Iterative Methods for Image Deblurring

JAN BIEMOND, SENIOR MEMBER, IEEE, REGINALD L. LAGENDIJK, STUDENT MEMBER, IEEE, AND RUSSELL M. MERSEREAU, FELLOW, IEEE

This tutorial paper discusses the use of iterative restoration algorithms for the removal of linear blurs from photographic images which may also be assumed to be degraded by pointwise nonlinearities such as film saturation and additive noise. Iterative algorithms are particularly attractive for this application because they allow for the incorporation of various types of prior knowledge about the class of feasible solutions, because they can be used to remove nonstationary blurs, and because they are fairly robust with respect to errors in the approximation of the blurring operator. Special attention is given to the problem of convergence of the algorithms, and classical solutions such as inverse filters, Wiener filters, and constrained least-squares filters are shown to be limiting solutions of variations of the iterations. Regularization is introduced as a means for preventing the excessive noise magnification that is typically associated with ill-conditioned inverse problems such as the deblurring problem, and it is shown that noise effects can be minimized by terminating the algorithms after a finite number of iterations. The role and choice of constraints on the class of feasible solutions are also discussed. Ringing artifacts are common with most image restoration methods. It is shown that these artifacts can be significantly reduced both by using constraints and also by making the algorithms spatially adaptive. Some variations on the basic iterations that accelerate the rate of convergence are discussed and numerous examples are presented.

I. Introduction

Images are produced to record or display useful information, but the process of image formation and recording is imperfect. The recorded image invariably represents a degraded version of the original scene. Three major types of degradations can occur—blurring, pointwise nonlinearities, and noise. Blurring is a form of bandwidth reduction of the image owing to the image formation process. It can be caused by relative motion between the camera and the original scene, or by an optical system that is out of focus. When aerial photographs are produced for remote sensing, 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. Electron micrographs are corrupted by the spherical aberrations of the electron lenses. The second type of image degradation is a pointwise nonlinearity introduced by the nonlinear response of the recording medium. An important example of such a sensor nonlinearity is the sensitivity of photographic film. The density of silver grains on developed film varies approximately logarithmically with the incident light intensity with saturation in both the black and white regions. The final source of degradation in recorded imagery is noise. This corrupts both the image formation and recording processes. It can be introduced by the transmission medium (such as a noisy channel), the recording medium (such as filmgrain noise), measurement errors, and quantization of the data for digital storage.

The field of image restoration is concerned with the reconstruction or estimation of an uncorrupted image from a distorted and noisy one. It is important in fields such as astronomy, where resolution and recording limitations are severe, for enhancing historically important photographs, and for analyzing images of unique events such as medical images, satellite photographs, and the result of scientific experiments. In recent years the commercial photographic industry has also shown an interest in consumer applications of image restoration.

This paper discusses an iterative approach to the problem of restoration of blurred images. This is a special case of the more general problem of iterative signal restoration, which has had a very active recent history [1]-[21]. It has been consistently demonstrated that these iterative procedures can be especially powerful when prior knowledge about the underlying signal or image is available in the form of constraints on the allowable restorations, when the blurring function is only approximately known, and when the user elects to vary the degree of blur and noise removal with the local information content in the image. This tutorial paper discusses many of these recent developments and shows that these iterative algorithms are particularly well suited to the problem of image restoration.

This paper is arranged into several sections. Section II discusses mathematical models for images and blur operators. Motion blur is introduced as an example of a stationary blur, and out-of-focus (defocussing) blur is presented as an example of a nonstationary blur. Stationary approximations for defocussing blurs are also introduced. Procedures for deblurring require complete knowledge of the blurring function. As this is rarely available, Section III

Manuscript received June 3, 1988; revised October 19, 1989. R. M. Mersereau was partially supported by the Joint Services Electronics Program under Contract DAAL-03-87-K-0059.

J. Biemond and R. L. Lagendijk are with the Delft University of Technology, Dept. of Electrical Engineering, 2600 GA Delft, The Netherlands.

R. M. Mersereau is with the Georgia Institute of Technology, School of Engineering, Atlanta, GA 30332, U.S.A.

The authors are listed alphabetically.

IEEE Log Number 8933935.

0018-9219/90/0500-0856$01.00 © 1990 IEEE

reviews both cepstral and spatial domain procedures for the estimation of the blurring operator from the blurred image itself.

The image deblurring problem is a classical example of an ill-conditioned problem; its solution is highly sensitive to measurement errors. Many of the early solutions were concerned with the problem of noise sensitivity. Some of these classical solutions are discussed in Section IV. These include inverse filters, least squares or Wiener filters [22], Kalman filters [23], [24], and constrained least squares solutions [25]–[27]. This section continues by introducing the basic iterative deblurring algorithm forming the basis for most of the algorithms discussed in the remainder of the paper. Variations on this iteration are presented which asymptotically produce the inverse and constrained least squares solutions as the number of iterations is increased. The issue of convergence of the iterations is discussed carefully and it is shown analytically that terminating these iterations prior to convergence is one important method for preventing noise magnification.

Section V of the paper introduces the concept of regularization, a formalism by which the ill-conditioned deblurring problem is converted into a well-conditioned problem with less sensitivity to measurement noise. Both iterative and noniterative regularized restoration procedures are presented and several examples are given which clearly demonstrate the power of the approach.

One of the strong motivations for using iterative procedures is the fact that they provide a mechanism for limiting the set of feasible solutions to the inversion problem by requiring that the restorations lie in a closed convex space. Section VI is concerned with the problem of constrained restoration. The earlier iterations are modified to allow for constraints and several examples are presented which demonstrate how the tightness of the constraints can affect the resulting restorations.

A common artifact associated with any of these restorations is ringing, a Gibbs-like oscillation introduced in the vicinity of abrupt changes in intensity. Methods for reducing noise magnification, such as regularization, tend to make this problem worse. The imposition of constraints can greatly reduce ringing in some cases. It is also shown that making the iterations spatially adaptive is even more effective. In Section VII we show that both techniques can be applied together.

