35.0$). Figure 7.3a shows the noisy blurred image with an SNR of 40 dB. For the purpose of comparison, the Tikhonov-Miller restoration result using the ideal parameters is shown in Figure 7.3b. The SNR improvement of this restored image is 7.7 dB, which constitutes an upper bound to the improvements in SNR when using estimated parameters.

In identifying the image and blur model we assume that $a(k,l)$ has a first order NSHP support, and that $d(m,n)$ is symmetric. Further, the correct size of the PSF support, S_d, is used (i.e. 1×9). For this reason the number of PSF coefficients to be estimated is known *a priori*. In general, however, this size must be determined prior to or simultaneously with estimating the PSF coefficients.

Selection of Initial Parameters

In order to identify the noisy blurred image using the EM-algorithm, we need to initialize the various parameters in $\hat{\theta}^{(0)}$ (see Figure 7.2). Since the EM-algorithm is guaranteed to converge only to a stationary point of the likelihood function, which may not be the global optimum that we are

actually looking for, the identification process is generally sensitive to the choice for $\hat{\theta}^{(0)}$. Furthermore, a bad set of initial parameters may needlessly increase the number of iterations required to achieve convergence.

If we consider the parameters to be initialized, we observe that:

(i) an initial estimate for the noise variance can be obtained from a smooth portion of the image (e.g. from the background),

(ii) an initial image model can be derived from the concept of regularization, as described in Chapter 3, or can be obtained from a prototype image. Throughout this section, we use the following (robust) initialization for the image model coefficients:

$$\hat{a}(k,l) = \begin{bmatrix} -0.300 & 0.500 & 0.100 \\ 0.700 & & \end{bmatrix},$$

and $\sigma_v^2 = 200.0$,

(iii) for some types of blur the initial PSF coefficients can be obtained by using the spectral or cepstral methods of Section 6.1, but in general this is not possible. In other words, the initialization of the PSF is most difficult, and requires special attention.

Motivated by the above observations, we select an initial guess for the PSF coefficients by evaluating the likelihood function $\mathcal{L}(\theta)$ for a number of relevant PSFs, using the above fixed image model coefficients and a fixed noise variance. Table 7.1 lists the value of the normalized likelihood function (i.e. $\mathcal{L}(\theta)$ divided by MN) for the blurred image with an SNR of 40 dB, for a number of relevant PSFs. The estimated noise variance is $\hat{\sigma}_w^2 = 1.0$. Since the PSF is assumed to be symmetric, only the unique PSF coefficients plus $\hat{d}(0,0)$ are listed. From this table we can select the initial PSF as the one with the minimal value of the likelihood function.

For the purpose of comparison, we consider three different initializations of the PSF. The identified parameters which are obtained by initializing the EM-algorithm in this way are listed in the Tables 7.2 and 7.3:

(i) If the initial PSF is set to "no blur" (i.e. $\hat{d}(m,n) = \delta(m,n)$), the algorithm converges to a local optimum. The restoration result associated with these parameters is shown in Figure 7.4a.

Table 7.1: Likelihood function as a function of the initial PSF coefficients (fixed image model and noise variance).

$\hat{d}(0,0)$	$\hat{d}(0,1)$	$\hat{d}(0,2)$	$\hat{d}(0,3)$	$\hat{d}(0,4)$	$\mathcal{L}(\hat{\theta})$
1.000	0.000	0.000	0.000	0.000	3.62
0.750	0.125	0.000	0.000	0.000	3.30
0.500	0.250	0.000	0.000	0.000	2.85
0.334	0.333	0.000	0.000	0.000	2.81
0.400	0.200	0.100	0.000	0.000	2.63
0.200	0.200	0.200	0.000	0.000	2.74
0.150	0.150	0.150	0.125	0.000	2.58
0.112	0.111	0.111	0.111	0.111	2.61

(ii) From Table 7.1, we select the PSF which yields the minimal value of the likelihood function. The identification results based on this initialization are very good. The restoration using the identified parameters is shown in Figure 7.4b. The SNR improvement of this image is 7.4 dB, compared with 7.7 dB for the image restored using the ideal parameters (shown in Figure 7.3b).

(iii) Finally, we initialized all PSF coefficients as $\frac{1}{9}$, i.e. the initial PSF assumes linear motion blur over 9 pixels. Although the identified parameters are again close to the optimum values, the quality of the associated restored image is less than obtained in case (ii). This restored image is shown in Figure 7.4c, and has an SNR improvement of 3.9 dB.

From the above experiment it is seen that the PSF needs to be initialized relatively close to the global optimum of the likelihood function in order to prevent convergence to an erroneous stationary point. The solutions obtained in the cases (ii) and (iii) must be considered to be both at the global optimum of the likelihood function, because the computed value of the likelihood function (using DFTs) is accurate only up to ± 0.01. Apparently, even if the initial parameters are close to the global optimum the identified solution is not unique, because the solutions in the cases (ii) and (iii) differ slightly. The reason for this is that near the global optimum, the likelihood function is relatively insensitive to variations in the various parameters. Therefore small numerical inaccuracies, which are unavoid-

(a)

(b)

(c)

(d)

Figure 7.4: Restoration results from Figure 7.3a: (a)-(c) Parameters estimated by the EM-algorithm; (d) Parameters estimated by a least-squares method.

Table 7.2: Identified PSF and σ_w^2 for 3 different initializations of the PSF.

	$\hat{d}(0,0)$	$\hat{d}(0,1)$	$\hat{d}(0,2)$	$\hat{d}(0,3)$	$\hat{d}(0,4)$	$\hat{\sigma}_w^2$	$\mathcal{L}(\hat{\theta})$
Case (i)							
initial	1.000	0.000	0.000	0.000	0.000	1.0	3.62
final	0.308	0.196	0.042	0.054	0.053	0.34	2.11
Case (ii)							
initial	0.150	0.150	0.150	0.125	0.000	1.0	2.58
final	0.128	0.126	0.125	0.125	0.060	0.32	2.00
Case (iii)							
initial	0.112	0.111	0.111	0.111	0.111	1.0	2.61
final	0.116	0.118	0.125	0.134	0.066	0.34	2.01

Table 7.3: Identified image model coefficients and σ_v^2 for 3 different initializations of the PSF.

	$\hat{a}(1,1)$	$\hat{a}(1,0)$	$\hat{a}(1,-1)$	$\hat{a}(0,1)$	$\hat{\sigma}_v^2$
Case (i)					
initial	-0.300	0.500	0.100	0.700	200.0
final	-0.653	0.647	0.133	0.871	36.1
Case (ii)					
initial	-0.300	0.500	0.100	0.700	200.0
final	-0.369	0.633	0.156	0.573	188.6
Case (iii)					
initial	-0.300	0.500	0.100	0.700	200.0
final	0.456	0.642	0.155	0.655	149.8

able, prohibit convergence to the exact global optimum. The errors due to the interaction between the numerical inaccuracies and the insensitivity of the likelihood function depend on the direction in which the optimum is approached. As a consequence the initialization of the EM-algorithm will be of influence on the identified parameter values and hence the quality of the restored image. In general the numerical inaccuracies become of more importance when the number of unknown parameters increases. For this reason we will study methods to identify PSFs with many coefficients (i.e. with large support sizes) in more detail in Chapter 8.

Bias and Estimation Variance

In the preceding experiment we have shown that it is possible to select $\hat{\theta}^{(0)}$ by a (heuristic) search mechanism for the PSF. The results may, however, be biased due to numerical inaccuracies, where the bias depends on $\hat{\theta}^{(0)}$. Besides these numerical biases, however, the ML solution may also be biased as a result of using a finite number of data points, and the solutions obtained may vary depending on the actual noise realization. Therefore, in this section we consider the effects of noise on the identified parameters. Since such an evaluation depends strongly on the image and PSF under consideration, the results presented here do not provide absolute answers but are only indicative, and must therefore be interpreted with caution.

The noisy motion blurred cameraman image was identified for 10 different noise realizations at the SNRs of 20, 40 and 60 dB. The image model was initialized as before, and the initial noise variance was set to 100.0 in all cases. The initial PSF coefficients were chosen as follows:

$$\hat{d}(0,n) = [\cdots, 0.150, 0.150, 0.150, 0.125, 0.000].$$

For each parameter we computed (i) the averaged identified value, (ii) the standard deviation (in percent) of the 10 different identification results, and (iii) the (absolute value of the) bias (in percent) relative to the actual parameter value. The results are listed in the Tables 7.4 and 7.5. Figure 7.5 shows the transfer function of the ideal PSF and the identified PSFs at the SNRs of 20, 40 and 60 dB.

From the results obtained by the above simulations we derive three conclusions:

Table 7.4: Averaged value, standard deviation and bias of the identified PSF and σ_w^2 for 3 different SNRs.

	$\hat{d}(0,0)$	$\hat{d}(0,1)$	$\hat{d}(0,2)$	$\hat{d}(0,3)$	$\hat{d}(0,4)$	$\hat{\sigma}_w^2$
	SNR=20 dB					
averaged	0.131	0.119	0.116	0.170	0.029	30.1
stan.dev.	8%	11%	19%	22%	23%	1%
bias	5%	5%	7%	36%	53%	14%
	SNR=40 dB					
averaged	0.126	0.125	0.124	0.128	0.060	0.31
stan.dev.	1%	$< 1\%$	$< 1\%$	1%	$< 1\%$	2%
bias	1%	$< 1\%$	1%	2%	4%	10%
	SNR=60 dB					
averaged	0.134	0.131	0.126	0.118	0.058	0.006
stan.dev.	$< 1\%$	$< 1\%$	$< 1\%$	$< 1\%$	$< 1\%$	45%
bias	7%	5%	1%	6%	7%	54%

Table 7.5: Averaged value, standard deviation and bias of the identified image model for 3 different SNRs.

	$\hat{a}(1,1)$	$\hat{a}(1,0)$	$\hat{a}(1,-1)$	$\hat{a}(0,1)$	$\hat{\sigma}_v^2$
	SNR=20 dB				
averaged	-0.570	0.688	0.122	0.755	95.4
stan.dev.	2%	2%	5%	1%	2%
bias	50%	7%	15%	30%	48%
	SNR=40 dB				
averaged	-0.400	0.643	0.151	0.650	171.7
stan.dev.	1%	$< 1\%$	1%	$< 1\%$	1%
bias	8%	$< 1\%$	6%	4%	8%
	SNR=60 dB				
averaged	-0.302	0.669	0.142	0.482	268.5
stan.dev.	6%	1%	1%	3%	10%
bias	18%	3%	1%	16%	45%

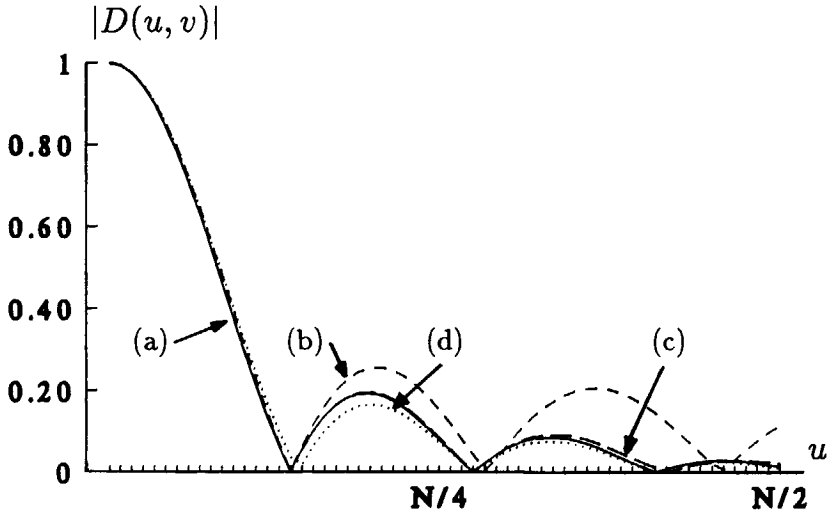


Figure 7.5: Transfer function of (a) Motion blur over 8 pixels, (b) Identified motion blur at 20 dB, (c) 40 dB, and (d) 60 dB.

(i) The presence of observation noise mainly affects the high frequency components of the identified PSF (see Figure 7.5),

(ii) Most of the estimated parameters do not change very much if different noise realizations are used, as can be seen from the small values of the standard deviation. For decreasing SNRs, however, the variation in the identified values of the PSF coefficients becomes significant.

(iii) If the SNR is increased from 20 to 40 dB, the bias of most identified parameters decreases significantly. The remaining bias at an SNR of 40 dB is negligible for practical purposes, hence we can conclude that at this SNR there are data points enough to correctly identify the parameters. However, if the SNR increases to 60 dB, the biases start to increase again. The reason for this is that as $\sigma_w^2 \to 0$, the conditional likelihood function $\mathcal{L}(\theta; \hat{\theta}^{(k)})$ becomes discontinuous in the parameters to be identified. In fact, the PDF of $\boldsymbol{g}$ given $\boldsymbol{f}$ and σ_w^2 (Eq. (6.10)) degenerates in this case. This will prevent the convergence of the EM-algorithm to the global optimum.

In conclusion, the above experiment shows that the EM-algorithm is particularly suitable for identifying noisy blurred images at medium to low

Table 7.6: Identified PSF using least-squares.

SNR	$\hat{d}(0,0)$	$\hat{d}(0,1)$	$\hat{d}(0,2)$	$\hat{d}(0,3)$	$\hat{d}(0,4)$
20	0.081	0.150	0.072	0.189	0.048
40	0.124	0.132	0.119	0.129	0.058
60	0.128	0.126	0.124	0.123	0.061

Table 7.7: Identified image model coefficients using least-squares.

SNR	$\hat{a}(1,1)$	$\hat{a}(1,0)$	$\hat{a}(1,-1)$	$\hat{a}(0,1)$	$\hat{\sigma}_v^2$
20	-0.115	0.430	0.269	0.395	311.1
40	-0.246	0.553	0.325	0.363	238.5
60	-0.317	0.552	0.194	0.570	225.2

SNRs. For low SNRs, however, the ML estimates of the PSF coefficients might be biased significantly. In general it is observed that the biases become larger as the number of unknown coefficients increases. Hence, in order to identify a large PSF at a low SNR, alternative identification procedures are required, which will be discussed in Chapter 8.

Comparison with the Least-Squares Method

In Section 6.3 two identification algorithms were described which are derived under the assumption that the observed image is noiseless. For the purpose of comparison, the Tables 7.6 and 7.7 list the identified parameters using the least-squares method of Tekalp *et al.* (see Section 6.3.1). The results are taken from reference [59]. Note that the least-squares method does not compute an estimate of the noise variance, because it assumes that the blurred image is noise free. As expected, for high SNRs the least-squares method performs well and the results are in fact better than the ones obtained by the EM-algorithm. For low SNRs the least-squares identification method does not provide satisfactory results. This shows that the least-squares method and EM-algorithm are essentially complementary methods for image identification. For high SNRs the least-squares method is most attractive, while for medium to low SNRs preference is given to the EM-algorithm. The restoration result using the parameters identified at 40 dB is shown in Figure 7.4d.

7.5.2 Defocusing Blur

The second experiment deals with a defocusing blur. The scope of this experiment is somewhat more restricted, since many of the conclusions derived from the identification of 1-D PSFs in the previous experiment carry over to 2-D PSFs. In the previous experiment we assumed that the extent of the PSF, i.e. S_d was known prior to identifying the PSF coefficients $d(m,n)$. In that case the number of coefficients to be estimated is exactly known. In general, however, the size of the PSF must be determined simultaneously with the estimation of the PSF coefficients. In this section we show that it is possible to determine the correct PSF support size by first identifying the PSF coefficients for a number of support sizes, and next selecting the best one on the basis of the value of the likelihood function.

The cameraman image was artificially blurred by defocusing blur using the following 2-D PSF:

$$d(m,n) = \begin{bmatrix} 0.018 & 0.031 & 0.031 & 0.031 & 0.018 \\ 0.031 & 0.062 & 0.062 & 0.062 & 0.031 \\ 0.031 & 0.062 & 0.062 & 0.062 & 0.031 \\ 0.031 & 0.062 & 0.062 & 0.062 & 0.031 \\ 0.018 & 0.031 & 0.031 & 0.031 & 0.018 \end{bmatrix}.$$

The blurred image with an SNR of 40 dB is shown in Figure 7.6a. This image was identified by running the EM-algorithm 3 times using three different PSF support sizes, namely 3×3, 5×5, and 7×7. In this controlled experiment the initial conditions were chosen in such a way that the EM-algorithm converged to the global optimum of the likelihood function. The results obtained in each of these cases are given next.

- Identification results when S_d is assumed to have the size 3×3:

$$\begin{aligned} \hat{a}(k,l) &= \begin{bmatrix} -0.682 & 0.680 & 0.142 \\ 0.861 & & \end{bmatrix}, \\ \hat{d}(m,n) &= \begin{bmatrix} 0.056 & 0.159 & 0.042 \\ 0.136 & 0.214 & 0.136 \\ 0.042 & 0.159 & 0.056 \end{bmatrix}, \\ \hat{\sigma}_v^2 &= 22.2, \\ \hat{\sigma}_w^2 &= 0.30, \\ \mathcal{L}(\hat{\theta}) &= 1.78. \end{aligned}$$

(a)

(b)

(c)

(d)

Figure 7.6: (a) Defocused image; (b)-(d) Restoration using identified parameters: (b) 3×3 PSF support, (c) 5×5 PSF support, (d) 7×7 PSF support.

- Identification results when S_d is assumed to have the size 5×5:

$$\hat{a}(k,l) = \begin{bmatrix} -0.406 & 0.594 & 0.170 \\ 0.633 & & \end{bmatrix},$$

$$\hat{d}(m,n) = \begin{bmatrix} 0.019 & 0.031 & 0.032 & 0.029 & 0.018 \\ 0.033 & 0.060 & 0.060 & 0.061 & 0.030 \\ 0.030 & 0.065 & 0.064 & 0.065 & 0.030 \\ 0.030 & 0.061 & 0.060 & 0.060 & 0.033 \\ 0.018 & 0.029 & 0.032 & 0.031 & 0.019 \end{bmatrix},$$

$$\hat{\sigma}_v^2 = 203.4,$$
$$\hat{\sigma}_w^2 = 0.32,$$
$$\mathcal{L}(\hat{\theta}) = 1.67.$$

- Identification results when S_d is assumed to have the size 7×7:

$$\hat{a}(k,l) = \begin{bmatrix} -0.450 & 0.574 & 0.141 \\ 0.725 & & \end{bmatrix},$$

$$\hat{d}(m,n) = \begin{bmatrix} 0.002 & 0.000 & 0.004 & 0.006 & 0.002 & 0.007 & 0.001 \\ -0.003 & 0.018 & 0.030 & 0.031 & 0.032 & 0.017 & 0.001 \\ -0.002 & 0.029 & 0.058 & 0.059 & 0.059 & 0.030 & 0.000 \\ -0.003 & 0.030 & 0.062 & 0.063 & 0.062 & 0.030 & -0.003 \\ 0.000 & 0.030 & 0.059 & 0.059 & 0.058 & 0.029 & -0.002 \\ 0.001 & 0.017 & 0.032 & 0.031 & 0.030 & 0.018 & -0.003 \\ 0.001 & 0.007 & 0.002 & 0.006 & 0.004 & 0.000 & 0.002 \end{bmatrix}$$

$$\hat{\sigma}_v^2 = 226.2,$$
$$\hat{\sigma}_w^2 = 0.32,$$
$$\mathcal{L}(\hat{\theta}) = 1.66.$$

The restoration results associated with the above identification results are shown in Figure 7.6b through d. As we change the size of the PSF support from 3×3 to 5×5, the value of the likelihood function decreases significantly. However, making the size even larger does not yield a significant decrease of $\mathcal{L}(\hat{\theta})$. Adding to this that the coefficients at the borders of the 7×7 PSF are relatively small, we conclude that the optimal PSF support size is 5×5.

The visual quality of the restoration results using the identified parameters can also be used to determine the correct PSF support size. From

Figure 7.6b it is seen that the PSF of size 3×3 does not restore the image enough, while the restoration results using the 5×5 and 7×7 PSF are visually identical. In conclusion, we have illustrated that both the numerical value of the likelihood function, as well as heuristic arguments can be employed to determine the correct size of the PSF support. In the next chapter more elegant ways of estimating this size will be discussed.

Chapter 8

Methods for Improved Image Identification

In the previous chapter an efficient iterative image identification algorithm has been introduced which is based on the EM-algorithm. In the practical use of this identification method a number of restrictions were encountered.

In the first place, the likelihood function $\mathcal{L}(\theta)$ becomes insensitive to variations in θ if the PSF contains more than just a few free coefficients, or – in other words – the extremum in $\mathcal{L}(\theta)$ becomes less pronounced. Since the EM-algorithm is basically a hill-climbing algorithm, inaccuracies in solving the E-step and M-step will prohibit convergence to the exact ML estimator. It is worthwhile noticing that similar problems will occur in any maximum likelihood based image identification algorithm.

Secondly, in the development of the EM-algorithm, the extent of the supports S_a and S_d have been assumed to be known *a priori*. This is, however, hardly ever true for the support of the PSF. As was experimentally shown, it is possible to use a number of differently sized supports in the identification process, and decide upon the one which leads to the optimal likelihood value. On the other hand, it is also suggested to overestimate S_d initially, and to remove small coefficients from the PSF after a preliminary identification, thus reducing S_d. In both strategies, however, the number of parameters can grow very large, which leads to the accuracy problems mentioned previously.

In the third place, the EM-algorithm may converge only to a local optimum of $\mathcal{L}(\theta)$. Therefore, the identification algorithm needs to be initialized with reasonable initial parameters $\hat{\theta}^{(0)}$. For the image model and noise variance this can be done quite reliably. However, choosing an ini-

tial PSF is very critical and therefore not a trivial problem. A heuristic search, which evaluates the likelihood function for a number of different PSFs, may be employed. However, for an increasing number of PSF coefficients, the number of suboptimal solutions, and hence the number of initial guesses $\hat{\theta}^{(0)}$ that need to be evaluated, may grow unacceptably large.

In order to circumvent the above restrictions improved strategies are required which are specifically designed to allow for an easy initialization of the PSF, and which avoid local optima as much as possible. In this chapter we will describe two such methods. In the first method (Sections 8.1 and 8.2) the number of parameters to be estimated is reduced significantly by assuming *a priori* that the PSF and image model can be described by a parametric model indexed by a restricted number of parameters. In the second approach (Sections 8.3 and 8.4) resolution pyramids are employed to estimate the coefficients and support size of the PSF in a hierarchical manner.

The use of parametric models and/or hierarchical identification procedures are essentially independent of the way the optimization of the likelihood function has been implemented. Here we will focus in particular on the EM-algorithm based image identification method.

This chapter is concluded with an evaluation of the status quo in image identification, and a look into possible future developments.

8.1 Parametric Image Identification

8.1.1 Parametric Modeling

In many practical situations of interest, there is more information available about the image model and PSF than has been used in the ML image identification methods described so far. More specifically, a certain image model and PSF structure can often be assumed, simply because this information in particular is the key to concluding that an image is blurred indeed.

One way to capture structural knowledge is by the use of parametric models, which describe the image model and PSF coefficients as a function of one or two parameters. Instead of estimating the values of the individual coefficients, the parameters of the selected parametric models are estimated. If appropriate models are used of low order, the dimension of

the solution space (i.e. the number of unknowns) is reduced significantly, and is independent of the support size S_d as well.

In the following we describe how parametric image and blur models can be incorporated directly into the EM-algorithm [57]. The E-step of the algorithm is not influenced by using these parametric models; the modified M-step will be discussed for the parametric models suggested.

8.1.2 Image Model

The structure of an original image is often such that its autocorrelation function resembles a 2-D separable exponentially decaying function [40]:

$$r_{ff}(p,q) = \sigma_f^2 \rho_v^{|p|} \rho_h^{|q|}. \tag{8.1}$$

Here ρ_v and ρ_h denote the vertical and horizontal correlation coefficients, respectively. Since in image restoration image models are merely used to regularize the inversion of an ill-conditioned PSF, it is not necessary to have the exact MMSE coefficients $a(k,l)$ as obtained by (7.20). In order to reduce the number of unknown parameters and to enforce an image model structure prior to the identification process, we assume that the image model coefficients fit the autocorrelation function (8.1) exactly. For an image model with quarter plane support, this requirement leads to the following parametric description [40]:

$$\begin{aligned} a(1,0) &= \rho_v, \quad a(0,1) = \rho_h, \text{ and } a(1,1) = -\rho_v \rho_h, \\ \sigma_v^2 &= \sigma_f^2 (1-\rho_v^2)(1-\rho_h^2). \end{aligned} \tag{8.2}$$

In the identification process the image model coefficients are forced to satisfy the above relations. Therefore, instead of identifying the coefficients $a(k,l)$, the horizontal and vertical correlation coefficients, ρ_h and ρ_v, are identified. Substitution of (8.2) into (7.19) yields after some straightforward manipulations:

$$\begin{aligned} J(\rho_v,\rho_h) = &-\hat{r}_{ff}(0,0) + 2\rho_v \hat{r}_{ff}(1,0) + 2\rho_h \hat{r}_{ff}(0,1) - 2\rho_v\rho_h\{\hat{r}_{ff}(1,1) \\ &+ \hat{r}_{ff}(1,-1)\} - \rho_v^2 \hat{r}_{ff}(0,0) - \rho_h^2 \hat{r}_{ff}(0,0) + 2\rho_v^2 \rho_h \hat{r}_{ff}(0,1) + \\ &+ 2\rho_v \rho_h^2 \hat{r}_{ff}(1,0) - \rho_v^2 \rho_h^2 \hat{r}_{ff}(0,0). \end{aligned} \tag{8.3}$$

It is not possible to obtain explicit expressions for ρ_v and ρ_h which minimize (8.3). But because $J(\rho_v,\rho_h)$ is a quadratic expression in ρ_v for a fixed value

of ρ_h, and vice versa, the following partial gradients based iteration is an efficient way to optimize (8.3):

$$\hat{\rho}_v^{(k)} = \frac{\hat{r}_{ff}(1,0) - \hat{\rho}_h^{(k-1)}\{\hat{r}_{ff}(1,1) + \hat{r}_{ff}(1,-1)\} + \hat{\rho}_h^{(k-1)^2}\hat{r}_{ff}(1,0)}{\hat{r}_{ff}(0,0) - 2\hat{\rho}_h^{(k-1)}\hat{r}_{ff}(0,1) + \hat{\rho}_h^{(k-1)^2}\hat{r}_{ff}(0,0)},$$

$$\hat{\rho}_h^{(k)} = \frac{\hat{r}_{ff}(0,1) - \hat{\rho}_v^{(k)}\{\hat{r}_{ff}(1,1) + \hat{r}_{ff}(-1,1)\} + \hat{\rho}_v^{(k)^2}\hat{r}_{ff}(0,1)}{\hat{r}_{ff}(0,0) - 2\hat{\rho}_v^{(k)}\hat{r}_{ff}(1,0) + \hat{\rho}_v^{(k)^2}\hat{r}_{ff}(0,0)}. \quad (8.4)$$

Here $\hat{\rho}_v^{(k)}$ and $\hat{\rho}_h^{(k)}$ denote the optimized parameters after k iteration cycles. The iterations (8.4) must be run in each M-step, and converge usually within a few cycles. Observe that if $J(\rho_v, \rho_h)$ is optimized only approximately, the EM-algorithm turns into a Generalized EM-algorithm.

The number of image model parameters can be reduced even more by considering the case where $\rho = \rho_v = \rho_h$. After substituting this condition into (8.3), optimizing $J(\rho_v, \rho_h)$ becomes identical to solving a third order polynomial equation, which can be done analytically.

8.1.3 Blur Model

It can be argued that it is not very realistic to model a PSF as a set of independent coefficients $d(m,n)$, because only a restricted subset of all possible $d(m,n)$ combinations will be accepted as representing realistic PSFs. For instance, $d(m,n)$ should represent a low-pass filtering action, should be "smooth", and $d(m,n) \geq 0.0$. For these reasons it is appropriate to consider a class of continuous parametric functions $d(s,t;\psi)$ which model realistic point-spread functions, and are indexed by ψ. Here ψ denotes the vector consisting of only a few parameters characterizing the PSF.

Relevant PSFs which can be modeled by such low-order parametric functions, are, for example, linear motion and out-of-focus blur. A number of common parametric PSFs have been described in more detail in Section 2.3. The discrete PSF coefficients $d(m,n)$ are computed from the continuous PSF $d(s,t;\psi)$ through Eq. (2.24). It is pointed out that since $\mathcal{L}(\theta;\hat{\theta}^{(k)})$ has to be continuous in θ in order to guarantee convergence of the EM-algorithm, each $d(m,n)$ needs to depend continuously on ψ. This condition restricts the possible set of parametric functions $d(s,t;\psi)$, and asks for an accurate (numerical) evaluation of (2.24).

Once a parametric blur model has been selected, the PSF coefficients and

the size of the support S_d are immediately determined by the value of ψ. Therefore, in the M-step we now need to optimize (7.14) with respect to ψ. By first substituting the selected parametric blur model into (2.24), and next substituting (2.24) into $J(\boldsymbol{D}, \sigma_w^2)$ in (7.22), we arrive at a nonlinear optimization problem in ψ, to which explicit solutions can rarely be found.

In general, (7.22) must be solved by numerical methods, which can be done efficiently since ψ is of low dimensionality. We have employed a straightforward steepest descent iteration to minimize $J(\psi, \sigma_w^2)$, where the gradients were computed numerically. Convergence of these iterations is normally achieved within 5-10 iteration steps. Again, as in the case with the image model coefficients, exact minimization of $J(\psi, \sigma_w^2)$ is not required as long as ψ is chosen in such a way that the likelihood function $\mathcal{L}(\theta)$ is increased in each M-step (Generalized EM-algorithm).

It goes without saying that for practical images more than one parametric PSF can be used. On the basis of the value of the likelihood function the PSF model which fits the image data best can be selected. It is also possible to consider the above parametric identification procedure as a means of finding a reasonable initial PSF. After ψ has been estimated, the non-parametric EM-algorithm described in Chapter 7 can be initialized using the PSF coefficients determined by $\hat{\psi}$.

8.2 Experimental Results Using Parametric Models

8.2.1 Linear Motion Blur

As in Section 7.5.1 we consider the identification of the motion blurred cameraman image. The length of motion is 8 pixels, and noise was added with the SNRs of 60, 50, 40, 30, 20, 10, 3 dB. Instead of identifying all PSF coefficients independently, the parametric model (2.28) is used (with $\phi = 0$). As a result only the parameter L, which defines the length of motion, needs to be estimated. In order to illustrate the behavior of the likelihood function, we evaluated $\mathcal{L}(\theta)$ for a large number of L values using a fixed image model ($\rho = 0.85$, $\sigma_v^2 = 200.0$) and several fixed noise variances. The resulting likelihood functions are shown in Figure 8.1. Since in this case θ is of low dimensionality such a search strategy is computationally feasible and can be used to find a reasonable initial guess for ψ.

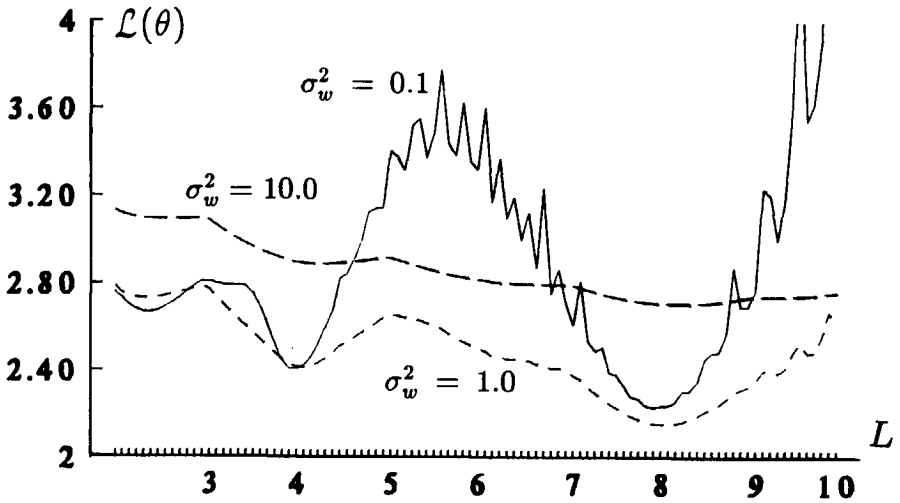
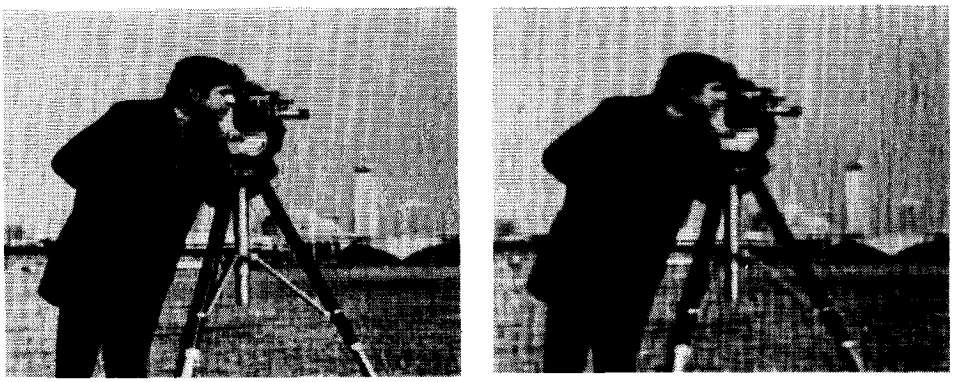


Figure 8.1: Likelihood function for linear motion blur over 8 pixels and SNR=40 dB.

Table 8.1: Identified parameters for motion blur with $L = 8.00$ at various SNRs.

SNR	$\hat{L}$	$\hat{\sigma}_w^2$	$\hat{\rho}_v$	$\hat{\rho}_h$	$\hat{\sigma}_v^2$	η_{SNR}	$\hat{\eta}_{\text{SNR}}$
3	10.41	1571.5	0.95	0.70	232.9	6.3	4.1
10	8.65	309.9	0.95	0.77	115.8	3.2	2.3
20	7.93	31.5	0.94	0.79	82.9	2.8	2.4
30	7.99	3.27	0.92	0.70	130.4	4.8	4.3
40	8.03	0.32	0.90	0.61	172.8	7.5	7.1
50	8.05	0.034	0.89	0.60	202.9	10.6	9.9
60	8.06	0.009	0.89	0.60	207.0	12.8	10.5

Table 8.1 lists the identification results for various SNRs using an initial guess of $\hat{L}^{(0)} = 6.0$. From the original cameraman we identified the optimal values for ρ_v and ρ_h as $\rho_v = 0.88, \rho_h = 0.60$, and $\sigma_v^2 = 201.2$. Table 8.1 also lists the signal-to-noise ratio improvement using the estimated parameters, denoted by $\hat{\eta}_{\text{SNR}}$, and using the ideal parameters (η_{SNR}). From this table



(a) (b)

Figure 8.2: Restoration of the motion blurred cameraman image (L=8) using identified parameters: (a) SNR=40 dB; (b) SNR=20 dB.

it becomes apparent that the parametric identification method is capable of obtaining good identification results up to an SNR of 20 dB. Figure 8.2a and 8.2b show the restored image using the identified parameters at an SNR of 40 dB and 20 dB, respectively.

8.2.2 Atmospheric Turbulence Blur

The cameraman image was blurred by turbulence blurs (Gaussian PSFs) with various standard deviations σ_G. The experiment included PSFs with support sizes ranging from 5×5 pixels to 13×13 pixels. This support size was determined by truncation of the PSF at coefficients smaller than approximately 0.1 percent of $d(0,0)$. Noise was added up to the SNR level of 30 dB ($\sigma_w^2 \approx 3.5$).