In Section VIII, the iteration is extended to include the removal of the pointwise nonlinearities introduced in the recording process. Finally, Section IX is concerned with procedures for increasing the rate of convergence of the iterative algorithms. Two different procedures are introduced for this purpose—one based on the method of conjugate gradients from optimization theory, and one that replaces the iterations by a higher-order iteration whose convergence is accelerated.

II. Models for Blurred Image Formation

A. 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. If the image formation process is linear, the recorded image can be modeled as the output of the system

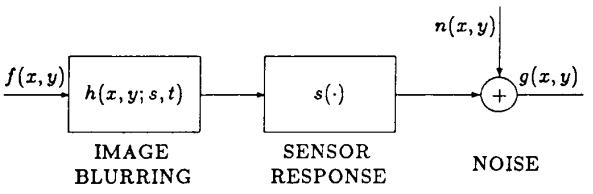


Fig. 1. Model for the processes of image formation and recording.

shown in Fig. 1, which is given mathematically by

$$g(x, y) = s\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y; s, t)\, f(s, t)\, ds\, dt\right\} + n(x, y). \tag{1}$$

Here and throughout this paper $g(x, y)$ will be used to denote the recorded image, and $f(x, y)$ will be used to denote the ideal image, which is a 2-D mapping of the 3-D input scene. The goal of the restoration is to produce a good estimate of f. Here $h(x, y; s, t)$ is the 2-D impulse response (point-spread function) of the linear blurring system and $s\{\cdot\}$ is the sensor nonlinearity which has been modeled as a point operator. The noise contribution is shown as an additive random process which is statistically uncorrelated with the image. This is a simplification because noises such as film-grain noise and the noise caused by photon statistics, which often corrupt images, are not uncorrelated with the input. This simplification nonetheless leads to reasonable and useful results.

If the impulse response is stationary across the image and object fields, it becomes a function of only the argument differences $x - s$ and $y - t$. In this case the superposition integral in (1) becomes a more familiar convolution integral

$$g(x, y) = s\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x - s, y - t)\, f(s, t)\, ds\, dt\right\} + n(x, y) \tag{2}$$

$$g(x, y) = s\{h(x, y) * f(x, y)\} + n(x, y) \tag{3}$$

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

In a discrete implementation the functions with continuous arguments f, g, h, and n are replaced by arrays of samples taken on $N \times N$ 2-D rectangular lattices of equi-spaced samples. The sampled arrays are related by

$$g(i, j) = s\left\{\sum_{\forall(k,l)} h(i, j; k, l)\, f(k, l)\right\} + n(i, j) \qquad 0 \le i, j \le N - 1. \tag{4}$$

For the spatially invariant (stationary) system, the convolution integral (2) becomes a convolution sum

$$g(i, j) = s\left\{\sum_{\forall(k,l)} h(i - k, j - l)\, f(k, l)\right\} + n(i, j) \tag{5}$$

$$g(i, j) = s\{h(i, j) * f(i, j)\} + n(i, j) \tag{6}$$

where the asterisk $(*)$ is now used to denote a discrete convolution. Often the sensor nonlinearity is conveniently neglected (or linearized) to justify the use of a linear restoration filter. When this nonlinearity is ignored, (6) reduces to the linear convolution model

$$g(i, j) = h(i, j) * f(i, j) + n(i, j) \tag{7}$$

for which discrete Fourier transforms (see Appendix) can be used to yield the frequency domain model

$$G(m, n) = H(m, n)\, F(m, n) + N(m, n). \tag{8}$$

Here $H(m, n)$ represents samples of the frequency response of the blurring system and m and n are the discrete horizontal and vertical spatial frequency variables. Because imperfections in an image formation system normally act as passive operations on the image data, all energy arising from the point (k, l) should be preserved. Thus, $h(i, j; k, l)$ is constrained to satisfy

$$\sum_{(i,j)\in S_1} h(i, j; k, l) = 1, \qquad \forall (k, l) \tag{9}$$

where S_1 is the support of the PSF.

For further simplification it is also convenient to use the matrix-vector notation

$$g = Hf + n \tag{10}$$

where f, g, and n are lexicographically ordered vectors [28] of size $N^2 \times 1$, and H is the blurring operator of size $N^2 \times N^2$ (see Appendix).

These expressions were presented for monochromatic (black and white) images. A color image is usually described by a vector with three components corresponding to the tristimulus values red, blue, and green, each of which is itself a monochromatic image.

B. Image Models

Certain linear image restoration techniques including Wiener filters [22] and Kalman filters [23], [24] make use of *a priori* statistical knowledge of the original (undistorted) image. This takes the form of a power density spectrum for the Wiener filter and the form of a stochastic difference equation for the Kalman filter. These quantities can be derived by using an autoregressive model for the image. A large class of real-world images can be modeled as the following 2-D autoregressive process of low order:

$$f(i, j) = \sum_{(p,q)\in W_1} a(p, q)\, f(i - p, j - q) + u(i, j), \qquad \forall (i, j) \tag{11}$$

Here $u(i, j)$ can be viewed as either an innovation process or as the error in approximating $f(i, j)$ using a linear combination of neighboring sample values contained in a neighborhood W_1. Different models result for different choices of the set W_1. Some common choices for W_1 are

$$W_1 = \begin{cases} \{(p, q): (p \geq 0, q < 0) \cup (p > 0, q \geq 0)\}, \\ \qquad \text{Nonsymmetric halfplane causal models} \\ \{(p, q): (p > 0, \forall q) \cup (p = 0, q \neq 0)\}, \\ \qquad \text{semicausal models} \\ \{(p, q): (p, q) \neq (0, 0)\}, \qquad \text{noncausal models.} \end{cases} \tag{12}$$

These three neighborhoods are illustrated in Fig. 2. A comprehensive survey of these three image models has been given by Jain [29]. Other relevant literature on image modeling can be found in [30]–[35].

Computational considerations usually restrict the nonzero values of the model parameters $\{a(p, q)\}$ to a finite window W, called the *prediction window*, which is a subset of W_1.

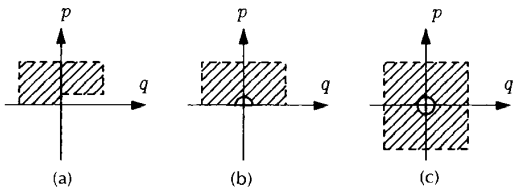


Fig. 2. Model support corresponding to (a) nonsymmetric halfplane image model; (b) semicausal image model; (c) noncausal image model.