The parametric EM-identification algorithm was run on each of the cases tested with the following initial conditions: $\hat{\sigma}_G = 0.0$ (i.e. no blur), $\hat{\sigma}_w^2 = 10.0$, $\hat{\rho}_v = \hat{\rho}_h = 0.8$, and $\hat{\sigma}_v^2 = 200.0$. Table 8.2 lists the identification results. It is again observed that the SNR improvement of the restoration results using the identified parameters is in the same range as the SNR improvement of the restoration results using the true parameters. Figure 8.3 shows the blurred image with $\sigma_G = 1.00$, and the restored image with

Table 8.2: Identified parameters for turbulence blurs at SNR=30 dB.

σ_G	S_d	$\hat{\sigma}_G$	$\hat{\sigma}_w^2$	$\hat{\rho}_v$	$\hat{\rho}_h$	$\hat{\sigma}_v^2$	η_{SNR}	$\hat{\eta}_{\text{SNR}}$
0.500	5×5	0.530	4.45	0.86	0.61	257.0	5.9	5.6
0.750	7×7	0.774	3.47	0.85	0.68	249.1	4.1	3.9
1.000	9×9	0.978	3.24	0.87	0.71	203.0	2.8	2.6
1.500	11×11	1.475	3.15	0.89	0.75	185.7	2.1	2.0
2.000	13×13	1.988	3.02	0.92	0.74	181.9	2.0	1.9

(a) (b)

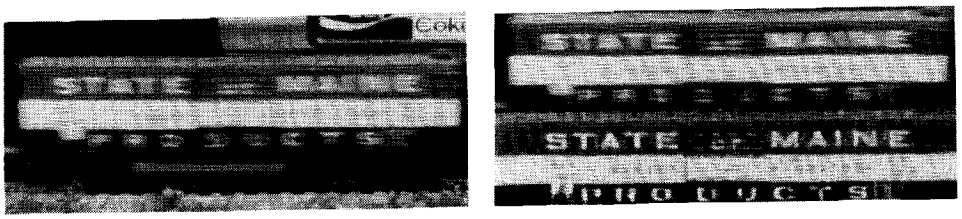
Figure 8.3: (a) Cameraman image blurred by atmospheric turbulence ($\sigma_G = 1.0$, SNR=30 dB); (b) Restoration using the identified parameters.

$\hat{\sigma}_G = 0.978$, respectively.

8.2.3 Photographic Motion Blur

The last experiment deals with the blurred train image in Figure 8.4a. The blur in this image is not artificial, but is due to a real photographic process[1]. Obviously the exposure time used to record the moving train is too large, which has led to a locally blurred image. Since the models used in this thesis assume that the entire image has been blurred in a similar

[1]Courtesy of Eastman Kodak Co., Rochester NY.



(a) (b)

Figure 8.4: (a) Blur introduced by real motion (b) Blurred section and restoration result using the identified parameters.

manner, we cannot immediately handle the situation at hand. Therefore we first cut out the blurred section of the image, and use only this part of the image in the identification and restoration process.

Since the blur in the image is obviously due to relative motion between the camera and the object of interest, a parametric model for horizontal linear motion blur was used in the EM-algorithm. The estimated length of motion is $\hat{L} = 8.31$ pixels, and $\hat{\sigma}_w^2 = 0.40$, $\hat{\rho}_v = 0.78$, $\hat{\rho}_h = 0.77$, and $\hat{\sigma}_v^2 = 59.2$. This result for L confirms the result of $\hat{L} = 8$ or $\hat{L} = 9$ pixels reported in Section 6.1. The restoration result of the blurred section of the train image using the identified parameters is shown in Figure 8.4b.

8.3 Hierarchical Image Identification

8.3.1 Use of Resolution Pyramids

Blur is a resolution dependent phenomenon, because the severity of the blur present in a discrete image depends on how the continuous (blurred) image was sampled. Images sampled at a very high resolution are more likely to be "blurred" than less densely sampled images. We can illustrate this by considering the discretization of the continuous PSF associated with linear motion blur over an absolute distance L_a (e.g. millimeters). If we assume that the sample distance is X_0 (millimeters), and that a suitable low-pass filter is used to prevent aliasing effects (see Section 8.3.2), an approximate

expression for the relative length of motion L_r is given by:

$$L_r = \frac{L_a}{X_0} \quad \text{(pixels)}. \tag{8.5}$$

For a decreasing sample distance X_0 the relative length of motion grows larger and larger. On the other hand, if $X_0 \geq L_a$, the relative length of motion is less than one, which means that there is no blur present in the image because the PSF falls within one pixel.

Reducing the resolution of a blurred image also affects the image identification problem. This can be illustrated by considering the likelihood functions of the cameraman image blurred by linear motion over 8 pixels (SNR=40 dB), which has been (sub-)sampled by different sample distances, namely $X_0 = 1, 2, 4$ and 8 pixels. For the purpose of simplicity we assume a fixed image model, a fixed value for σ_w^2, and use a parametric description of the (subsampled) PSFs. The resulting likelihood functions are shown in Figure 8.5. It is clear that for decreasing resolutions the optimum of $\mathcal{L}(\hat{\theta})$ shifts to smaller lengths of motion, and that at the resolution level 3 ($X_0 = 8$ pixels) the optimum appears for $L < 1$ (no blur). It is also observed that for decreasing resolutions the likelihood function becomes

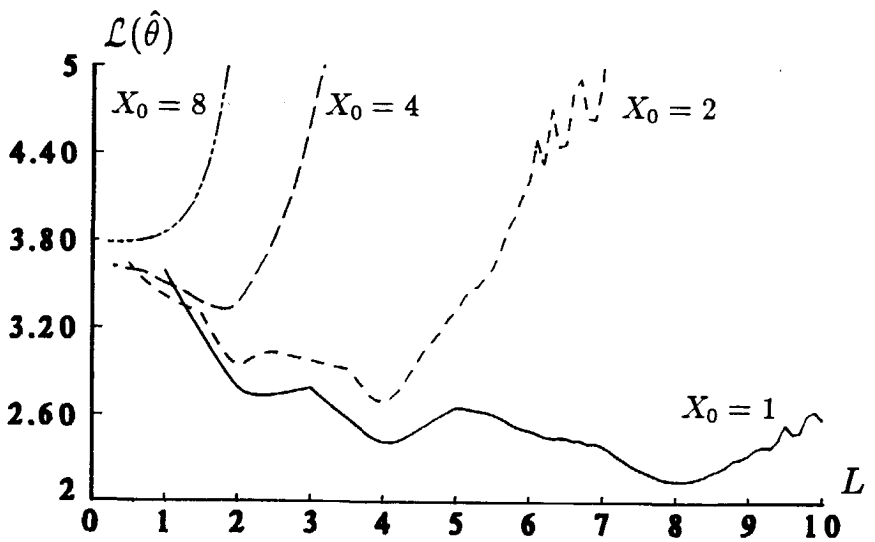


Figure 8.5: Likelihood functions for motion blur over 8 pixels at 4 different resolution levels (using a parametric PSF description).

"smoother", and contains fewer local optima.

The fact that for larger sample distances X_0 the severity of the blur decreases, implies that the support size S_d of the discrete PSF decreases as well. Hence, if we need to identify a discrete image blurred by a PSF with a large support size, the dimensionality of the problem can (initially) be reduced by considering the identification of a subsampled version of this image. In this way the effective PSF support size is reduced, and fewer PSF coefficients have to be identified. In addition to this, the likelihood function contains fewer local optima, and the identification process is computationally less expensive because it operates on less data.

Once the PSF and the image model of the subsampled image have been identified, these parameters and/or the restoration result using these parameters can be used to initialize the identification or restoration algorithm operating on the (full resolution) blurred image originally considered. Extending this procedure to multiple subsampled versions of the blurred image available (using a subsample factor of 2^p for the p-th subsampled image), we obtain a hierarchical identification method that "builds up" a possibly very large PSF by gradually increasing its resolution [58]. The identification results at the resolution level p in the resolution pyramid depend on the results at level $p+1$ via the initialization of the identification procedure.

In Figure 8.6 the principle of a hierarchical identification scheme is outlined in which the EM-algorithm is used. At first, the available blurred image is filtered and downsampled a number of times in order to generate the "blurred image resolution pyramid", denoted by $\{\boldsymbol{g}, \boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_P\}$. At a certain level P, the image $\boldsymbol{g}_P$, which is a downsampled version of $\boldsymbol{g}$ by a factor of 2^P, is identified using the EM-algorithm. The level P should be chosen in such a way that the global optimization of the likelihood function at that resolution level, denoted by $\mathcal{L}_P(\theta_P)$, can be done easily. Therefore the resolution of $\boldsymbol{g}$ needs to be reduced far enough to warrant that $d(m,n) \approx \delta(m,n)$, i.e. the PSF falls almost within a single pixel (no blur).

Associated with the identified parameters $\hat{\theta}_P$ at level P is the restoration result $\hat{\boldsymbol{f}}_P$ of $\boldsymbol{g}_P$. Either the identified parameters or the restored image can be used to initialize the EM-algorithm at the resolution level $P-1$, which operates on the downsampled version $\boldsymbol{g}_{P-1}$ of $\boldsymbol{g}$. By repeating the above procedure, we finally obtain the required parameters at the resolution level 0, i.e. for the blurred image originally considered.

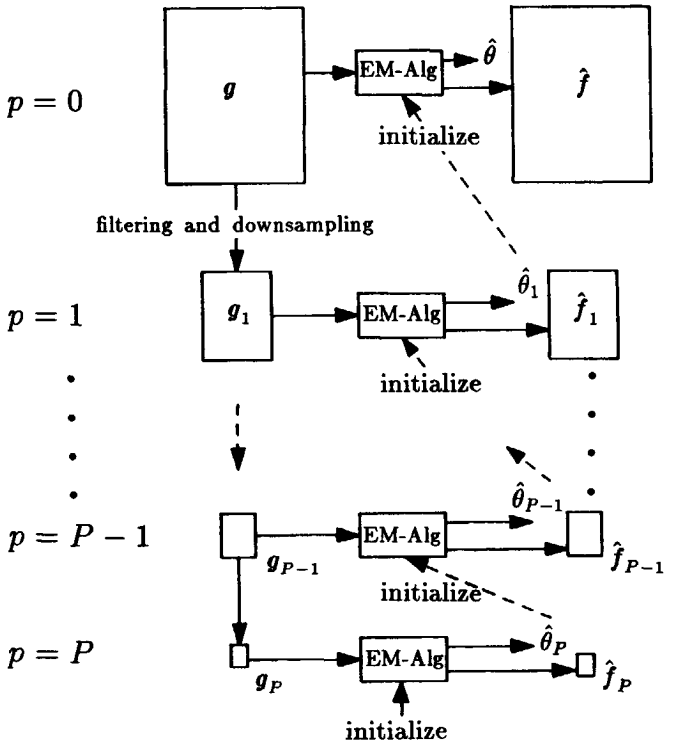


Figure 8.6: Identification of blurred images using a hierarchical structure.

Before applying the hierarchical identification method outlined above, we first have to pay more attention to the following three aspects:

- In order to prevent aliasing effects the blurred image available needs to be bandlimited prior to subsampling. However, the low-pass filter to be used must be chosen in such a way that the support size of the PSF after subsampling is guaranteed to be smaller than the one before subsampling,
- To initialize the identification process at level p using the results achieved at level $p+1$, either the identified PSF or the restored image needs to be interpolated. Which of the two options works best, and how should the interpolation be done?
- The most important effect of downsampling the blurred image is

that the PSF support size is reduced proportionally to the subsample factor. As a result we need to increase the support of the PSF as we go from level $p+1$ to level p in the hierarchical identification method. Hence, how can the size of the PSF support be established at each of the resolution levels?

The above non-trivial questions will be addressed in the next sections.

8.3.2 Downsampling of Blurred Images

The filtered and downsampled blurred images $\boldsymbol{g}_p$ are related to the full resolution image $\boldsymbol{g}$ as follows (c.f. Eq. (2.3)):

$$g_p(k,l) = \{l_p(i,j) * g(i,j)\}_{(i,j)=(2^p k, 2^p l)}, \tag{8.6}$$

where $l_p(i,j)$ is the low-pass filter which restricts the bandwidth of $g(i,j)$ before downsampling it in order to prevent aliasing effects. If we neglect the noise contribution $w(i,j)$, we can write

$$g_p(k,l) = \{l_p(i,j) * d(i,j) * f(i,j)\}_{(i,j)=(2^p k, 2^p l)}. \tag{8.7}$$

Associated with $\{\boldsymbol{g}, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_P\}$ are the downsampled original images $\{\boldsymbol{f}, \boldsymbol{f}_1, \cdots, \boldsymbol{f}_P\}$ and the downsampled PSFs $\{d(m,n), d_1(m,n), \cdots, d_P(m,n)\}$. We would like to model the blurred image at each resolution level as the convolution of a downsampled PSF with a downsampled original image (see Figure 8.7):

$$\begin{aligned} g_p(k,l) &= d_p(k,l) * f_p(k,l) \\ &= \{l_p(i,j) * d(i,j)\}_{(i,j)=(2^p k, 2^p l)} * \{l_p(i,j) * f(i,j)\}_{(i,j)=(2^p k, 2^p l)}. \end{aligned} \tag{8.8}$$

With the exception of the case where $l_p(i,j) * d(i,j) = d(i,j)$ and $l_p(i,j) * f(i,j) = f(i,j)$, we can conclude that (8.7) and (8.8) are equivalent only if $l_p(i,j)$ is an ideal low-pass filter. In this case the PSF $d_p(m,n)$, which is identified at resolution level p, is an exact downsampled version of the PSF $d(m,n)$ at level 0. The use of such an ideal filter $l_p(i,j)$ has two serious drawbacks. In the first place, the PSF coefficients may take negative values, which is in contradiction with the PSF modeling assumptions employed throughout this thesis. Secondly, the size of the support of the downsampled PSF $d_p(m,n)$ is not necessarily smaller than the support

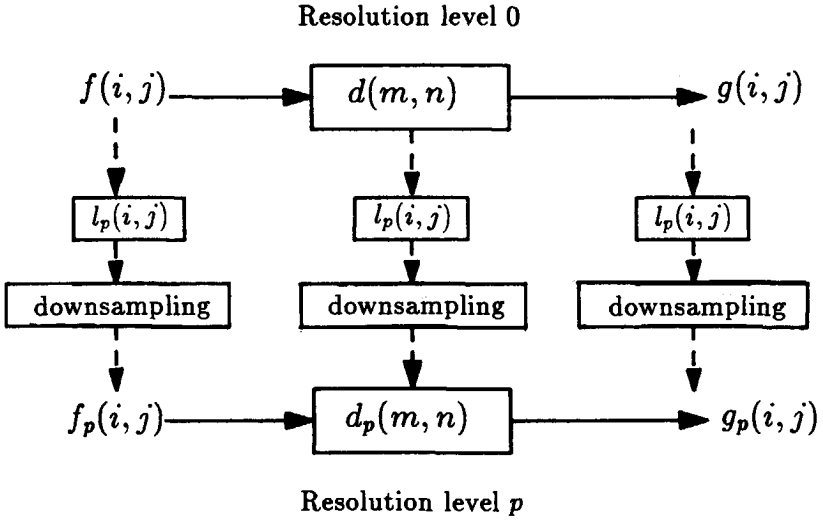


Figure 8.7: Theoretical model for downsampled blurred images.

size of the original PSF $d(m,n)$. Even more, $d_p(m,n)$ may have an infinite support. This leads to the conclusion that an ideal low-pass filter should not be used to bandlimit the blurred image.

To design a more suitable low-pass filter, we require that three conditions must be satisfied, namely

(i) If the image $\boldsymbol{g}$ is downsampled by a factor 2^p to get the reduced resolution image $\boldsymbol{g}_p$, the size of the PSF support S_d needs to be reduced by a factor 2^p as well. (The minimum support size, however, is 3, because it cannot be expected that a downsampled PSF will exactly fall within one pixel.) In general, the extent of the low-pass filter $l_p(i,j)$ must therefore be less than $2^{p+1} \times 2^{p+1}$,

(ii) In order for $d_p(m,n) \geq 0$, we need to have $g_p(i,j) \geq 0$. Hence, the filter coefficients $l_p(i,j)$ must be positive,

(iii) The low-pass filter must have a cut-off frequency of approximately $\omega_c = \pi 2^{-p}$.

A reasonable low-pass filter that satisfies the above constraints, is a separable truncated sinc-function, windowed by a Hamming function. In a 1-D

notation, the filter coefficients are given by:

$$l_p(k) = \frac{\sin(\pi k 2^{-p})}{\pi k 2^{-p}}\{0.54+0.46\cos(\pi k 2^{-p})\}, \quad -(2^p-1) \leq k \leq 2^p-1. \quad (8.9)$$

Table 8.3 lists the filter coefficients for $p = 1, 2, 3$. The coefficients given in this table are assumed throughout this chapter.

As a consequence from using a suboptimal low-pass filter, the PSF identified from $g_p(i,j)$ cannot be related to the PSF at higher resolution levels. The discrepancy between the optimally and suboptimally filtered PSF is mainly due to aliasing effects, and depends in general on the power spectral density of the image $\boldsymbol{g}$ in the "stopband" of the filter $l_p(i,j)$. Since for low resolution images the power in this stopband is relatively larger than for higher resolution images, the difference between the optimally and suboptimally filtered images increases as the resolution is decreased.

8.3.3 Image and Parameter Interpolation

The parameters identified at the resolution level $p+1$ are used to initialize the EM-algorithm at the resolution level p. Since the EM-algorithm can start with either an E-step or an M-step, we have the following two options:

- Initialize $\hat{\boldsymbol{\theta}}_p^{(0)}$ at level p using the image model, PSF and σ_w^2 identified at level $p+1$. To this end the identified PSF has to be represented at a higher resolution, i.e. it needs to be interpolated. The identified image model does not need to be interpolated because we assume a first order (NSHP) image model at all resolution levels. As in Chapter 7 the EM-algorithm is started with an E-step,

- Use the identified parameters at level $p+1$ to compute a restoration result $\hat{\boldsymbol{f}}_{p+1}$ from $\boldsymbol{g}_{p+1}$. Next, this result is interpolated in order to initialize $\hat{\boldsymbol{f}}_p^{(0)}$ at level p, from which the auto- and crosscorrelation

Table 8.3: Coefficients of the low-pass filter $l_p(k) = l_p(-k)$.

p	$l(0)$	$l(1)$	$l(2)$	$l(3)$	$l(4)$	$l(5)$	$l(6)$	$l(7)$
1	1.0000	0.3438						
2	1.0000	0.7790	0.3438	0.0644				
3	1.0000	0.9404	0.7790	0.5615	0.3438	0.1713	0.0644	0.0160

coefficients $\hat{r}_{ff}^{(0)}(p,q)$ and $\hat{r}_{fg}^{(0)}(p,q)$, respectively, can be computed. In this case the EM-algorithm starts essentially with an M-step.

In interpolating the PSF we can take advantage of the *a priori* knowledge that $d_p(m,n) \geq 0$, and that $d_p(m,n)$ has a finite support. This implies that the PSF cannot be interpolated using sinc functions. A similar reasoning holds for the interpolation of the restored image. However, the amount of data in an image is much larger than the number of PSF coefficients. Therefore the interpolation of the restored image is less critical and less sensitive to errors than the interpolation of the PSF. In fact, experimental evidence shows that a simple 0-th order interpolation of the restored image (i.e. interpolating the image by repeating pixels) leads to results which are superior to the results achieved by far more complex methods to interpolate the PSF, such as iterative interpolation procedures which enforce the PSF to be positive and of finite extent. For this reason, in our hierarchical identification approach the EM-algorithm is initialized at a certain resolution level by the interpolated restoration result obtained at the previous level using a simple pixel repeating method.

8.3.4 Decision Tree for PSF Support Size

Nearly all identification methods known in the literature require explicit knowledge about the size of the PSF support S_d. This size is, however, hardly ever known *a priori*, and needs to be estimated simultaneously with θ. One way of doing this has been addressed experimentally in Section 7.5.2, namely to identify the PSF for a number of different support sizes. The value of the likelihood function was used to discriminate between the various possibilities. Alternatively we can include a strategy for estimating the size of S_d into the hierarchical identification method described here.

Consider the identification of the class of 1-D horizontal PSFs which have a maximal support size of 1×17 (this is a reasonable value for images of size 256×256). If we use the low-pass filter described by (8.9), then the resulting support sizes of the downsampled PSFs $d_p(m,n)$ at the resolution level p are given by Table 8.4. Since we assume that the PSF is symmetric, only odd values for S_d appear in this table. Further, the minimal support size is 1×3, because a downsampled PSF will never fall entirely within one pixel. From this table it can be seen that at the resolution levels 3 and 4 we can always use a PSF of size 1×3. Further, a 1×3 PSF is

Table 8.4: Resulting PSF support size (S_d) after downsampling.

	Resolution level			
S_d	1	2	3	4
1×3	1×3	1×3	1×3	1×3
1×5	1×3	1×3	1×3	1×3
1×7	1×5	1×3	1×3	1×3
1×9	1×5	1×3	1×3	1×3
1×11	1×7	1×5	1×3	1×3
1×13	1×7	1×5	1×3	1×3
1×15	1×9	1×5	1×3	1×3
1×17	1×9	1×5	1×3	1×3

always associated with another 1×3 PSF or with a 1×5 PSF at the next resolution level. In general, a PSF of size $1 \times (2v+1)$ $(v = 2, 3, \cdots)$ at the resolution level $p+1$ leads to a support size of either $1 \times (4v-1)$ or $1 \times (4v+1)$ at the resolution level p.

By "inverting" Table 8.4 the decision tree in Figure 8.8 is obtained, which shows how the PSF support size can propagate as we perform the hierarchical identification. In this tree a solid line between two support sizes at subsequent resolution levels indicates that these supports are associated with each other according to Table 8.4. A dashed line, on the other hand, implies a (heuristic) correction possibility of the support size by leaving out the outermost PSF coefficients if they have a relatively small value (PSF truncation). Further, a solid line implies running the EM-algorithm, while for a dashed line we only need to evaluate the likelihood function for the truncated PSF.

Starting at level 4 we always identify a 1×3 PSF. At the resolution level 3 again a PSF with support 1×3 is identified. Next, at level 2 we have to choose between either a 1×3 or a 1×5 PSF. Therefore the EM-algorithm needs to be run for both these cases. We can decide upon the best support size by considering (i) the value of the likelihood function, (ii) the value of the outermost PSF coefficients, or any combination of these two measures. Here we base our decision on the value of the likelihood function. Observe that if the PSF support size has been chosen too large at a certain resolution level, there is always a correction possibility on the next level (dashed

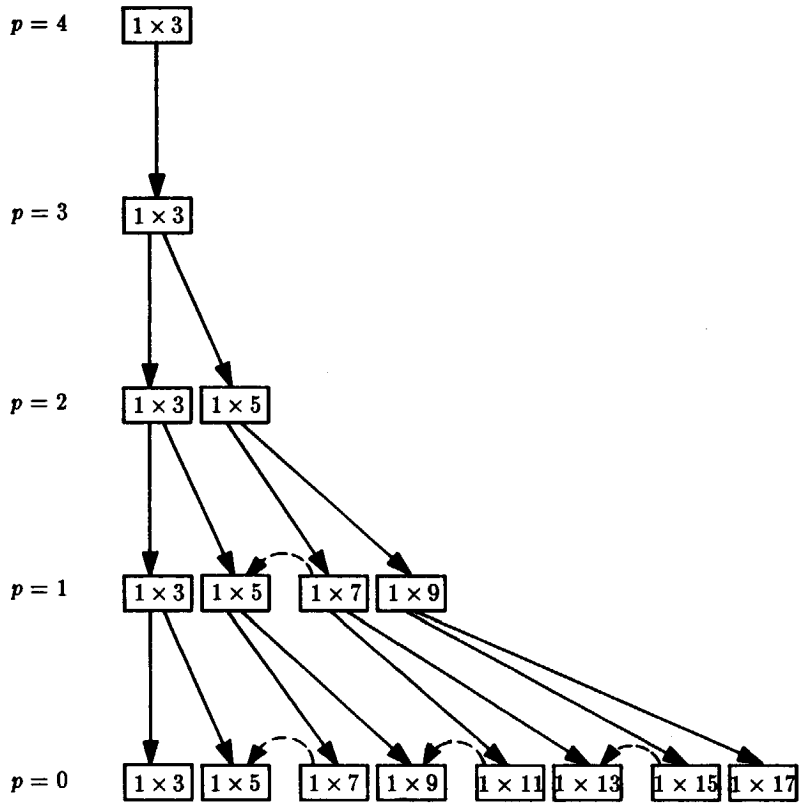


Figure 8.8: Decision tree for 1-D horizontal PSFs.

lines).

The above way of simultaneously identifying the PSF support and its coefficients requires running the EM-algorithm 8 times in total, with image sizes varying from 16×16 to 256×256. On the whole this takes significantly less time than running the EM-algorithm 8 times (1×3 to 1×17 PSFs) on the full resolution image of size 256×256.

The decision tree in Figure 8.8 can immediately be extended to 2-D PSFs, but then we need to consider 4 different PSF sizes at each resolution level.

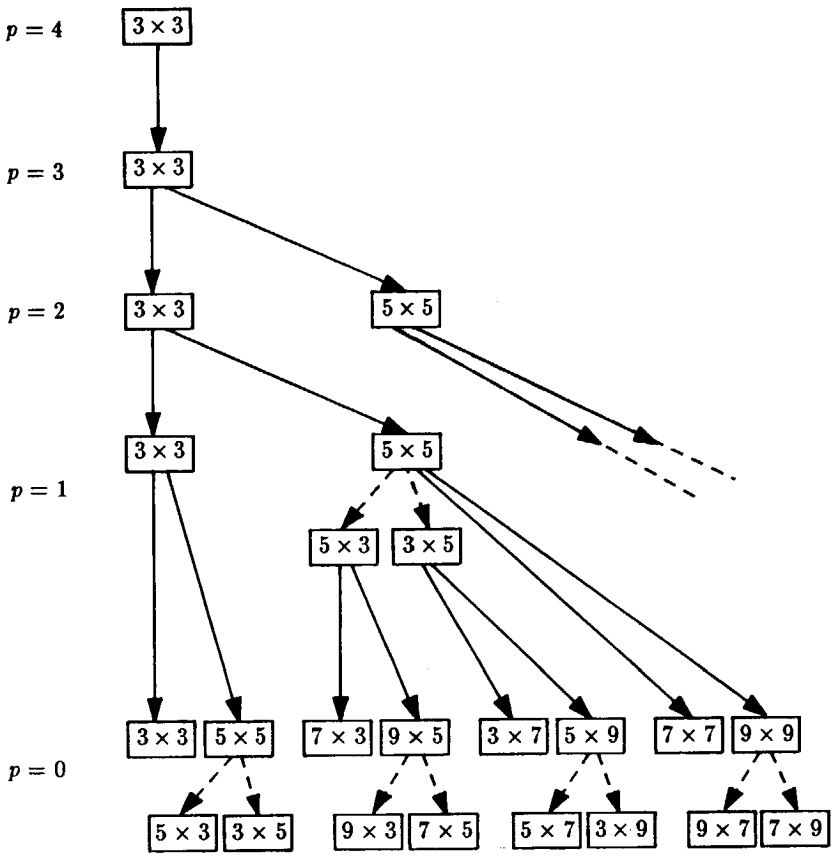


Figure 8.9: Decision tree for 2-D PSFs. (The branch leaving the 5×5 support size at level 2 is not shown in order to keep the figure readible.)

For instance, starting with a 5×5 PSF, we may get a PSF of size 7×7, 9×7, 7×9 or 9×9. In order to reduce the computational complexity, we only allow the increase of the vertical and horizontal extent by the same factor; hence we initially disregard the 7×9 and 9×7 PSFs. These support sizes can be obtained later on by truncating the 9×9 support if the outermost coefficients are relatively small. Figure 8.9 shows the resulting decision tree for 2-D blurs. Again, the solid lines imply running the EM-algorithm, while the dashed lines imply leaving out the outermost rows or

columns of the PSF when they contain small entries. Such a decision can again be based on the value of the likelihood function computed from the truncated PSF.

8.4 Experimental Results Using the Hierarchical Method

8.4.1 Linear Motion Blur

We consider the identification of the motion blurred image in Figure 7.3a using the hierarchical method described in the preceding section. The resolution pyramid generated from this image is shown in Figure 8.10a. Table 8.5 shows the identified PSF and observation noise variance at each level. Also shown in this table are the (initial) PSF and noise variance, computed by the M-step in the first iteration from the interpolated restoration result. In Table 8.6 the initial and identified image model coefficients are listed. Observe that for the levels 2, 1, and 0 we computed the PSF for 2 different PSF support sizes, which is in conformity with the decision tree in Figure 8.8. At the resolution level 2 the value of the likelihood function (see Table 8.5[2]) indicates that the 1×3 PSF is to be preferred to the 1×5 PSF. This conclusion can also be reached by considering the value of the coefficient $\hat{d}(0,2)$ at this level. At the resolution level 1 the likelihood function for the 1×5 PSF support is significantly smaller than for the 1×3 PSF. Hence, at the next level we need to consider a 1×7 and a 1×9 PSF support size. Finally, for the full resolution image, preference has to be given to the support of size 1×9.

Another way to illustrate the hierarchical identification method is to consider the transfer function $|D_p(u,v)|$ of the identified PSFs at the various resolution levels. In a combined plot Figure 8.11 shows the transfer function of the ideal PSF and the transfer function of the PSFs identified at the levels 3 through 0. It is observed that at low resolution levels the low frequency components of $d(m,n)$ are estimated, while the identification of higher resolution blurred images adds high frequency components to the PSF. The images obtained by restoring the downsampled blurred images using the identified parameters, are shown in Figure 8.10b. Table 8.7 gives

[2]It is recalled that the accuracy of the computed value of the likelihood function is approximately ± 0.01.

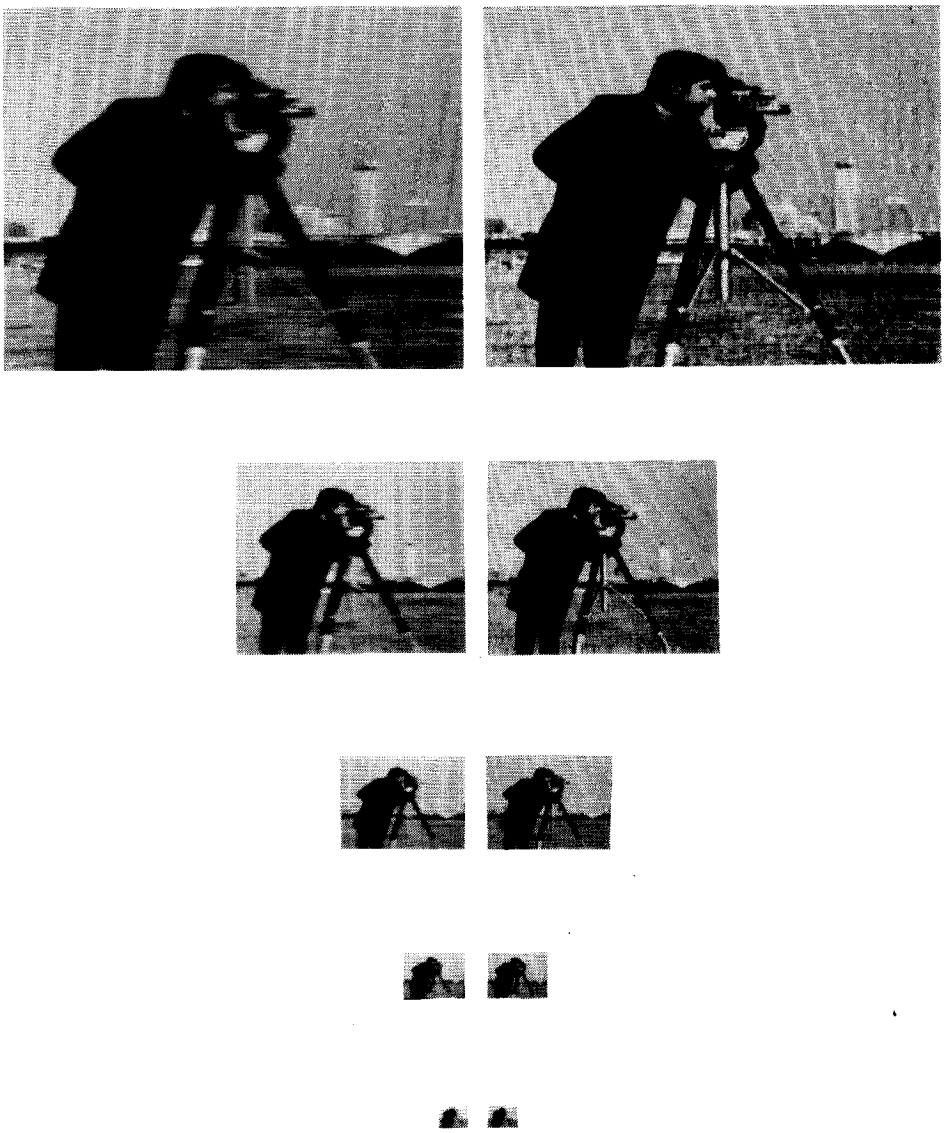
Table 8.5: Initial and identified PSF and observation noise variance at 5 resolution levels for horizontal motion blur at SNR=40 dB.

		Point-spread function						
	S_d	$\hat{d}(0,0)$	$\hat{d}(0,1)$	$\hat{d}(0,2)$	$\hat{d}(0,3)$	$\hat{d}(0,4)$	$\hat{\sigma}_w^2$	$\mathcal{L}_p(\hat{\theta}_p)$
Resolution level 4								
initial	1×3	1.000	0.000				10.0	
final		0.575	0.213				1.10	3.56
Resolution level 3								
initial	1×3	0.456	0.272				282.6	
final		0.581	0.210				1.53	3.69
Resolution level 2								
initial	1×3	0.397	0.301				189.2	
final		0.506	0.247				0.10	3.13
initial	1×5	0.400	0.288	0.012			189.2	
final		0.483	0.247	0.011			1.10	3.14
Resolution level 1								
initial	1×3	0.282	0.359				108.3	
final		0.554	0.223				0.36	2.73
initial	1×5	0.302	0.232	0.117			96.7	
final		0.242	0.257	0.122			0.11	2.53
Resolution level 0								
initial	1×7	0.098	0.143	0.129	0.179		62.6	
final		0.140	0.101	0.141	0.188		0.37	2.09
initial	1×9	0.120	0.124	0.138	0.108	0.069	59.2	
final		0.123	0.124	0.124	0.129	0.061	0.29	1.97

an impression of the computation time required for each of the stages in the hierarchical identification procedure.

Table 8.6: Initial and identified image model at 5 resolution levels for horizontal motion blur at SNR=40 dB.

		Image model				
	S_d	$\hat{a}(1,1)$	$\hat{a}(1,0)$	$\hat{a}(1,-1)$	$\hat{a}(0,1)$	$\hat{\sigma}_v^2$
Resolution level 4						
initial	1×3	-0.100	0.600	0.100	0.400	200.0
final		-0.486	0.678	0.159	0.608	296.5
Resolution level 3						
initial	1×3	-0.751	0.902	0.010	0.819	83.59
final		-0.328	0.710	0.140	0.449	369.9
Resolution level 2						
initial	1×3	-0.650	0.881	0.017	0.739	101.5
final		-0.341	0.731	0.103	0.490	291.2
initial	1×5	-0.650	0.881	0.017	0.739	101.5
final		-0.283	0.737	0.104	0.422	345.2
Resolution level 1						
initial	1×3	-0.673	0.886	0.013	0.767	80.5
final		-0.647	0.690	0.114	0.840	56.2
initial	1×5	-0.673	0.886	0.013	0.767	80.5
final		-0.305	0.696	0.133	0.464	269.0
Resolution level 0						
initial	1×7	-0.654	0.878	0.016	0.755	72.7
final		-0.623	0.617	0.158	0.847	50.0
initial	1×9	-0.654	0.878	0.016	0.755	72.7
final		-0.423	0.656	0.145	0.616	162.3



(a) (b)

Figure 8.10: Resolution pyramid of (a) Motion blurred images (L=8, SNR=40 dB), (b) Restored images.