C. Blur Models

Motion Blur: Many types of motion blur [36] can be distinguished, all of which are caused by relative motion between the camera and the object. This can be in the form of a translation, a rotation, a sudden change of scale, or to some combination of these. Here only the important case of a translation will be considered. When the object translates at a constant horizontal velocity V during the exposure interval $[0, T]$, the distortion is one-dimensional and its point-spread function is given by [36]

$$h(x, y; s, t) = h(x - s) = \begin{cases} \frac{1}{VT}\,\delta(y - t) & 0 \leq x - s \leq VT \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The discrete equivalent point-spread function makes use of the blurring distance L, which is the number of additional points in the image resulting from a single point in the original scene.

$$h(i, j; k, l) = h(i - k) = \begin{cases} \frac{1}{L + 1} & 0 \leq i - k \leq L, \quad j = l \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

The frequency response corresponding to this blur is given by

$$H(m, n) = \frac{1}{L + 1}\, e^{-j(L\pi/N)m}\, \frac{\sin\left(\pi(L + 1)m/N\right)}{\sin\left(\pi m/N\right)} \quad \forall n \tag{15}$$

These impulse and frequency responses are seen in Fig. 3. In that figure it is readily seen that the frequency response is zero on lines parallel to the n-axis with an interline spacing of $N/(L + 1)$. If the linear motion is in some other direction, the blurring frequency response will have the same form but will be rotated in frequency. The presence of these parallel zeros in the frequency domain, which are also present in the blurred image (in the absence of noise), not only indicates the presence of a linear motion blur, but also indicates the direction of motion, and the blurring distance.

Out-of-Focus Blur: When a three-dimensional scene is imaged by a camera onto a two-dimensional image field, some parts of the scene are in focus while other parts are not. The degree of defocus depends upon the effective lens diameter and the distance between the object and the cam-

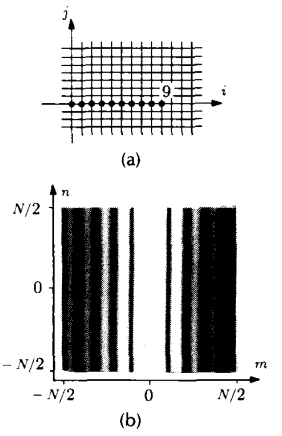


Fig. 3. (a) The impulse response and (b) magnitude of the frequency response of a horizontal linear blur of $L = 9$.

era. To describe this inherently spatially varying blur, consider a camera consisting of a lens and an aperture that limits the lens diameter. When the film is located at the focal plane of the lens, objects infinitely far away are in perfect focus in the resulting image. As the lens is moved relative to the image plane, objects at other distances are brought into focus. In Fig. 4, an object at distance D is focussed

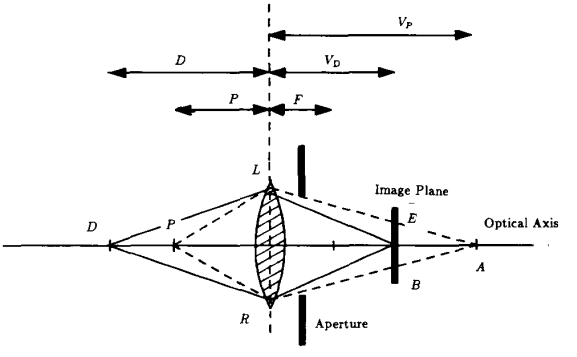


Fig. 4. Geometry of an imaging system.

sharply. More distant object points come into focus in front of the imaging plane, and converging rays from nearer objects are intercepted by the film before they reach a sharp focus. If the aperture is circular, the image of any point source is a small disk, known as the *circle of confusion* (COC).

The diameter of the circle of confusion is a function of the distance P of the observed point [37]. Let V_D and V_P be the image distances corresponding to objects at distances D (in focus) and P (out-of-focus) respectively. The point at V_D lies in the image plane, but the point at P projects onto a circle as it converges a distance $|V_P - V_D|$ away. From simple geometry (see Fig. 4), it follows that

$$\frac{\overline{LR}}{V_P} = \frac{\overline{EB}}{|V_P - V_D|} \tag{16}$$

where $\overline{EB}$ is the diameter of the circle of confusion and $\overline{LR}$ is the effective lens diameter, defined as the focal length divided by the aperture number (*f*-stop) n. From the lens law

$$\frac{1}{P} + \frac{1}{V_P} = \frac{1}{F} \tag{17}$$

$$\frac{1}{D} + \frac{1}{V_D} = \frac{1}{F} \tag{18}$$

where F is the focal length of the lens, it follows that the diameter of the circle of confusion $C(P)$ can be written as

$$C(P) = \begin{cases} \dfrac{FD}{n(D-F)} - \dfrac{F^2D}{nP(D-F)} - \dfrac{F}{n}, & \text{for } D < P < \infty \\ \dfrac{F}{n} - \dfrac{FD}{n(D-F)} + \dfrac{F^2D}{nP(D-F)}, & \text{for } F < P < D. \end{cases} \tag{19}$$

This function is sketched in Fig. 5. As $P \to D$ the plane comes into focus and the diameter of the circle of confusion

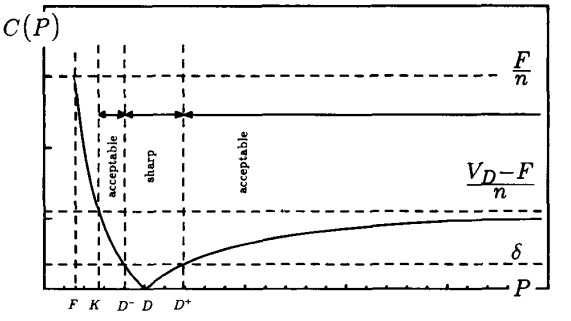


Fig. 5. Diameter of the circle of confusion.

approaches zero. The diameter of the COC varies asymmetrically with P.