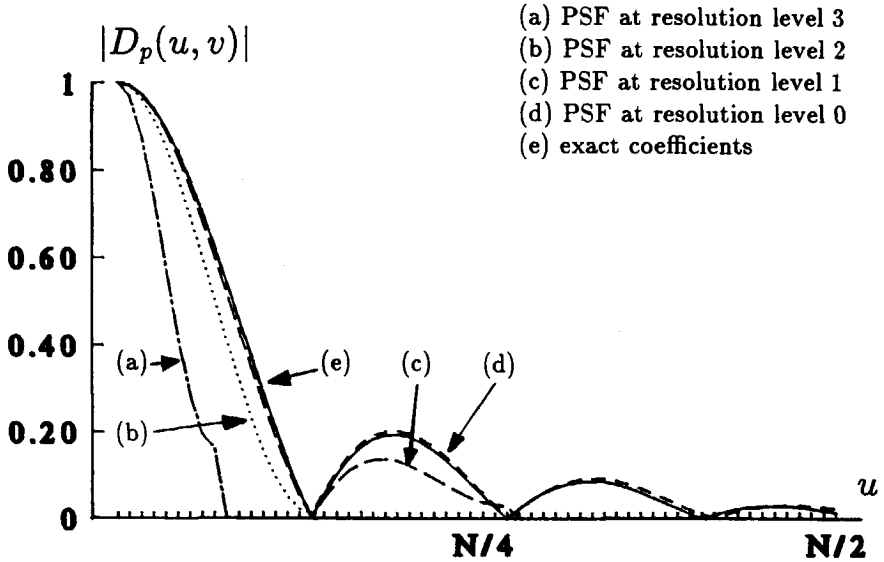


Figure 8.11: 1-D Transfer function of the exact and identified PSF at the resolution levels 3, 2, 1, and 0.

Table 8.7: Computation time for the hierarchical identification method.

		"CPU" (sec)					"CPU" (sec)	
level	$M = N$	E	M	total	iter.	runs	total	relative
4	16	0.6	0.3	0.9	100	1	90.0	13%
3	32	0.7	0.4	1.1	50	1	55.0	8%
2	64	0.7	0.4	1.1	50	2	110.0	16%
1	128	1.0	0.4	1.4	50	2	140.0	20%
0	256	2.6	0.4	3.0	50	2	300.0	43%
					300	8	695.0	

8.4.2 Defocusing Blur

The hierarchical identification procedure was applied to the out-of-focus cameraman image in Figure 7.6a (SNR=40 dB). The blurred image resolution pyramid and the associated restoration results are shown in Figure

8.12. We restrict ourselves here to listing the identified PSFs at the 5 different resolution levels. Starting out with the assumption of no blur at the resolution level 4, the PSFs identified at the levels 4 and 3 (both of which must have a 3×3 support) are given by:

$$d_4(m,n) = \begin{bmatrix} 0.036 & 0.137 & 0.039 \\ 0.114 & 0.349 & 0.114 \\ 0.039 & 0.137 & 0.036 \end{bmatrix}, \quad d_3(m,n) = \begin{bmatrix} 0.040 & 0.135 & 0.035 \\ 0.114 & 0.355 & 0.114 \\ 0.035 & 0.135 & 0.040 \end{bmatrix}.$$

At resolution level 2 we must decide between a 3×3 or 5×5 PSF support. The identified PSFs are

$$d_2(m,n) = \begin{bmatrix} 0.038 & 0.133 & 0.039 \\ 0.114 & 0.351 & 0.114 \\ 0.039 & 0.133 & 0.038 \end{bmatrix},$$

$$d_2(m,n) = \begin{bmatrix} 0.007 & 0.017 & 0.002 & -0.003 & 0.010 \\ 0.011 & 0.048 & 0.130 & 0.024 & 0.005 \\ -0.013 & 0.089 & 0.346 & 0.089 & -0.013 \\ 0.005 & 0.024 & 0.130 & 0.048 & 0.011 \\ 0.010 & -0.003 & 0.002 & 0.017 & 0.007 \end{bmatrix}.$$

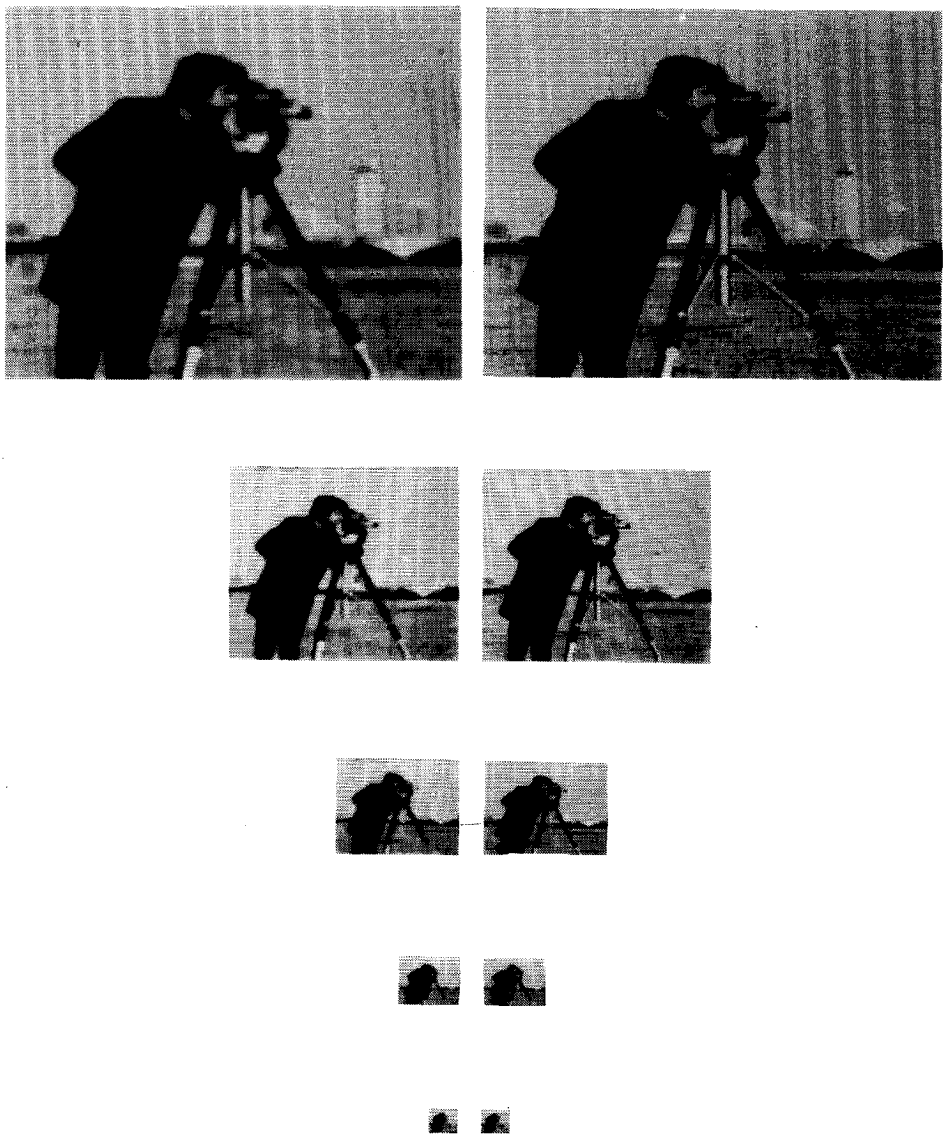
Since the likelihood value for both the 3×3 and 5×5 PSF support are equal to 3.48, preference is given to the smallest support size. On the next resolution level the identification results are given by:

$$d_1(m,n) = \begin{bmatrix} 0.061 & 0.125 & 0.060 \\ 0.123 & 0.265 & 0.123 \\ 0.060 & 0.125 & 0.061 \end{bmatrix},$$

$$d_1(m,n) = \begin{bmatrix} 0.012 & 0.017 & 0.010 & 0.014 & 0.009 \\ 0.009 & 0.060 & 0.107 & 0.057 & 0.007 \\ 0.000 & 0.095 & 0.205 & 0.095 & 0.000 \\ 0.007 & 0.057 & 0.107 & 0.060 & 0.009 \\ 0.009 & 0.014 & 0.010 & 0.017 & 0.012 \end{bmatrix}.$$

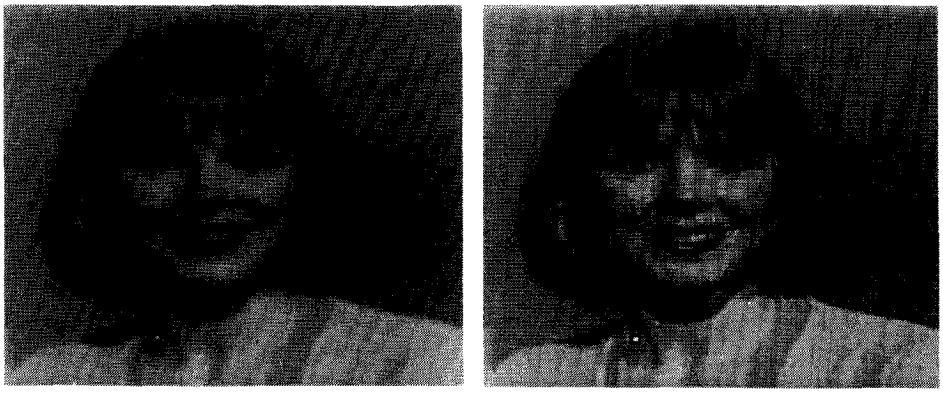
Since the value of the likelihood function for the two cases are very close to each other, namely 2.43 (3×3 PSF support) and 2.46 (5×5 PSF support), preference is given to the 3×3 PSF. Therefore, at the final resolution level 0 we consider again a 3×3 and 5×5 PSF support size:

$$d_0(m,n) = \begin{bmatrix} 0.064 & 0.149 & 0.053 \\ 0.130 & 0.208 & 0.130 \\ 0.053 & 0.149 & 0.064 \end{bmatrix},$$



(a) (b)

Figure 8.12: Resolution pyramid; (a) Defocused images, (b) Restored images.



(a) (b)

Figure 8.13: (a) Photographic out-of-focus blur; (b) Identified and restored image.

$$d_0(m,n) = \begin{bmatrix} 0.018 & 0.033 & 0.032 & 0.030 & 0.016 \\ 0.032 & 0.060 & 0.059 & 0.061 & 0.030 \\ 0.029 & 0.067 & 0.065 & 0.067 & 0.029 \\ 0.030 & 0.061 & 0.059 & 0.060 & 0.032 \\ 0.016 & 0.030 & 0.032 & 0.033 & 0.018 \end{bmatrix}.$$

In this case the likelihood value associated with the 5×5 PSF support is considerably smaller than the one computed from the 3×3 PSF, namely 1.64 compared to 1.81. Therefore, the PSF with support size 5×5 is considered to fit the blurred image best.

8.4.3 Photographic Out-of-Focus Blur

Figure 8.13a shows a portrait in which the out-of-focus blur is not synthetically introduced, but is due to a deliberately defocused camera lens[3]. Application of the hierarchical identification method to this image yields the following identified PSF:

[3]Courtesy of Eastman Kodak Co., Rochester NY.

$$\hat{d}(m,n) = \begin{bmatrix} 0.003 & 0.008 & 0.012 & 0.017 & 0.015 & 0.007 & 0.001 \\ 0.008 & 0.017 & 0.022 & 0.016 & 0.019 & 0.017 & 0.010 \\ 0.014 & 0.021 & 0.036 & 0.053 & 0.045 & 0.022 & 0.011 \\ 0.013 & 0.018 & 0.053 & 0.087 & 0.053 & 0.018 & 0.013 \\ 0.011 & 0.022 & 0.045 & 0.053 & 0.036 & 0.021 & 0.014 \\ 0.010 & 0.017 & 0.019 & 0.016 & 0.022 & 0.017 & 0.008 \\ 0.001 & 0.007 & 0.015 & 0.017 & 0.012 & 0.008 & 0.003 \end{bmatrix}.$$

This PSF was used in the constrained adaptive iterative restoration method described in Chapter 5. The restored image is shown in Figure 8.13b. Since the original image associated with the blurred image is not available, it is not possible to numerically evaluate the quality of the identified PSF and the restored image.

8.5 Status Quo and Prospective Developments

This thesis has discussed several modern approaches towards the image identification and restoration problem. A number of novel algorithms were introduced which outperform more conventional identification and restoration methods. More specifically, the methods that were developed are especially suited for blurred images with a medium to low SNR. Though numerically acceptable results are obtained which improve upon the results achieved by more conventional methods, this does, however, not mean that the visual quality is yet acceptable for all situations. Particularly at very low SNRs the quality improvements which can be achieved are limited because of the inherent (local) noise magnification of the restoration process.

The image formation model used in this thesis assumes that the nonlinear response of the sensor can either be neglected or can be inverted in a straightforward manner. Thes assumptions, however, do not hold at low SNRs. Furthermore, in order to invert the nonlinear response of the sensor, it needs to be known *a priori* which is not always the case. In some situations the sensor response can be modeled parametrically. For instance, if photographic film is used, a reasonable description between the intensity I_l of the incident light and the density D_g of the film grains deposited is given by:

$$D_g = c_1 \log(I_l) + c_2, \tag{8.10}$$

where c_1 and c_2 are parameters which define the behavior of the film. If

this relation is incorporated into the image formation model, the image identification problem now also encompasses the identification of the parameters c_1 and c_2. The complexity of this identification problem is much larger than the complexity of the ML image identification problem described in this thesis, because it is no longer possible to give an analytic expression for the probability density of the observed data. Approximations to the nonlinear model need to be made in order to be able to use an ML approach.

The PSF of the blur has been assumed stationary throughout this thesis. Non-stationary PSFs can be handled by restricting the identification and restoration process to a local region of the image. This can be done, for instance, by applying the EM-algorithm in a sliding block mode. It should be noted, however, that if the PSF changes too rapidly, the above approach will lead to very small blocks in which the PSF cannot be estimated due to the lack of data. In such a situation parametric PSFs which model the local variations of the PSF may be helpful, although this kind of modeling requires a lot of *a priori* knowledge.

An open problem in image identification is the identification of PSFs with arbitrary phases. The PSFs identified by the ML approach studied in this thesis do not have a unique phase in general. We circumvented this problem by assuming that the PSF is symmetric or has minimum phase. Obviously, practical blurs of interest may have any kind of phase. In order to be able to identify the phase of a PSF, we need to assume that the original image has a non-Gaussian character (i.e. the higher order moments of the modeling error $v(i,j)$ must be assumed to be non-zero). Again, this assumption will make the identification problem significantly more difficult due to the lack of criteria and the increased computational complexity.

The work in this thesis considered the identification and restoration of grey-valued images. These methods can also be applied to color images and image sequences by considering each color component or frame of the image sequence, respectively, as an individual grey-valued image. Such an approach does, however, not take into account the cross-color correlation between the components of the color image, or the interframe correlation between the frames of the image sequence. An improved approach should take into account these correlation properties by extending the 2-D image models currently employed to 3-D models. Since such an approach does not change the structure of the identification and restoration problem, iterative

identification and restoration methods can again be used succesfully.

Bibliography

[1] J.B. Abbiss, C. DeMol, and H.S. Dhadwal, "Regularized Iterative and Non-Iterative Procedures for Object Restoration from Experimental Data", *Optica Acta*, vol. 30, no. 1, pp. 107–124, 1983.

[2] G.L. Anderson and A.N. Netravali, "Image Restoration Based on a Subjective Criterion", *IEEE Trans. Syst., Man, Cybern.*, vol. 6, pp. 845–853, 1976.

[3] H.C. Andrews and B.R. Hunt, *Digital Image Restoration*, Prentice Hall, Englewood Cliffs NJ, 1977.

[4] D.L. Angwin, *Adaptive Image Restoration Using Reduced Order Model Based Kalman Filters*, Ph.D. Thesis Rensselaer Polytechnic Institute, Dep. ECSE, 1989.

[5] D.L. Angwin and H. Kaufman, "Image Restoration Using Reduced Order Models", *Signal Processing*, vol. 16, pp. 21–28, 1989.

[6] K.J. Åstrom, "Maximum Likelihood and Prediction Error Methods", *Automatica*, vol. 16, pp. 551–574, 1980.

[7] M. Bertero, C. DeMol, and G.A. Viano, "On the Regularization of Linear Inverse Problems in Fourier Optics," in *Applied Inverse Problems*, P.C. Sabatier (ed.), vol. 85, Springer-Verlag, Berlin, 1978.

[8] H. Bially, "Iterative Behandlung Linearer Functionalgleichungen", *Arch. Ration. Mech. Anal.*, vol. 4, pp. 166–176, 1959.

[9] J. Biemond, *Image Restoration: a Linear Stochastic Filtering Approach*, Ph.D. Thesis Delft University of Technology, Dep. EE, 1982.

[10] J. Biemond, J. Rieske and J.J. Gerbrands, "A Fast Kalman Filter for Images Degraded by both Blur and Noise", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 31, pp. 1248–1256, 1983.

[11] J. Biemond, *Stochastic Linear Image Restoration*, in: Advances in Computer Vision and Image Processing, vol. 2, pp. 213–273, JAI Press, Inc., 1986.

[12] J. Biemond and R.L. Lagendijk, "Digital Restoration of Noisy Blurred Images", *Proc. 1st Euro. Conf. Imaging Visual Documents Med.*, in *Excerpta Medica*, K. Wamsteker *et al.* (eds.), Amsterdam, pp. 65–76, 1987.

[13] J. Biemond, F.G. van der Putten and J.W. Woods, "Identification and Restoration of Images with Symmetric Noncausal Blurs", *IEEE Trans. Circuit and Systems*, vol. 23, no. 3, pp. 385–394, 1988.

[14] J. Biemond, R.L. Lagendijk and R.M. Mersereau, "Iterative Methods for Image Deblurring", *Proc. IEEE*, June 1990.

[15] L. Blanc-Féraud, M. Barlaud and P. Mathieu, "Amélioration de la restauration d'images floues utilisant une image miroir", *Revue Traitement du Signal*, 1988.

[16] L. Blanc-Féraud, M. Barlaud and P. Mathieu, "Image Restoration and Blur Estimation Using a Constrained Maximum Likelihood Method", *Proc. 3-rd International Workshop on Time-Varying Image Processing and Moving Object Recognition*, Florence, 1989.

[17] L.M. Bregman, "The Method of Successive Projections for Finding a Common Point of Convex Sets", *Soviet Math. Doklady*, no. 6, pp. 688–692, 1965.

[18] M. Cannon, "Blind Deconvolution of Spatially Invariant Image Blurs with Phase", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 24, pp. 58–63, 1976.

[19] M.R. Civanlar and H.J. Trussell, "Digital Image Restoration Using Fuzzy Sets", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, pp. 919–936, 1986.

[20] A. Dembo and O. Zeitouni, "Maximum A Posteriori Estimation of Time-Varying ARMA Processes from Noisy Data", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 471–476, 1988.

[21] A. Dembo, "Signal Reconstruction from Noisy Partial Information of Its Transform", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no.1, pp. 65–72, 1989.

[22] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood from Incomplete Data", *J. Royal Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[23] S.N. Efstratiadis and A.K. Katsaggelos, "Fast Adaptive Iterative Image Restoration Algorithms", *Proc. SPIE Conf. Visual Comm. Image Processing 1988*, Cambridge MA, pp. 10–17.

[24] M.P. Ekstrom and J.W. Woods, "Two-dimensional Spectral Factorization with Applications in Recursive Digital Filtering", *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 24, no. 2, pp. 115–128, 1976.

[25] C.L. Fales, F.O. Huck, J.A. McCormick and S.K. Park, "Wiener Restoration of Sampled Image Data: End-to-end Analysis", *J. Opt. Soc. Am. A*, vol. 5, no. 3, pp. 300–314, 1988.

[26] M. Feder, *Statistical Signal Processing Using a Class of Iterative Estimation Algorithms*, Ph.D. Thesis, Massachusetts Institute of Technology, Dep. of EECS, Research Laboratory of Electronics, 1987.

[27] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 95–103, 1988.

[28] R. Fletcher, *Practical Methods of Optimization, vol. 1: Unconstrained Optimization*, Wiley, Chichester, 1980.

[29] D. Girard, "Practical Optimal Regularization of Large Linear Systems", *Mathematical modeling and numerical analysis*, vol. 20, no. 1, pp. 75–87, 1986.

[30] R.C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison Wesley, Reading Mass, 1987.

[31] J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Book Co., New York, 1968.

[32] S.F. Gull and J. Skilling, "Maximum Entropy Method in Image Processing", *IEE Proc. F.*, vol. 131, no. 6, pp. 646–659, 1984.

[33] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.

[34] T.S. Huang, W.F. Schreiber and O.J. Tretiak, "Image Processing", *Proc. IEEE*, vol. 59, no. 11, pp. 1586–1609, 1972.

[35] T.S. Huang and P.M. Narendra, "Image Restoration by Singular Value Decomposition", *Applied Optics*, vol. 14, no. 9, pp. 2213–2216, 1975.

[36] B.R. Hunt, "The Application of Constrained Least-squares Estimation to Image Restoration by Digital Computer", *IEEE Trans. Computers*, vol. 22, no. 9, pp. 805–812, 1973.

[37] B.R. Hunt, "Bayesian Methods in Nonlinear Digital Image Restoration", *IEEE Trans. Comp.*, vol. 26, no. 3, pp. 219–229, 1977.

[38] Y. Ichioka and N. Nakajima, "Image Restoration Considering Visibility", *J. Opt. Soc. Amer.*, vol. 71, no. 8, pp. 983–988, 1981.

[39] Y. Ichioka, Y. Takabo, M. Matsuoka and T. Suzuki, "Iterative Image Restoration by a Method of Steepest Descent", *J. Optics*, vol. 12, no. 1, pp. 35–41, 1981.

[40] A.K. Jain, "Advances in Mathematical Models for Image Processing", *Proc. IEEE*, vol. 69, no. 5, pp. 502–528, 1981.

[41] F.-C. Jeng and J.W. Woods, "Inhomogeneous Gaussian Image Models for Estimation and Restoration", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol, 36, no. 8, pp. 1305–1312, 1988.

[42] N.B. Karayiannis and A.V. Venetsanopoulos, "Regularization Theory in Discrete Image Restoration", *Proc. SPIE Conf. Visual Comm. and Image Processing 1988*, Cambridge MA, pp. 25–36.

[43] R.L. Kashyap, "Image Models", in *Handbook of Pattern Recognition and Image Processing*, T.Y. Young and K.S. Fu (eds.), Academic Press, 1986.

[44] T. Katayama, T. Hirai and K. Okamura, "A Fast Kalman Filter Approach to Restoration of Blurred Images", *Signal Processing*, vol. 14, pp. 165–175, 1988.

[45] A.K. Katsaggelos, J. Biemond, R.M. Mersereau and R.W. Schafer, "An Iterative Method for Restoring Noisy Blurred Images", *Circ. Syst. Signal Proc.*, vol. 3, no. 2, pp. 139–160, 1984.

[46] A.K. Katsaggelos, *Constrained Iterative Image Restoration Algorithms*, Ph.D. Thesis, Georgia Institute of Technology, School of EE., 1985.

[47] A.K. Katsaggelos, "A General Formulation of Adaptive Iterative Image Restoration Algorithms", *Proc. 1986 Princeton Conf. Info. Sciences and Systems*, pp. 42–46.

[48] A.K. Katsaggelos and K.T. Lay, "Simultaneous Identification and Restoration of Images Using Maximum Likelihood Estimation", *Proc. 1989 IEEE Int. Conf. Control and Applic.*, Jeruzalem, pp. TP-5-1.

[49] S. Kawata and Y. Ichioka, "Iterative Image Restoration for Linearly Degraded Images, I. Basis & II. Reblurring Procedure", *J. Opt. Soc. Am.*, vol. 70, pp. 762–772, 1980.

[50] H. Kwakernaak and R. Sivan, *Linear Optimal Control*, Wiley, New York, 1972.

[51] R.L. Lagendijk, R.M. Mersereau and J. Biemond, "On Increasing the Convergence Rate of Regularized Iterative Image Restoration Algorithms", *Proc. 1987 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Dallas, pp. 1183–1186.

[52] R.L. Lagendijk, A.K. Katsaggelos and J. Biemond, "Iterative Identification and Restoration of Images", *Proc. 1988 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, New York, pp. 992–995.

[53] R.L. Lagendijk, D.L. Angwin, H. Kaufman and J. Biemond, "Recursive and Iterative Methods for Image Identification and Restoration", *Proceedings Fourth European Signal Processing Conf. EUSIPCO'88*, J.G. Lacoume (ed.), Grenoble, pp. 235–238.

[54] R.L. Lagendijk, J. Biemond and D.E. Boekee, "Simultaneous Image Identification and Restoration Using the EM-Algorithm", *Proc. SPIE Int. Conf. Visual Comm. and Image Processing 1988*, Cambridge, Mass., pp. 2–9.

[55] R.L. Lagendijk, J. Biemond and D.E. Boekee, "Regularized Iterative Image Restoration with Ringing Reduction", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 12, pp. 1874–1888, 1988.

[56] R.L. Lagendijk, J. Biemond and D.E. Boekee, "Blur Identification Using the Expectation-Maximization Algorithm", *Proc. 1989 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Glasgow, pp. 1397–1400.

[57] R.L. Lagendijk, J. Biemond and D.E. Boekee, "Identification and Restoration of Noisy Blurred Images Using the Expectation-Maximization Algorithm", *IEEE Trans. Acoustics, Speech and Signal Processing*, July 1990.

[58] R.L. Lagendijk, J. Biemond and D.E. Boekee, "Hierarchical Blur Identification", *Proc. 1990 IEEE Int. Conf. Acoustics, Speech and Signal Proc.*, Albuquerque NM.

[59] R.L. Lagendijk, A.M. Tekalp and J. Biemond, "Maximum Likelihood Image and Blur Identification: A Unifying Approach", *J. Optical Engineering*, May 1990.

[60] L. Landweber, "An Iteration Formula for Fredholm Integral Equations of the First Kind", *Am. J. Math.*, vol. 73, pp. 615–624, 1951.

[61] A. Lannos, S. Roques, and M.J. Casonove, "Resolution and Robustness in Image Processing: A New Regularization Principle", *J. Opt. Soc. Amer.*, vol. 4, no. 1, pp. 189–199, 1987.

[62] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, The MIT Press, Cambridge, Mass, 1983.

[63] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison Wesley, Reading MA, 1973.

[64] J. Maeda and K. Murata, "Restoration of Band-Limited Images by an Iterative Regularized Pseudo-Inverse Method", *J. Opt. Soc. Am. A*, vol. 1, no. 1, pp. 28–34, 1984.

[65] R. Marucci, R.M. Mersereau and R.W. Schafer, "Constrained Iterative Deconvolution with Optimized Rate of Convergence", *Proc. 1982 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 1845–1848.

[66] B.L. McGlamery, "Restoration of Turbulence Degraded Images", *J. Opt. Soc. Am.*, vol. 57, no. 3, pp. 293–297, 1967.

[67] K. Miller, "Least Squares Methods for Ill-Posed Problems with a Prescribed Bound", *SIAM J. Math. Anal.*, vol. 1, no. 1, pp. 52–74, 1970.

[68] C.E. Morris, M.A. Richards and M.H. Hayes, "Iterative Deconvolution Algorithm with Quadratic Convergence", *J. Opt. Soc. Amer. A.*, 1987.

[69] C.E. Morris, M.A. Richards and M.H. Hayes, "Fast Reconstruction of Linearly Distorted Signals", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 7, pp. 1017–1025, 1988.

[70] J.B. Morton and H.C. Andrews, "*A Posteriori* Method of Image Restoration", *J. Opt. Soc. Am.*, vol. 69, no. 2, pp. 280–290, 1979.

[71] B.R. Musicus and J.S. Lim, "Maximum Likelihood Parameter Estimates of Noisy Data", *Proc. 1979 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 224–227.

[72] M.Z. Nashed, "Aspects of Generalized Inverses in Analysis and Regularization," in *Generalized Inverses and Applications*, Academic Press, New York, pp. 193–244, 1976.

[73] M.Z. Nashed, "Operator Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory", *IEEE Trans. Antennas Prop.*, vol. 29, no. 2, pp. 220–231, 1981.

[74] M.Z. Nashed (ed.), *Ill-posed Problems: Theory and Practice*, Reidel, Dordrecht, 1981.

[75] D.L. Phillips, "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind", *J. Ass. Comput. Mach.*, vol. 9, pp. 84–97, 1962.

[76] E.R. Pike, J.G. McWhirter, M. Bertero, and C. DeMol, "Generalized Information Theory for Inverse Problems in Signal Processing", *IEE Proc. F*, vol. 131, no. 6, pp. 660–667, 1984.

[77] M. Potmesil and I. Chakravarty, "Synthetic Image Generation with a Lens and Aperture Camera Model", *ACM Trans. Graph.*, vol. 1, no. 2, pp. 25–108, 1982.

[78] W.K. Pratt, "Vector Space Formulation of Two-Dimensional Signal Processing Operations", *Comp. Graph. and Image Processing*, no. 4, pp. 1–24, 1975.

[79] R. Prost and R. Goutte, "Discrete Constrained Iterative Deconvolution with Optimized Rate of Convergence", *Signal Processing*, pp. 209–230, 1984.

[80] M.V. Ranganath, A.P. Dhawan and N. Mullani, "A Multigrid Expectation-Maximization Reconstruction Algorithm for Positron Emission Tomography", *IEEE Trans. Med. Imaging*, vol. 7, no. 4, pp. 273–277, 1988.

[81] J.L.C. Sanz and T.S. Huang, "Unified Hilbert Space Approach to Iterative Least-squares Linear Signal Restoration", *J. Opt. Soc. Amer.*, vol. 73, no. 11, pp. 1455–1465, 1983.

[82] T.K. Sarkar, K.R. Siarkiewicz and R.F. Stratton, "Survey of Numerical Methods for Solution of Large Systems of Linear Equations for Electromagnetic Field Problems", *IEEE Trans. Antenn. Prop.*, vol. 29, no. 6, 1981.

[83] T.K. Sarkar, F.I. Tseng, S.A. Dianat and B.Z. Hollmann, "Deconvolution by the Conjugate Gradient Method", *Proc. 1985 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Tampa, pp. 445–448.

[84] R.W. Schafer, R.M. Mersereau and M.A. Richards, "Constrained Iterative Signal Restoration Algorithms", *Proc. IEEE*, vol. 69, no. 4, pp. 4322–450, 1981.

[85] F.C. Schweppe, *Uncertain Dynamic Systems*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1973.

[86] M. Segal and E. Weinstein, "The Cascade EM Algorithm", *Proc. IEEE* , vol. 76, no. 10, pp. 1388–1390, 1988.

[87] M.I. Sezan and H. Stark, "Image Restoration by the Method of Convex Projections: Part 2 — Applications and Numerical Results", *IEEE Trans. Med. Imaging*, vol. 1, no. 2, pp. 95–101, 1982.

[88] M.I. Sezan, A.M. Tekalp and C. Chen, " Regularized Signal Restoration Using the Theory of the Convex Projections", *Proc. 1987 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Dallas, pp. 1565–1568.

[89] M.I. Sezan and A.M. Tekalp, "Iterative Image Restoration with Ringing Suppression Using POCS", *Proc. 1988 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, New York, pp. 1300–1303.

[90] S. Singh, S.N. Tandon and H.M. Gupta,"An Iterative Restoration Technique", *Signal Processing*, vol. 11, pp. 1–11, 1986.

[91] H. Stark (ed.), *Image Recovery: Theory and Application*, Academic Press, Orlando, 1987.

[92] K. Stewart and T.S. Duranni, "Constrained Signal Reconstruction — A Unified Approach", *Signal Processing III: Theories and Applications*, I.T. Young *et al.* (eds.), Elsevier North Holland, Amsterdam, 1986.

[93] T.G. Stockham, T.M. Cannon and R.B. Ingebretsen, "Blind Deconvolution through Digital Signal Processing", *Proc. IEEE* , vol. 64, no. 4, pp. 678–692, 1975.

[94] P.A. Stokseth, "Properties of a Defocussed Optical System", *J. Opt. Soc. Am.*, vol. 59, pp. 1314–1321, 1969.

[95] O.N. Strand, "Theory and Methods Related to the Singular Function Expansion and Landweber's Iteration for Integral Equations of the First Kind", *Siam J. Numer. Anal.*, vol. 11, no. 4, pp. 798–825, 1974.

[96] A.M. Tekalp, H. Kaufman and J.W. Woods, "Identification of Image and Blur Parameters for the Restoration of Noncausal Blurs", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, pp. 963–972, 1986.

[97] A.M. Tekalp and H. Kaufman, "On Statistical Identification of a Class of Linear Space-Invariant Blurs Using Non-Minimum-Phase

ARMA Models", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 8, pp. 1360–1363, 1988.

[98] A.M. Tekalp, H. Kaufman and J.W. Woods, "Edge-Adaptive Kalman Filtering for Image Restoration with Ringing Suppression", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no. 6, pp. 892–899, 1989.

[99] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Wiley, New York, 1977.

[100] V.T. Tom, T.F. Quartieri, M.H. Hayes and J.H. McClellan, "Convergence of Iterative Nonexpansive Signal Reconstruction Algorithms", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, pp. 1052–1058, 1981.

[101] H.J. Trussell and B.R. Hunt, "Improved Methods of Maximum A Posteriori Restoration", *IEEE Trans. Comp.*, vol. 27, pp. 57–62, 1979.

[102] H.J. Trussell, "Maximum Power Signal Restoration", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 5, 1981.

[103] H.J. Trussell and M.R. Civanlar, "The Initial Estimate in Constrained Iterative Restoration", *Proc. 1983 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Boston, pp. 643–646.

[104] H.J. Trussell and M.R. Civanlar, "The Feasible Solution in Signal Restoration", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 32, pp. 201–212, 1984.

[105] S. Twomey, "On the Numerical Solution of Fredholm Integral Equations of the First Kind by Inversion of the Linear Systems Produced by Quadrature", *J. Ass. Comput. Mach.* vol. 10, pp. 97–101, 1963.

[106] P.H. VanCittert, "Zum Einfluss der Spaltbreit auf die Intensitätsverteilung in Spektrallinien II", *Z. Physik*, vol. 69, pp. 298–308, 1931.

[107] G.R. Wagner, *Self-Tuning Algorithms for Two-Dimensional Signal Processing*, Ph.D. Thesis, Univ. of Manchester, Control Syst. Centre, 1987.

[108] F.C. Wagner, A. Macovski and D.G. Nishimura, "A Characterization of the Scatter Point-Spread Function in Term of Air Gaps", *IEEE Trans. Medical Imaging*, vol. 7, no. 4, pp. 337–344, 1988.

[109] G. Wahba, "Practical Approximate Solutions to Linear Operator Equations when the Data are Noisy", *SIAM J. Math. Anal.*, vol. 14, no. 4, pp. 651–667, 1977.

[110] J.W. Woods, "Two-Dimensional Discrete Markovian Fields", *IEEE Trans. Information Theory*, vol. 18, no. 2, pp. 232–240, 1972.

[111] J.W. Woods and C.H. Radewan, "Kalman Filtering in Two Dimensions", *IEEE Trans. Information Theory*, vol. 23, no. 4, pp. 473–482, 1977.

[112] J.W. Woods and V.K. Ingle, "Kalman Filtering in Two-Dimensions: Further Results", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 188–197, 1981.

[113] J.W. Woods, J. Biemond and A.M. Tekalp, "Boundary Value Problem in Image Restoration", *Proc. 1985 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Tampa, pp. 692–695.

[114] C.F.J. Wu, "On the Convergence Properties of the EM Algorithm", *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983

[115] D.C. Youla and H. Webb, "Image Restoration by the Method of Convex Projections: Part 1 — Theory", *IEEE Trans. Med. Imaging*, vol. 1, no. 2, pp. 81–94, 1982.