In practice an object can be said to be in focus whenever the diameter of its circle of confusion is less than δ, the resolution limit of the film. Appealing to Fig. 5, this means that objects at distances between D^- and D^+ are in focus. Blur will be visible only when the diameter of the circle of confusion exceeds the resolution limit. The term *depth of field* refers to the range of object distances $[D^-, D^+]$ that fall within the resolution limit. For increasing aperture number n, that is, a decreasing effective lens diameter, a greater depth of field can be realized.

Points infinitely far away have a limiting COC diameter given by $(V_D - F)/n$. Points at the distance $K = D/2$, known as the *critical distance*, also have this same COC diameter. If $(V_D - F)/n < \delta$, then all points in the range $[K, \infty)$ will be in focus.

To obtain a complete model for defocussing, we need to know the intensity distribution within the circle of confusion caused by a point object. From geometrical optics it follows that this intensity distribution should be roughly constant and nonzero within the circle of confusion and zero elsewhere [38]. This corresponds to the point-spread function

$$h(x, y) = \begin{cases} \dfrac{1}{\pi r^2} & x^2 + y^2 \le r^2 \\ 0 & \text{elsewhere.} \end{cases} \tag{20}$$

where r is the radius of the circle of confusion.

The frequency response (optical transfer function (OTF)) corresponding to this model for the blur is given by

$$H(m, n) = \frac{J_1(r\rho)}{r\rho}, \qquad \rho^2 = m^2 + n^2 \tag{21}$$

where J_1 is the first-order Bessel function. This frequency response is shown in Fig. 6. A more accurate calculation

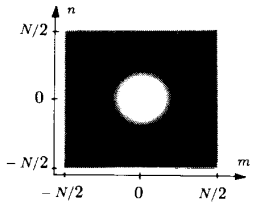


Fig. 6. Simplified frequency response corresponding to out-of-focus blur.

would involve the effect of diffraction [39]. It can be shown [38], [40] that when the degree of defocussing is large, the geometrical OTF closely approximates the diffraction OTF for low spatial frequencies.

A discrete equivalent point-spread function corresponding to (20) can be obtained by associating with each location (i, j) in the discrete plane the rectangular pixel shape shown in Fig. 7. The value of the discrete point-spread function

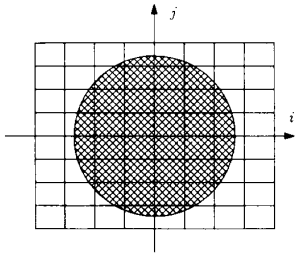


Fig. 7. Discrete approximation to an out-of-focus point-spread function.

(PSF) is then equal to the value of the continuous point-spread function weighted by the fraction of the pixel covered. The discrete PSF is constant for small radii, zero for large ones, and assumes intermediate values for radii close to the radius of the circle of confusion.

If the camera misadjustment and object position are known exactly, we can calculate the spatially varying point-spread function exactly. However, in most practical situations we will not have this much prior knowledge. The only assumption often to be made is that the image is unsharp because of defocussing. Then the degree of the blur should be estimated at each pixel from the blurred image itself.

A more accurate model reveals that the point-spread function corresponding to an out-of-focus blur is also wavelength dependent owing to diffraction and interference phenomena, and that the radius of the COC is also wavelength dependent because of the refractive index of the lens. This is known as chromatic aberration [41]. Thus, the three color components red, green, and blue (R, G, and B) of a color image, each originating from a different frequency band of the image scene, would generally have different point-spread functions.

III. Blur Identification

The first step in restoring a degraded image is the identification of the type of degradation. If the camera misadjustment, object distances, object motion, and camera motion are known exactly, we can calculate the point-spread function for the three primary color components. In practice the degradation is rarely known exactly, and the blur must be identified from the blurred image itself. In this situation it is helpful to have a parametric blur description such as that in (14) or (20). For linear motion blur, as given in (14) it is only necessary to estimate the direction of blur and the blurring distance. With the simplified model for an out-of-focus blur in (20) it is only necessary to estimate the radius of the circle of confusion. Because both of these blurs have an oscillatory frequency response with a characteristic zero-crossing pattern, it is advantageous to identify them in the spectral or cepstral domain under the assumption that the blur is locally space invariant. If this assumption does not hold, the blur must be identified in the spatial domain [42], [43].

A. Blur Identification in the Spectral/Cepstral Domain

The following technique for identifying the power spectrum of the blurring function was developed by Stockham, Cannon, and Ingebretsen [44]. As before, let $g(x, y)$, and $f(x, y)$ denote the blurred and original images, respectively. When the noise contribution is neglected, the power density spectra of the two images are then related by

$$|G(m, n)|^2 = |H(m, n)|^2 |F(m, n)|^2. \tag{22}$$

If the images g and f are divided into nonoverlapping subimages $\{g_k(x, y), f_k(x, y), k = 1, 2, \ldots, K\}$, the power density spectra of these subimages will approximately satisfy a relation similar to (22)

$$|G_k(m, n)|^2 \approx |H(m, n)|^2 |F_k(m, n)|^2. \tag{23}$$

This relationship is only approximately true for the subimages, because the convolution of $h(x, y)$ with $f_k(x, y)$ will extend beyond the boundaries of $g_k(x, y)$. If these boundary effects are negligible, however, which is the case if the subimages are large compared to the extent of the blurring function, then the approximation in (23) is a good one. Taking logarithms of both sides of (23) and adding the results for each of the subimages gives

$$\frac{1}{K}\sum_{k=1}^{K} \log |G_k(m, n)|^2 \approx \frac{1}{K}\sum_{k=1}^{K} \log |F_k(m, n)|^2 + \log |H(m, n)|^2. \tag{24}$$

The quantity on the left can be evaluated from the blurred image. The first sum on the right side of this equation, however, is unknown. Stockham *et al.* [44] argued that it could be approximated by an average power spectrum evaluated over a wide variety of images. This estimate can then be subtracted from the expression on the left-hand side to yield an approximation to the magnitude response of the blurring function.

For linear motion blur, such an estimate is often sufficient

to estimate the zero patterns in the frequency response from which one can estimate the direction of motion and the blurring distance. For an out-of-focus blur, the frequency zero patterns can be used to estimate the radius of the circle of confusion.