Appendix A

Eigenvalue Analysis for 2-D Systems

In this appendix we will briefly review several mathematical tools used in linear image processing, such as lexicographic ordering of image data, convolutions, block-circulant matrices, eigenvalues of 2-D matrix operations, and the 2-D discrete Fourier transform (DFT). More on these topics can be found in the paper by Pratt [78], and the textbooks by Andrews and Hunt [3], and Gonzalez and Wintz [30].

Lexicographic Ordering

Let the 2-D array $\mathcal{F}$ represent a discrete image with M rows and N columns:

$$\mathcal{F} = \{f(i,j)\}, \quad 0 \leq i \leq M-1,\ 0 \leq j \leq N-1. \tag{A.1}$$

This data can be written in a vector form through a 1-D mapping that is known as lexicographic ordering. In essence the $M \times N$ 2-D array is converted to a $MN \times 1$ vector $\boldsymbol{f}$ by concatenating its rows. Thus

$$\begin{aligned} \boldsymbol{f} \;=\; & [f(0,0), f(0,1), f(0,2), \cdots, f(0,N-1), \\ & f(1,0), f(1,1), \cdots, f(M-1,N-1)]^t . \end{aligned} \tag{A.2}$$

Block-Toeplitz and Block-Circulant Matrices

Neglecting the noise contribution $w(i,j)$ the image formation model is given by:

$$g(i,j) \;=\; \sum_{m,n \in S_d} d(m,n) f(i-m, j-n)$$

$$\begin{bmatrix} \boldsymbol{g}_0 \\ \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_{M-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_{-1} & \boldsymbol{D}_{-2} & & \boldsymbol{O} \\ \boldsymbol{D}_1 & \boldsymbol{D}_0 & \boldsymbol{D}_{-1} & & \\ \boldsymbol{D}_2 & \boldsymbol{D}_1 & \boldsymbol{D}_0 & \cdots & \\ & & \vdots & & \vdots \\ \boldsymbol{O} & & & \cdots & \boldsymbol{D}_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_0 \\ \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \vdots \\ \boldsymbol{f}_{M-1} \end{bmatrix}$$

$$\boldsymbol{D}_j = \begin{bmatrix} d(j,0) & d(j,-1) & d(j,-2) & & 0 \\ d(j,1) & d(j,0) & d(j,-1) & & \\ d(j,2) & d(j,1) & d(j,0) & \cdots & \\ & & \vdots & & \vdots \\ 0 & & & \cdots & d(j,0) \end{bmatrix}$$

Figure A.1: Block-Toeplitz structure of the matrix $\boldsymbol{D}$. The vectors $\boldsymbol{g}_j$ and $\boldsymbol{f}_j$ refer to the j^{th} row of $\boldsymbol{g}$ and $\boldsymbol{f}$, respectively.

$$= d(i,j) * f(i,j). \tag{A.3}$$

By lexicographically ordering both $f(i,j)$ and $g(i,j)$ we have

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{f}, \tag{A.4}$$

where $\boldsymbol{D}$ is the blurring matrix which is of size $MN \times MN$. (Observe that $\boldsymbol{D}$ will always be square if $\boldsymbol{f}$ and $\boldsymbol{g}$ are of the same size.) If the convolution in (A.3) is interpreted as a linear convolution (with zero boundaries), the matrix $\boldsymbol{D}$ has a block-Toeplitz structure. If it is partitioned into M^2 submatrices of size $N \times N$, each of these submatrices will be a Toeplitz matrix. Further the submatrices are arranged in a Toeplitz pattern. This is illustrated in Figure A.1.

A (block-)Toeplitz matrix is often approximated by a (block-)circulant one because these two matrix types are structurally closely related, and operations involving block-circulant matrices can be efficiently evaluated using 2-D discrete Fourier transforms. The errors introduced by the approximations are usually small, and mainly effect the boundaries of the image. With a block-circulant matrix $\boldsymbol{D}_j = \boldsymbol{D}_{j-M}$ and the elements $d(j,k)$ of $\boldsymbol{D}_j$ are replaced by $d(j,k-N)$. Figure A.2 illustrates this conversion on

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_{-1} & \boldsymbol{D}_{-2} & & \boldsymbol{D}_1 \\ \boldsymbol{D}_1 & \boldsymbol{D}_0 & \boldsymbol{D}_{-1} & & \vdots \\ \boldsymbol{D}_2 & \boldsymbol{D}_1 & \boldsymbol{D}_0 & & \\ & & & \cdots & \\ \vdots & & \vdots & & \vdots \\ \boldsymbol{D}_{-1} & & & \cdots & \boldsymbol{D}_0 \end{bmatrix}$$

$$\boldsymbol{D}_j = \begin{bmatrix} d(j,0) & d(j,-1) & d(j,-2) & & d(j,1) \\ d(j,1) & d(j,0) & d(j,-1) & & \\ d(j,2) & d(j,1) & d(j,0) & \cdots & \\ & & \vdots & & \vdots \\ d(j,-1) & & & \cdots & d(j,0) \end{bmatrix}$$

Figure A.2: Block-circulant approximation of the matrix $\boldsymbol{D}$.

the matrices shown in Figure A.1. In signal processing terms the approximation of a block-Toeplitz matrix by a block-circulant one is equivalent to replacing the linear convolution in (A.3) by a circulant one.

Eigensystem of a Matrix

A powerful tool in both the analysis and implementation of linear equations such as (A.3) is the set of eigenvalues and eigenvectors (eigensystem) of the matrix $\boldsymbol{D}$ involved. Let $\{\boldsymbol{z}_{uv}\}$ denote the normalized (complex-valued) eigenvectors associated with the blurring matrix $\boldsymbol{D}$, and let the (complex-valued) scalars λ_{uv} represent the corresponding eigenvalues, i.e.

$$\boldsymbol{D}\boldsymbol{z}_{uv} = \lambda_{uv}\boldsymbol{z}_{uv}, \quad 0 \leq u \leq M-1,\ 0 \leq v \leq N-1. \tag{A.5}$$

The double subscript notation is used since the eigenvectors when unstacked are, in fact, images. They are actually the set of images $\{z_{uv}(i,j)\}$ which are unaffected by the blurring operator except for a change of scale. Writing out the eigenvector equations we get

$$\sum_{m,n \in S_d} d(m,n) z_{uv}(i-m, j-n) = \lambda_{uv} z_{uv}(i,j), \quad \forall i,j,\ \forall u,v. \tag{A.6}$$

We now assume without loss of generality, that the matrices $\boldsymbol{D}$ and $\boldsymbol{D}'$ have the same set of eigenvectors[1]. As a result the eigenvectors $\boldsymbol{z}_{uv}$ are mutually orthogonal and the matrix $\boldsymbol{D}$ can then be expressed in terms of its eigenvectors as

$$\boldsymbol{D} = \sum_{u,v} \lambda_{uv} \boldsymbol{z}_{uv} \boldsymbol{z}_{uv}^{*}. \tag{A.7}$$

The advantage of using the eigensystem of a matrix is that linear relations such as (A.3), can be evaluated as a set of independent scalar equations. To this end we first decompose $\boldsymbol{f}$ in terms of the eigenvectors of $\boldsymbol{D}$, i.e.

$$\boldsymbol{f} = \sum_{u,v} (\boldsymbol{f}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv}, \tag{A.8}$$

where $(\boldsymbol{f}, \boldsymbol{z}_{uv})$ denotes the inner product between $\boldsymbol{f}$ and the uv^{th} eigenvector. Substituting (A.8) into (A.4) yields

$$\begin{aligned} \boldsymbol{g} &= \boldsymbol{D}\boldsymbol{f} = \boldsymbol{D} \sum_{u,v} (\boldsymbol{f}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv} \\ &= \sum_{u,v} (\boldsymbol{f}, \boldsymbol{z}_{uv}) \boldsymbol{D}\boldsymbol{z}_{uv} \\ &= \sum_{u,v} (\boldsymbol{f}, \boldsymbol{z}_{uv}) \lambda_{uv} \boldsymbol{z}_{uv}. \end{aligned} \tag{A.9}$$

By also expanding $\boldsymbol{g}$ in terms of the eigenvectors of $\boldsymbol{D}$ we get

$$\sum_{u,v} (\boldsymbol{g}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv} = \sum_{u,v} \lambda_{uv} (\boldsymbol{f}, \boldsymbol{z}_{uv}) \boldsymbol{z}_{uv}, \tag{A.10}$$

or

$$(\boldsymbol{g}, \boldsymbol{z}_{uv}) = \lambda_{uv} (\boldsymbol{f}, \boldsymbol{z}_{uv}), \quad \forall u, v. \tag{A.11}$$

Thus by virtue of using the eigensystem of $\boldsymbol{D}$, we have replaced the convolution (A.3) by a far simpler set of scalar equations (A.11).

2-D Discrete Fourier Transform

In general it is hardly ever possible to find the eigenvectors $\{\boldsymbol{z}_{uv}\}$ of an arbitrary matrix $\boldsymbol{D}$ of size $MN \times MN$ because of its size. This becomes possible only when $\boldsymbol{D}$ is highly structured. In the particular case where $\boldsymbol{D}$ is block-circulant, the eigenvectors are the following complex exponentials:

$$\boldsymbol{z}_{uv}(i,j) = \exp\left\{-2\pi\jmath\left(\frac{iu}{M} + \frac{jv}{N}\right)\right\}, \tag{A.12}$$

[1]Similar expressions can be derived if this assumption does not hold by making use of the singular value analysis of the matrix $\boldsymbol{D}$. Essentially, this involves an eigenvalue analysis of the matrix $\boldsymbol{D}^{t}\boldsymbol{D}$ instead of the matrix $\boldsymbol{D}$ [3,72].

and the eigenvalues are equal to the discrete Fourier transform (DFT) samples, which are computed as

$$\lambda_{uv} = \sum_{i,j} d(i,j) z_{uv}(i,j), \tag{A.13}$$

which can be evaluated efficiently using an FFT algorithm. Further the inner products $(\boldsymbol{f}, \boldsymbol{z}_{uv})$ and $(\boldsymbol{g}, \boldsymbol{z}_{uv})$ are the 2-D DFTs of $f(i,j)$ and $g(i,j)$, respectively. In this case (A.11) is usually written as:

$$\begin{aligned} (\boldsymbol{g}, \boldsymbol{z}_{uv}) &= \lambda_{uv}(\boldsymbol{f}, \boldsymbol{z}_{uv}), \quad \forall u, v, \\ \Rightarrow \quad G(u,v) &= D(u,v)F(u,v) \quad \forall u, v. \end{aligned} \tag{A.14}$$

As a result (A.14) is called the discrete Fourier transform of the discrete convolution (A.3) or the matrix-vector equation (A.4).

Appendix B

Properties of the Iteration (5.21)

In this appendix we will prove that the iterations (5.21) minimize the objective function $\Phi(f)$ in (5.15) on a closed convex set. To this end we will first give the definitions of a nonexpansive mapping and a contraction mapping, and we will state the contraction mapping theorem. Next it will be shown that $\mathcal{G}(f)$ in (5.17) is a contraction mapping under certain conditions. Finally the mentioned proof will be given.

Contraction Mapping

Consider the properties of a general mapping $I(x)$. Suppose that

$$\|I(x_i) - I(x_j)\| \leq r\|x_i - x_j\|, \tag{B.1}$$

for x_i and x_j in the considered solution space. If $0 \leq r < 1$, the mapping $I(x)$ is said to be a contraction mapping, or simply a contraction. If $r = 1$, the mapping is nonexpansive [100]. Contraction mappings essentially reduce the distance between two mapped elements x_i and x_j, while nonexpansive mappings do at least not increase this distance. Concatening nonexpansive mappings yields a nonexpansive mapping, while concatenating nonexpansive and contraction mappings yields a contraction mapping.

The fixed points of a mapping are defined as those points which are not affected by the mapping. If a mapping is a contraction, then is has a unique fixed point x^*, such that $x^* = I(x^*)$. Further, every sequence of successive approximations, defined as

$$x_{k+1} = I(x_k), \tag{B.2}$$

converges to the unique fixed point x^* for any initial starting point x_0 if I is a contraction mapping: this sentence states the contraction mapping theorem. Hence, for contraction mappings the limiting solution of (B.2) exists and is unique.

Unfortunately, many mappings are only nonexpansive, in which case any number of fixed points may exist and the method of successive approximations may not converge.

Mappings which are associated with a projection onto a closed convex set are always nonexpansive [115]. The elements of the convex set are the fixed points of such a mapping.

Contractiveness of $\mathcal{G}$

We next show that the mapping $\mathcal{G}(\boldsymbol{f})$ in Eq. (5.17) is a contraction mapping [55]. This mapping has been defined as

$$\begin{aligned} \mathcal{G}(\boldsymbol{f}) &= (\boldsymbol{I} - \alpha\beta\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\boldsymbol{f} + \beta\boldsymbol{D}^t\boldsymbol{R}(\boldsymbol{g} - \boldsymbol{D}\boldsymbol{f}) \\ &= \boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\boldsymbol{f} + \beta\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{g}. \end{aligned} \tag{B.3}$$

Substitution of $\mathcal{G}(\boldsymbol{f})$ into (B.1) yields:

$$\begin{aligned} &\|\mathcal{G}(\boldsymbol{f}_1) - \mathcal{G}(\boldsymbol{f}_2)\| \\ &= \|\{\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C}\}\boldsymbol{f}_1 - \{\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C}\}\boldsymbol{f}_2\| \\ &\leq \|\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C}\|\,\|\boldsymbol{f}_1 - \boldsymbol{f}_2\| \leq r\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|. \end{aligned} \tag{B.4}$$

If we now require that $0 \leq r < 1$, then $\mathcal{G}(\boldsymbol{f})$ is contractive. This leads to the following inequalities:

$$0 \leq \|\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\| < 1, \tag{B.5}$$

where $\|\cdot\|$ denotes the norm of the matrix. Since $(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$ is a symmetric matrix, the eigenvalues of this matrix, denoted by ρ_{uv}, are real valued, as well as the eigenvalues of $\boldsymbol{I} - \beta(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$. We can therefore replace (B.5) by

$$0 \leq \max_{u,v}\{|1 - \beta\rho_{uv}|\} < 1. \tag{B.6}$$

The above inequalities are always satisfied if β satisfies

$$0 < \beta < \frac{2}{|\rho_{max}|}, \tag{B.7}$$

where μ_{max} is the largest eigenvalue of the matrix $(\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D}+\alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})$. Hence, $\mathcal{G}$ is a contraction if (B.7) is satisfied. It is observed that (B.7) is identical to the condition (5.19), which ensured the convergence of the unconstrained iterations (5.18).

Properties of the Iteration (5.21)

We now consider the properties of the limiting solution of the constrained iterative process (5.21). We will show that (5.21) minimizes the objective function $\Phi(\boldsymbol{f})$ in (5.15) on the closed convex set $\mathcal{C}$.

First, we define the gradient of $\Phi(\boldsymbol{f})$ with respect to $\boldsymbol{f}$ as:

$$\boldsymbol{r}(\boldsymbol{f}) = -\frac{1}{2}\nabla_{\boldsymbol{f}}\Phi(\boldsymbol{f}) = \boldsymbol{D}^t\boldsymbol{R}_{\mathcal{C}} - (\boldsymbol{D}^t\boldsymbol{R}\boldsymbol{D} + \alpha\boldsymbol{C}^t\boldsymbol{S}\boldsymbol{C})\boldsymbol{f}. \tag{B.8}$$

According to (5.17) or (5.22), the mapping $\mathcal{G}(\boldsymbol{f})$ is created as follows:

$$\mathcal{G}(\boldsymbol{f}) = \boldsymbol{f} + \beta\boldsymbol{r}(\boldsymbol{f}). \tag{B.9}$$

The iterative algorithm (5.21) is then given by:

$$\hat{\boldsymbol{f}}_{k+1} = \mathcal{P}\mathcal{G}(\hat{\boldsymbol{f}}_k) = \mathcal{P}\left[\hat{\boldsymbol{f}}_k + \beta\boldsymbol{r}(\hat{\boldsymbol{f}}_k)\right]. \tag{B.10}$$

Since the mapping $\mathcal{G}$ is assumed to be a contraction (provided that (B.7) is satisfied), and the mapping $\mathcal{P}$ is nonexpansive, the concatenated mapping $\mathcal{P}\mathcal{G}$ is a contraction mapping. By the contraction mapping theorem, the iterations in (B.10) converge to the unique fixed point $\hat{\boldsymbol{f}}_\ell$ of the mapping $\mathcal{P}\mathcal{G}$. As a consequence, the following relation has to hold for $\hat{\boldsymbol{f}} = \hat{\boldsymbol{f}}_\ell$:

$$\hat{\boldsymbol{f}}_\ell = \mathcal{P}\left[\hat{\boldsymbol{f}}_\ell + \beta\boldsymbol{r}(\hat{\boldsymbol{f}}_\ell)\right]. \tag{B.11}$$

This relation is satisfied in three possible situations, namely:

1) The parameter $\beta = 0$. This, however, contradicts the bounds on β, as given in (B.7).

2) The gradient $\boldsymbol{r}(\hat{\boldsymbol{f}}_\ell) = 0$. This solution corresponds to the unconstrained minimization of $\Phi(\boldsymbol{f})$. Obviously, the limiting solution of (B.10) coincides with the Tikhonov-Miller solution obtained from (5.16). Therefore, in this case the Tikhonov-Miller solution satisfies the constraint $\mathcal{C}$.

3) The gradient is not equal to zero for $\hat{f} = \hat{f}_\ell$. To satisfy the relation (B.11), the projection operator P must exactly compensate the modifications in $\hat{f}_\ell$ in the direction $r(\hat{f}_\ell)$. Consequently, $r(\hat{f}_\ell)$ has to be orthogonal to the surface of the convex set C at the solution $\hat{f}_\ell$. By definition, $r(\hat{f}_\ell)$ is orthogonal to the surface of the set K, defined by

$$K = \{h | \Phi(h) \leq \Phi(\hat{f}_\ell)\}. \tag{B.12}$$

Since (B.12) defines a closed strictly convex set [115] and the constraint C a closed convex set, and the surfaces of these two sets are both orthogonal to $r(\hat{f}_\ell)$ at $\hat{f} = \hat{f}_\ell$, we conclude that their intersection consists of exactly one element (namely, $\hat{f}_\ell$). Therefore, the objective function $\Phi(f)$ attains its smallest value on the set C at the solution $\hat{f}_\ell$.