An alternative to the above for identifying *linear motion* blur involves the computation of the two-dimensional *cepstrum* of $g(x, y)$ [45]. The (power) cepstrum is the inverse Fourier transform of the logarithm of the magnitude of $G(m, n)$. Thus

$$\hat{g}(x, y) = \mathfrak{F}^{-1}\{\log |G(m, n)|\} \tag{25}$$

where $\mathfrak{F}^{-1}$ is the inverse Fourier transform operator. One of the important properties of the cepstrum is that if two signals are convolved, their cepstra add. Thus, if the noise is again neglected

$$\hat{g}(x, y) = \hat{h}(x, y) + \hat{f}(x, y). \tag{26}$$

For horizontal, linear motion the frequency response of the blur can be expressed in terms of the Fourier variables (m, n) as

$$H(m, n) = \frac{1}{L + 1} e^{-j(L\pi/N)m} \frac{\sin(\pi(L + 1)m/N)}{\sin(\pi m/N)} \quad \forall n. \tag{27}$$

This response has zeros at integer multiples of $N/L + 1$. As a result, $\hat{h}$ has a large negative spike at a distance L from the origin. This spike is a prominent feature in $\hat{g}(x, y)$. Its presence indicates the presence of motion blur and its position indicates the direction and extent. As an example, consider the blurred image of a train shown in Fig. 8. This image demonstrates horizontal motion blur. The rowwise summed log spectrum, formed from 32 rows taken at the level of the centers of the cars, is shown in Fig. 8(b), and the cepstrum is shown in Fig. 8(c). The cepstrum displays a prominent spike at $L = 7$ samples.

B. Blur Estimation in the Spatial Domain

The blur estimation techniques described in the previous section relied on a parametric description of the blur, for which the missing parameters were estimated using either the spectrum or cepstrum of the blurred image. These deterministic techniques can only be used to estimate a certain class of frequency responses—those having zeros on the unit bi-circle. Not all important blurs have such characteristics. For example, a Gaussian blur, which is commonly used to model the degradation introduced in an x-ray recording system, could not be identified using these techniques. This section will present a spatial domain procedure for simultaneously estimating both the blurring operator and the image model coefficients without assuming a specific functional form for the blur. These estimated model and blur coefficients can then be used for the subsequent restoration of noisy blurred images. An additional advantage of the spatial domain technique is its ability to track slowly varying image statistics and spatially varying blurs.

The technique begins with the assumption that the undistorted original image can be described by the autoregressive model (11) with causal support (12). That is,

$$f(i, j) = \sum_{(p,q)\in W} a(p, q)\, f(i - p, j - q) + u(i, j) \tag{28}$$

and that the noisy, blurred image with noncausal support can be described by

$$g(i, j) = \sum_{(k,l)\in S_1} h(k, l)\, f(i - k, j - l) + n(i, j). \tag{29}$$

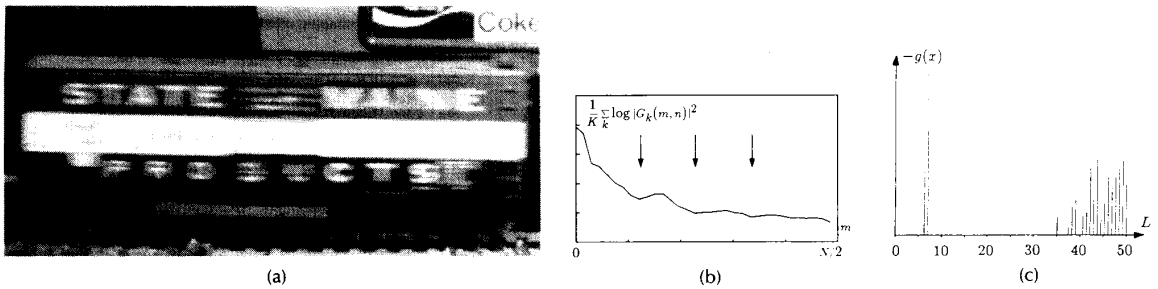
(Notice that the point-wise nonlinearity from (1) has been omitted.) This so-called state-space pair is not suitable for the identification of the unknown parameters in the model, because the undistorted image $f(i, j)$ is not available. By eliminating $f(i, j)$ from these equations and neglecting the effect of the observation noise on the estimation of the coefficients [42], we arrive at the equation

$$g(i, j) = \sum_{(p,q)\in W} a(p, q)\, g(i - p, j - q) + \sum_{(k,l)\in S_1} h(k, l)\, u(i - k, j - l). \tag{30}$$

This represents a 2-D ARMA model for the observations, where the image model coefficients form the autoregressive (AR) portion of the model, and the blur coefficients $h(k, l)$ form the moving average (MA) part.

In [42], Tekalp *et al.* derive conditional maximum likelihood estimates of these unknown coefficients in the absence of observation noise. Biemond *et al.* [43] followed the same procedure, but first decomposed the 2-D ARMA model rowwise into $N/2 + 1$ complex 1-D ARMA column sequences by using the DFT and an assumed semicausal model support W. This gave

$$A_0(n)\, G(i, n) = -\sum_{p=1}^{P} A_p(n)\, G(i - p, n) + \sum_{k=0}^{2K} H_k(n)\, U(i - k, n), \quad v = 0, 1, \ldots, N - 1, \tag{31}$$



(a) (b) (c)

Fig. 8. (a) A natural image displaying motion blur. (b) The log spectrum computed from 32 rows in the center of the train. (c) The cepstrum displaying a prominent spike at 7 samples.

where $A_p(n)$ and $H_k(n)$ are defined as the 1-D DFTs of the defining sequence $a_p(i)$ and $h_k(i)$. These are given by

$$a_p(i) = \{-a(p, -P), \ldots, -a(p, 0), \ldots, -a(p, P)\} \quad (32)$$

$$h_k(i) = \{h(k, -L), \ldots, h(k, 0), \ldots, h(k, L)\}. \quad (33)$$

Here capitals denote transform domain quantities and n denotes the discrete horizontal frequency variable. With this decomposition the parameter estimation can be performed in parallel using simple 1-D recursive estimation techniques [46]. An estimation procedure that offers the potential of being relatively fast, while still estimating the MA portion of the model accurately, uses a high-order AR approximation as an intermediate step [43], [47].

As an example of this identification procedure, consider the blurred cameraman image in Fig. 10, which was obtained by a computer simulated blurring of the image in Fig. 9 with

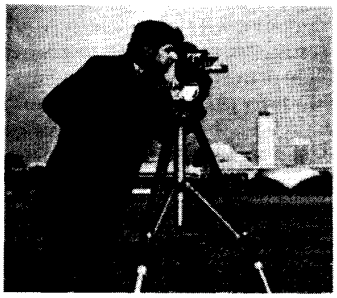


Fig. 9. Original cameraman image with 256 × 256 pixels quantized to 8 bits per pixel.

Fig. 10. Motion-blurred cameraman image with noise added at an SNR of 50 dB.

motion ($L = 8$) and noise. The image model computed from the original image is given by [43]:

$$a(p, q) = \begin{bmatrix} a(2, -1) & a(2, 0) & a(2, 1) \\ a(1, -1) & a(1, 0) & a(1, 1) \\ a(0, -1) & a(0, 0) & a(0, 1) \end{bmatrix} = \begin{bmatrix} -0.0614 & 0.1740 & -0.0614 \\ 0.2440 & -0.7018 & 0.2440 \\ -0.4605 & 1.0000 & -0.4605 \end{bmatrix}. \quad (34)$$

The following estimates were calculated for the image model and blur parameters using the blurred image in Fig. 10:

$$\hat{a}(p, q) = \begin{bmatrix} -0.0497 & 0.1654 & -0.0497 \\ 0.1951 & -0.6896 & 0.1951 \\ -0.4169 & 1.0000 & -0.4169 \end{bmatrix} \quad (35)$$

$$\hat{h}(0, j) = [h(0, -4), \cdots, h(0, 0), \cdots, h(0, 4)] = [0.1110, 0.1109, 0.1092, 0.1124, 0.1131, 0.1124, 0.1092, 0.1109, 0.1110]. \quad (36)$$

IV. The Classical and Basic Iterative Solutions

Sections II and III addressed the problem of modeling and estimating the blurring function. This section begins by assuming that these are satisfactorily known. It looks at the problem of blur removal using a linear restoration filter, neglecting any pointwise nonlinearities that might be corrupting the image. In the space-varying case the original and blurred images are related by

$$g(i, j) = \sum_{\forall(k,l)} h(i, j; k, l)\, f(k, l) + n(i, j). \quad (37)$$

and in the space-invariant case they are related by

$$g(i, j) = \sum_{\forall(k,l)} h(i - k, j - l)\, f(k, l) + n(i, j). \quad (38)$$

This section will compare a number of methods for estimating f from g.

A. The Inverse Filter Solution

An inverse filter is a linear filter whose point-spread function $h_{inv}(i, j; k, l)$ is the inverse of the blurring function $h(i, j; k, l)$ in the sense that

$$\sum_{\forall(p,q)} h_{inv}(i, j; p, q)\, h(p, q; k, l) = \delta(i - k, j - l) \quad (39)$$

where

$$\delta(i, j) = \begin{cases} 1 & \text{if } i = j = 0 \\ 0 & \text{elsewhere.} \end{cases} \quad (40)$$

These filters are virtually impossible to design in the spatially varying case. Therefore, in the remainder of this section only the space-invariant case will be considered.

The space-invariant inverse filter $h_{inv}(i, j)$ is the convolutional inverse of $h(i, j)$. Thus,

$$h_{inv}(i, j) * h(i, j) = \delta(i, j) \quad (41)$$

which can be expressed in the discrete frequency domain as

$$H(m, n)\, H_{inv}(m, n) = 1. \quad (42)$$

If the blurred image is passed through the inverse filter, the discrete Fourier transform of the output is given by

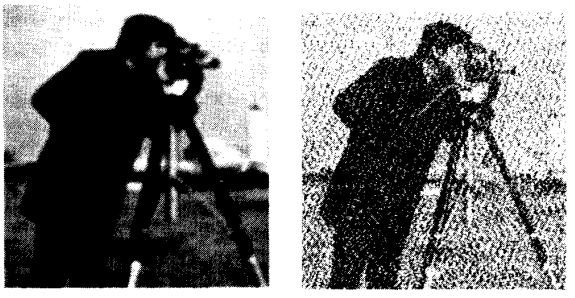
$$\begin{aligned} \hat{F}(m, n) &= H_{inv}(m, n)\, G(m, n) \\ &= H_{inv}(m, n)[H(m, n)\, F(m, n) + N(m, n)] \\ &= F(m, n) + H_{inv}(m, n)\, N(m, n). \end{aligned} \quad (43)$$

The restored image is thus equal to the desired image plus the inverse filtered noise.

Unfortunately, there are several problems with this approach. First, the inverse filter may not exist. Such is the

case if $H(m, n)$ comes from an ideal lowpass filter, or if $H(m, n)$ is zero at selected frequencies. Recall that this is the case with both linear motion blur and shift-invariant approximations to out-of-focus blur. Second, even when the blurring frequency response does not actually go to zero, there are usually problems caused by excessive noise amplification at high frequencies. This is because the power spectrum of the blurred image is typically highest at low frequencies and rolls off significantly for higher ones. The spectrum of the additive noise, on the other hand, typically contains relatively more high frequency components. Thus, at high frequencies, $\hat{F}(m, n)$ is dominated by the inverse filtered noise, which yields useless solutions. The inverse filter may also be difficult to realize, and when the blurring function is known only approximately, the resulting uncertainty in $H_{inv}(m, n)$ may be intolerable. With hindsight it can also be noted that the inverse filter suffers because it makes no use of the properties of f.