Hence, we have shown that the iteration (5.21) minimizes the objective function $\Phi(f)$ on the closed convex set C.

Appendix C

Derivation of Equation (7.14)

In this appendix the derivation of (7.14) is given. To this end we first substitute (7.10) into (7.9), yielding

$$\begin{aligned}
\mathcal{L}(\theta;\hat{\theta}^{(k)}) &= \mathrm{E}\left\{\log p(\boldsymbol{f},\boldsymbol{g};\theta)/\boldsymbol{g};\hat{\theta}^{(k)}\right\} \\
&= -\frac{1}{2}\log(2\pi^{2MN}) - \frac{1}{2}\log\det|\boldsymbol{Q}_w\boldsymbol{Q}_v| + \frac{1}{2}\log\det|\boldsymbol{I}-\boldsymbol{A}|^2 + \\
&\quad -\frac{1}{2}\mathrm{E}\left\{(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})^t\boldsymbol{Q}_w^{-1}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f}) + \boldsymbol{f}^t(\boldsymbol{I}-\boldsymbol{A})^t\boldsymbol{Q}_v^{-1}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)}\right\} \\
&= c_1 - \frac{1}{2}MN\log(\sigma_w^2\sigma_v^2) + \frac{1}{2}\log\det|\boldsymbol{I}-\boldsymbol{A}|^2 + \\
&\quad -\frac{1}{2}\mathrm{E}\left\{(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})^t\boldsymbol{Q}_w^{-1}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})/\boldsymbol{g};\hat{\theta}^{(k)}\right\} \\
&\quad -\frac{1}{2}\mathrm{E}\left\{\boldsymbol{f}^t(\boldsymbol{I}-\boldsymbol{A})^t\boldsymbol{Q}_v^{-1}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)}\right\}. \qquad \text{(C.1)}
\end{aligned}$$

If we have a closer look at the 4-th term of (C.1), we find

$$\begin{aligned}
&\mathrm{E}\left\{(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})^t\boldsymbol{Q}_w^{-1}(\boldsymbol{g}-\boldsymbol{D}\boldsymbol{f})/\boldsymbol{g};\hat{\theta}^{(k)}\right\} \\
&= \mathrm{E}\left\{\boldsymbol{g}^t\boldsymbol{Q}_w^{-1}\boldsymbol{g}/\boldsymbol{g};\hat{\theta}^{(k)}\right\} - 2\mathrm{E}\left\{\boldsymbol{f}^t\boldsymbol{D}^t\boldsymbol{Q}_w^{-1}\boldsymbol{g}/\boldsymbol{g};\hat{\theta}^{(k)}\right\} + \mathrm{E}\left\{\boldsymbol{f}^t\boldsymbol{D}^t\boldsymbol{Q}_w^{-1}\boldsymbol{D}\boldsymbol{f}/\boldsymbol{g};\hat{\theta}^{(k)}\right\} \\
&= \frac{1}{\sigma_w^2}\boldsymbol{g}^t\boldsymbol{g} - \frac{2}{\sigma_w^2}\mathrm{tr}\left\{\boldsymbol{D}\,\mathrm{E}(\boldsymbol{f}\boldsymbol{g}^t/\boldsymbol{g};\hat{\theta}^{(k)})\right\} + \frac{1}{\sigma_w^2}\mathrm{tr}\left\{\boldsymbol{D}\,\mathrm{E}(\boldsymbol{f}\boldsymbol{f}^t/\boldsymbol{g};\hat{\theta}^{(k)})\,\boldsymbol{D}^t\right\}, \qquad \text{(C.2)}
\end{aligned}$$

where tr $(\boldsymbol{A})$ denotes the trace of the matrix $\boldsymbol{A}$. A similar derivation holds for the last term of (C.1):

$$\mathrm{E}\left\{\boldsymbol{f}^t(\boldsymbol{I}-\boldsymbol{A})^t\boldsymbol{Q}_v^{-1}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)}\right\} = \frac{1}{\sigma_v^2}\mathrm{tr}\left\{(\boldsymbol{I}-\boldsymbol{A})\,\mathrm{E}(\boldsymbol{f}\boldsymbol{f}^t/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)})\,(\boldsymbol{I}-\boldsymbol{A})^t\right\}. \tag{C.3}$$

Now we define the $MN \times MN$ conditional correlation matrices $\hat{\boldsymbol{R}}_{ff}^{(k)}$ and $\hat{\boldsymbol{R}}_{fg}^{(k)}$ as follows:

$$\begin{aligned}\hat{\boldsymbol{R}}_{ff}^{(k)} &= \mathrm{E}(\boldsymbol{f}\boldsymbol{f}^t/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\boldsymbol{f}\boldsymbol{f}^t\,p(\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)})\,df(1)\cdots df(MN)\\ &= \mathrm{Cov}(\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)}) + \mathrm{E}(\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)})\mathrm{E}(\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)})^t\\ &= \hat{\boldsymbol{V}}^{(k)} + \hat{\boldsymbol{f}}^{(k)}\hat{\boldsymbol{f}}^{(k)t}, \end{aligned} \tag{C.4}$$

$$\begin{aligned}\hat{\boldsymbol{R}}_{fg}^{(k)} &= \mathrm{E}(\boldsymbol{f}\boldsymbol{g}^t/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\boldsymbol{f}\boldsymbol{g}^t\,p(\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)})\,df(1)\cdots df(MN)\\ &= \mathrm{E}(\boldsymbol{f}/\boldsymbol{g};\hat{\boldsymbol{\theta}}^{(k)})\,\boldsymbol{g}^t = \hat{\boldsymbol{f}}^{(k)}\boldsymbol{g}^t. \end{aligned} \tag{C.5}$$

Here $\hat{\boldsymbol{f}}^{(k)}$ and $\hat{\boldsymbol{V}}^{(k)}$ are, respectively, the conditional mean and covariance of $\boldsymbol{f}$, given the observed image $\boldsymbol{g}$ and the estimate of $\boldsymbol{\theta}$ at the k-th iteration. By first substituting $\hat{\boldsymbol{R}}_{ff}^{(k)}$ and $\hat{\boldsymbol{R}}_{fg}^{(k)}$ into (C.2) and (C.3), next substituting (C.2) and (C.3) into (C.1), and finally multiplying the result by a factor 2 for notational simplicity we arrive at (7.14).

Samenvatting

Beelden worden gebruikt om belangrijke informatie vast te leggen of om deze zichtbaar te maken. Ten gevolge van onvolkomenheden in het beeldvormend proces (bijvoorbeeld camera, fotografische film) zijn de afbeeldingen die verkregen worden regelmatig een vervormde versie van de gewenste ideale afbeelding. In het algemeen kan er een groot aantal vervormingen zijn opgetreden in het beeldvormend proces, maar het blijkt dat de vervormingen die geïntroduceerd worden door onscherpte en ruis meestal overheersend zijn. De beeldidentificatie en -restauratie richt zich dan ook op het verminderen of zo mogelijk verwijderen van deze vervormingen. Dit heeft tot doel de (menselijke) interpretatie of de eventuele verdere verwerking (zoals beeldanalyse) van het beschikbare beeld te vergemakkelijken. In de beeldidentificatie wordt een schatting gemaakt van de eigenschappen van het niet-perfecte beeldvormende systeem (d.w.z. de onscherpte) vanuit het beschikbare onscherpe beeld. Verder worden enkele (statistische) eigenschappen bepaald van de in het beeld aanwezige ruis en van het originele (d.w.z. het scherpe) beeld. Op basis van deze gegevens wordt in de beeldrestauratie een schatting berekend van het originele beeld. Toepassingen van beeldidentificatie en -restauratie zijn bijvoorbeeld te vinden in de astronomie, medische beeldbewerking, en gerechtelijke onderzoek, en bij het analyseren van foto's van eenmalige gebeurtenissen,

Dit proefschrift concentreert zich op het gebruik van iteratieve methoden in de beeldidentificatie en -restauratie. Hierbij wordt verondersteld dat het onscherpe beeld vervormd is door lineaire plaats-invariante onscherpte en additieve witte ruis. In tegenstelling tot niet-iteratieve methoden kunnen iteratieve technieken het beeldrestauratieprobleem ook oplossen als dit geformuleerd is als een plaats-variante optimalisatie met randvoorwaarden. Er zal worden aangetoond dat de restauratieresultaten die op deze wijze verkregen worden, aanmerkelijk beter zijn dan die verkregen worden met meer conventionele technieken. Voor de beeldidentificatie zal

een iteratief algoritme geïntroduceerd worden dat bekend staat als het "expectation-maximization" (EM) algoritme, om een gecompliceerde niet-lineaire waarschijnlijkheids- ("likelihood") functie te optimaliseren. Vrijwel al het onderzoek waarover dit proefschrift rapporteert, is in de vorm van wetenschappelijke publicaties verschenen [14,55,57,59].

In Hoofdstuk 2 zal allereerst de modelvorming van het beeldvormend systeem worden besproken. Deze modellen vormen een essentiëel onderdeel van het proefschrift, aangezien vrijwel alle besproken beeldidentificatie en -restauratie methoden hier op gebaseerd zijn.

In Hoofdstuk 3 tot en met 5 komt een aantal aspecten van het gebruik van iteratieve methoden in de beeldrestauratie aan de orde. In Hoofdstuk 3 zal allereerst worden ingegaan op het slecht-geconditioneerd zijn van het beeldrestauratieprobleem. Dit betekent dat wanneer een inverse filter wordt gebruikt om een onscherp beeld te restaureren, dit altijd aanleiding zal geven tot een zeer sterke opslingering van de in het beeld aanwezige ruis. Vervolgens wordt een beknopte inleiding gegeven tot enkele welbekende restauratiemethoden, zoals de Wiener en Kalman filters, het kleinste-kwadraten filter met randvoorwaarde, en de methode van alternerende projecties op convexe deelruimten. Deze methoden hebben als gemeenschappelijke eigenschap dat ze vrijwel altijd gebruik maken van stochastische, algebraïsche of deterministische voorkennis over het te restaureren beeld om op deze wijze een middenweg te vinden tussen het opslingeren van de ruis en de betrouwbaarheid van de oplossing.

In Hoofdstuk 4 wordt een andere methode geïntroduceerd om rekening te houden met het feit dat het onscherpe beeld tevens ruis bevat. Hierbij wordt een iteratief schema dat in de limiet convergeert naar het inverse filter, voortijdig afgebroken. Het gevolg van dit afbreken is dat de onscherpte in het beeld slechts ten dele teniet wordt gedaan. Daar staat echter tegenover dat de ruis slechts weinig wordt opgeslingerd. Bijkomende voordelen van deze methode zijn dat er geen matrixinversies hoeven plaats te vinden, en dat het iteratieve algoritme redelijk eenvoudig is uit te breiden naar meer complexe schema's. Een aantal variaties op het voorgaande standaard iteratieve restauratiefilter wordt besproken. In eerste instantie zal een eenvoudig "steepest descent"-algoritme gebruikt worden om de iteraties te implementeren. Als alternatief hiervoor wordt ook een aantal efficiëntere implementaties besproken.

In Hoofdstuk 5 zal worden aangetoond dat als gevolg van het ruis-onderdrukkend karakter van geregulariseerde restauratiemethoden er een

storend artefact verschijnt in het opgescherpte beeld, namelijk ringvormige patronen. Het iteratieve schema dat wordt voorgesteld in dit hoofdstuk probeert deze "ringing" zo goed mogelijk te voorkomen door gebruik te maken van z.g. deterministische voorkennis, en door lokaal te bepalen of ruisopslingering wel of niet toegestaan is. In een aantal experimenten wordt geïllustreerd dat de restauratieresultaten van het voorgestelde iteratieve schema de voorkeur verdienen boven de meer conventionele restauratiefilters, zowel in numeriek als visueel opzicht.

Hoofdstuk 6, 7 en 8 gaan in op het beeldidentificatieprobleem zoals dat tegenwoordig geformuleerd wordt, waarbij specifiek gebruik wordt gemaakt van een "maximum likelihood" (ML) parameterschattingsmethode. In Hoofdstuk 6 wordt eerst een aantal reeds langer bekendstaande identificatie methoden besproken, dat echter maar in een klein aantal gevallen ook daadwerkelijk tot goede resultaten leidt. Vervolgens wordt het beeldidentificatieprobleem in de vorm van een ML-schattingsprobleem gegoten. Helaas blijkt dat het oplossen hiervan de optimalisatie vergt van een vrij complexe en niet-lineaire waarschijnlijkheidsfunctie. De methoden die tot dusver bekend zijn in de literatuur worden onder een gemeenschappelijk noemer geplaatst. Deze aanpak toont aan dat de verschillende algoritmen (en dus ook de resultaten hiervan) niet essentiëel verschillen, maar voortkomen uit een verschil in modelvorming (bijvoorbeeld: bevat het beeld wel of geen ruis), en/of overwegingen met betrekking tot algoritmische complexiteit of computerfaciliteiten (dit kan bijvoorbeeld leiden tot een voorkeur voor een recursieve, dan wel een matrix-vector formulering van het schattingsprobleem).

In Hoofdstuk 7 passen we het iteratieve EM-algoritme toe op het ML-beeldidentificatieprobleem. Deze aanpak leidt tot een zeer elegant algoritme dat tegelijkertijd het onscherpe beeld identificeert en restaureert. Vereist het oorspronkelijke ML-identificatieprobleem de optimalisatie van een niet-lineaire functie, bij het voorgestelde iteratieve identificatiealgoritme blijken slechts lineaire vergelijkingen opgelost te hoeven worden. De afleiding en de prestaties van bovengenoemde methode worden uitgebreid besproken.

Alhoewel de in Hoofdstuk 6 en 7 besproken algoritmen mathematisch adequaat geformuleerd zijn, blijkt er in praktische situaties een aantal problemen te onstaan door numerieke onnauwkeurigheden en door de afhankelijkheid tussen de verkregen oplossing en de initiële parameterschatting. Daarom wordt in Hoofdstuk 8 een tweetal praktisch georienteerde

identificatiestrategieën voorgesteld, die echter wel geënt zijn op het in Hoofdstuk 7 besproken identificatiealgoritme. De eerste methode maakt gebruik van voorkennis over de structuur van de onscherpte en het beeldmodel. De tweede methode maakt gebruik van resolutiepiramides om tot een hiërarchische schattingsprocedure te komen voor de onscherpte in het beeld. Er worden experimentele resultaten op zowel synthetische als fotografische bewegings- en defocusseringsonscherpte gegeven. Het hoofdstuk eindigt met een korte bespreking van de huidige stand van zaken rond de beeldidentificatie en schetst een perspectief voor de toekomst.

Acknowledgements

On the completion of this thesis I would like to express my appreciation to all those who have contributed in one or another way.

The work reported here has been strongly influenced and has been improved by the many lengthy and beneficial discussions I had with research collegues. I have found the contacts with research groups at universities abroad very valuable and profitable.

First I would like to thank Prof. Russell M. Mersereau from the Georgia Institute of Technology (Atlanta, USA), who was one of the first to recognize the importance of iterative methods for restoration purposes. As a result of a 3 month's visit of Prof. Jan Biemond to Georgia Tech. in 1983, cooperative research was initiated. The visit of Prof. Mersereau to Delft in the summer of 1986 first resulted in a joint conference paper, and later on it led to a joint tutorial paper in the Proceedings of the IEEE. I would like to acknowledge the effort Prof. Mersereau put into this paper, which helped me particularly in writing Chapters 2 to 4.

Further I would like to thank Prof. Aggelos K. Katsaggelos, who graduated from Georgia Tech. and is currently with Northwestern University (Evanston, USA), for the joint research we performed on iterative image identification and restoration during the exchange of visits.

I express much gratitude towards Prof. Howard Kaufman and Prof. John W. Woods of Rensselaer Polytechnic Institute (Troy, USA) for putting me on the track of the EM-algorithm. With them and with Dr. Denise L. Angwin I had many beneficial discussions about image restoration in general and about image identification methods in particular. I would also like to thank them for the pleasant social activities during my visits to Troy.

Special thanks are also due to Prof. A. Murat Tekalp of the University of Rochester (Rochester, USA). We had valuable and enlightening discussions about the (dis)advantages of iterative methods compared to recursive filtering methods. As a direct result of these discussions, a joint

paper was published in the Journal of Optical Engineering. The greater part of Chapter 6 of this thesis is based on this paper.

Of course, this thesis could not have been completed without a brisk and pleasant working environment. I greatly appreciate the support of my colleagues and the efforts of the students who have been working with me over the last five years. In particular I would like to thank Peter Westerink and Hans Driessen for their companionship both in Delft and on various occasions abroad.

I am obliged to Mr. J. van der Krogt for printing the photographs which are shown in this thesis, and to Mrs. J.B. Zaat-Jones for correcting the English text.

Finally, I am highly indebted to my wife Marleen for her support in general and in particular for the conscientiousness with which she read and corrected the manuscript of this thesis. Over the four months it took me to write this thesis I somewhat neglected our family life and our social contacts. Although not entirely without murmur, she accepted that in order to finish this work I needed to temporarily reverse my priorities.

Curriculum Vitae

Reginald Leendert Lagendijk was born in Leyden, the Netherlands, in 1962. In 1980 he obtained his Atheneum-B diploma from the Fioretti College in Lisse. Subsequently, he enrolled in the Department of Electrical Engineering of the Delft University of Technology, the Netherlands, from which he received his M.Sc. degree (cum laude) in 1985.

From 1985 to 1987 he worked as a Ph.D. student in the Laboratory for Information Theory of the Delft University of Technology. In 1987 he joined the faculty of the same group.

He is the (co-)author of more than 20 publications in international scientific journals and conference proceedings in the area of image identification and restoration, and image sequence coding. Since 1987 he has been involved in teaching regular courses in (multidimensional) stochastic signal processing and digital signal encoding. He is involved in co-supervising graduate students in the area of image (sequence) coding and restoration, in cooperation with Prof. Jan Biemond.