Figure 11 shows a blurred cameraman image and the corresponding inverse filtered restoration. The distortion here



(a) (b)

Fig. 11. (a) Image blurred by defocusing blur ($r = 3$) at an SNR of 40 dB. (b) Restoration by inverse filtering. SNR improvement = −16.5 dB.

was a defocusing blur with a COC radius of 3. The blurred image was computed from the unblurred original in Fig. 9 and Gaussian noise was added to the result at a signal-to-noise ratio (SNR) of 40 dB. Here the signal-to-noise ratio is calculated as

$$\text{SNR} = 10 \log_{10} \left(\frac{\text{variance of the noise}}{\text{variance of the blurred image}} \right). \quad (44)$$

As the undistorted image is available, it is possible to evaluate the improvement in SNR introduced by the process of restoration. This is calculated as

$$\text{Improvement in SNR} = 10 \log_{10} \frac{\sum_{i,j} (g(i, j) - f(i, j))^2}{\sum_{i,j} (\hat{f}(i, j) - f(i, j))^2}. \quad (45)$$

For this image the "improvement" in SNR was −16.5 dB, which is to say that the restored image was farther from the original image than the blurred one was. The noise amplification introduced by inverse filtering caused the restoration to lose ground.

B. Least-Squares Solutions

To overcome the noise sensitivity of the inverse filter, a number of restoration filters have been developed which we will collectively call least-squares filters. This section will explore two least-squares restoration methods—the direct methods (which are usually implemented in the frequency domain) and the recursive or Kalman filtering methods (which are usually implemented in the spatial domain.)

The Wiener Solution: The Wiener filter [22] is a linear space-invariant filter which makes use of the power spectrum of both the image and the noise to prevent excessive noise amplification. The frequency response of this restoration filter, $H_w(m, n)$, is chosen to minimize the mean squared restoration error $\mathcal{E}_2$ given by

$$\mathcal{E}_2 = E(|F(m, n) - \hat{F}(m, n)|^2) \quad (46)$$

$$= E(|F(m, n) - g(m, n)\, H_w(m, n)|^2) \quad (47)$$

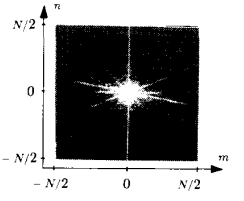
where $E(\cdot)$ denotes the expectation over an ensemble of images. The solution to this minimization problem is given by

$$H_w(m, n) = \frac{H^*(m, n)}{|H(m, n)|^2 + \dfrac{S_{nn}(m, n)}{S_{ff}(m, n)}} \quad (48)$$

where $S_{ff}(m, n)$ is the power spectrum of the original image, $S_{nn}(m, n)$ is the power spectrum of the noise, and $H^*(m, n)$ denotes the complex conjugate of $H(m, n)$. In the noiseless case the Wiener filter approximates the pseudo-inverse filter [22] defined by

$$H_{ps}(m, n) = \begin{cases} \dfrac{1}{H(m, n)} & \text{for } H(m, n) \neq 0 \\ 0 & \text{for } H(m, n) = 0. \end{cases} \quad (49)$$

An example of a Wiener filter restoration is shown in Fig. 12. The improvement in the SNR is 5.9 dB. The excessive noise amplification of the earlier example is no longer present because of the masking of the spectral zeros, but the image is still somewhat blurred. It has been a regular crit-



(a)

(b)

Fig. 12. Restoration of the image in Fig. 11(a) using a Wiener filter. (a) The power spectrum used in the Wiener filter. (b) The restored image. SNR improvement is 5.9 dB.

icism of Wiener filters that they act mainly to suppress measurement noise, while performing only minor deblurring.

Constrained Least-Squares Solution: Constrained least-squares filtering is another approach for overcoming some of the difficulties associated with the inverse filter, while still retaining the simplicity of using a single linear space-invariant filter to restore the image.

If the restoration is a good one, the blurred estimate should be approximately equal to the observed image. That is,

$$\hat{F}(m, n)\, H(m, n) \approx G(m, n).$$

With the inverse filter this approximation is made exact, which causes a problem when there are measurement errors because the inverse filter tries to get an exact fit to noisy data. It is, in fact, unreasonable to expect the restoration to match the observations any more closely than the ideal solution itself. Thus, a more reasonable expectation for the restoration is that it satisfies the relation

$$\|G(m, n) - H(m, n)\, \hat{F}(m, n)\| \approx \|N(m, n)\| = \sqrt{\sum_{m,n} |N(m, n)|^2}. \tag{50}$$

where $\|\cdot\|$ denotes the regular Euclidean norm. An estimate of the variance of the noise, and, hence $\|N(m, n)\|$, can easily be obtained from a smooth portion of the image. There are potentially many possible restorations which meet this criterion. Prior knowledge about the solution is one means for choosing among them or secondary optimization criteria can be used. One common secondary criterion, which acknowledges the tendency of the inverse filter to emphasize high frequency noise, is to require that the restoration be as "smooth" as possible.

This is the motivation for the constrained least-squares restoration [22], [25]. The restoration $\hat{F}(m, n)$ is chosen which minimizes the quantity $\Omega(\hat{F})$ defined by

$$\Omega(\hat{F}) = \|C(m, n)\, \hat{F}(m, n)\| \tag{51}$$

subject to the condition that

$$\|G(m, n) - H(m, n)\, \hat{F}(m, n)\| = \|N(m, n)\|. \tag{52}$$

Here $C(m, n)$ is the frequency response corresponding to the point-spread function $c(i, j)$ of an operator which measures the nonsmoothness of the restoration. A common choice for this operator is some form of second derivative, such as a discrete approximation to a 2-D Laplace filter [48].

The solution to the above minimization problem is again a linear space-invariant filter with the frequency response given by

$$H_{cl}(m, n) = \frac{H^*(m, n)}{|H(m, n)|^2 + \gamma |C(m, n)|^2} \tag{53}$$

where the Lagrange multiplier $1/\gamma$ is chosen so that the constraint in (52) is satisfied. Equation (53) is called the constrained least-squares solution [25], [49].

It should be noted that the formulations of the Wiener and constrained least-squares filters are very similar, although their motivations are quite different. The constrained least-squares filters can be viewed as a generalization of the direct least-squares solutions. In the limit as γ approaches 0, the limiting solution is again the pseudo-inverse solution (Eq. (49)).

Fig. 13. Restoration of the image in Fig. 11 using the constrained least-squares method. SNR improvement is 6.2 dB.

Figure 13 shows an example of a constrained least-squares restoration. The blurred image is the same as before with a defocusing blur. A Laplacian operator C was used with a value of $\gamma = 0.01$. In this case the improvement in SNR is 6.2 dB.

Recursive Solutions: Another solution to linear mean-squared error image restoration uses a Kalman filter. Once an ordering for the data has been chosen (causality condition), a Kalman filter can be defined which provides for a recursive solution to the restoration problem. Such a filter can track slowly varying image statistics and spatially varying blurs. The Kalman filter makes use of the autoregressive image model given in (11) and the causal support condition given in (12). Together with (29), these form a set of state-space equations which form the basis for a scalar Kalman filter, which filters the data one point at a time rowwise. The reduced update Kalman filter (RUKF) by Woods *et al.* [24] is a suboptimal but efficient alternative, which uses the following *state prediction* and *state update* equations.

$$\hat{f}_b^{(i,j)}(m, n) = \sum_{(p,q)\in W} a(p, q)\, \hat{f}_a^{(i-1,j)}(m - p, n - q) \tag{54}$$

$$\hat{f}_a^{(i,j)}(m, n) = \hat{f}_b^{(i,j)}(m, n) + k^{(i,j)}(i - m, j - n) \cdot \left[g(i, j) - \sum_{(k,l)\in S_1} h(k, l)\, \hat{f}_b^{(i,j)}(i - k, j - l) \right] \tag{55}$$

Here $\hat{f}(i, j)$ denotes the estimate of $f(i, j)$, and $k^{(m,n)}(i, j)$ denotes the Kalman gain. In the above expressions the superscripts refer to the step in the filtering and the arguments denote the position of the data. The subscripts b and a denote before and after the update. A Kalman filter requires the *a priori* knowledge of the image model and the blur coefficients. This identification problem was discussed in Section III-B as an ARMA identification problem.

Instead of using a scalar Kalman filter which recursively estimates one pixel at a time, Biemond *et al.* developed a Kalman filter for vector observations, in which the image is filtered one image line (row) at a time [23], [50]. By using a decorrelating row transform, under certain conditions the final algorithm reduces to a set of scalar 1-D Kalman filters suitable for parallel processing of the data in the column direction. In Fig. 14 such a system is shown, which uses row discrete Fourier transforms (DFTs) to decorrelate the column data.

By exploiting the symmetry properties of the Fourier transform for real input data, the number of Kalman filters

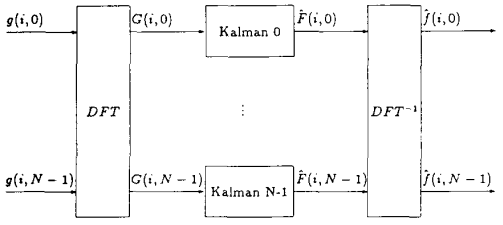


Fig. 14. Parallel Kalman filter scheme.

(channels) shown in that figure can be reduced to $N/2 + 1$. A restoration of the noisy blurred image in Fig. 11 made by this Kalman filter is shown in Fig. 15. The SNR improvement is 5.6 dB.

Fig. 15. Restoration of the defocused cameraman image using the parallel Kalman filter. SNR improvement is 5.6 dB.

This discussion of Kalman filtering for images is far from complete. It was presented in order that the Kalman restoration could be compared to those of the iterative methods, which are the real subject of this paper. A more complete discussion of this issue can be found in, for example, [23].

C. Iterative Solutions

Van Cittert's Method: The simplest of the iterative deconvolution methods has a long history. It goes back at least to the work of Van Cittert [51] in the 1930s and may, in fact, have even older antecedents. Iterative solution techniques have been applied to the image deconvolution problem by many researchers in recent years [3], [9]–[11], [14]–[21]. Although originally formulated for the space-invariant case, it can be applied to the spatially varying case as well. Neglecting, for a moment, the noise contribution and making use of the compact matrix-vector notation introduced in (10) to denote both the space-varying and space-invariant cases, the following identity is introduced, which must hold for all values of the parameter β:

$$f = f + \beta(g - Hf). \tag{56}$$

Applying the method of successive substitutions to this suggests the following iteration

$$\begin{aligned} \hat{f}_0 &= \beta g \\ \hat{f}_{k+1} &= \hat{f}_k + \beta(g - H\hat{f}_k) \\ &= \beta g + (I - \beta H)\hat{f}_k \\ &= \beta g + R\hat{f}_k \end{aligned} \tag{57}$$

where I is the identity operator. Different researchers refer to this iteration as the Van Cittert [51], Bially [52], or Landweber [53], [54] iteration, presumably because it has been independently discovered many times.

With any iterative algorithm there are two important concerns—does it converge and, if so, to what limiting solution? By direct enumeration it is seen that

$$\hat{f}_k = \beta \sum_{r=0}^{k} R^r g \tag{58}$$

which can be written notationally as

$$\hat{f}_k = \beta(I - R)^{-1}(I - R^{k+1})g \tag{59}$$

provided that the matrix $(I - R)$ is invertible, that is, H is invertible. If

$$\lim_{k \to \infty} R^{k+1} g = 0 \tag{60}$$

which is a sufficient condition for convergence, the limiting solution is

$$\hat{f}_\infty = \lim_{k \to \infty} \hat{f}_k = \beta(I - R)^{-1} g = H^{-1} g. \tag{61}$$

This is the inverse filter solution. Hence, continuing the iterations indefinitely will produce a solution which has many unsatisfactory properties. The iterative implementation of the inverse filter (57), however, does have two advantages over the direct implementation. First, it can be terminated prior to convergence, resulting in a partially deblurred image which will often not exhibit noise amplification. The second advantage is that the inverse operator does not need to be implemented. Each iteration requires only that the blurring operator itself be implemented. Other advantages of the iterative approach will become apparent in later sections.

Convergence Conditions and Properties of the Limiting Solution: We can gain a greater understanding of the iteration in (57) through an eigenvalue analysis of it. Not only will this provide a better understanding of the convergence condition in (60), but it will also explain why more satisfactory results occur when the iteration is terminated prior to convergence. It is also useful for understanding generalizations of this basic iteration in later sections.

To begin, consider the blurring operation in its matrix-vector form

$$g = Hf + n \tag{62}$$

where g and f are lexicographically stacked images and H is the blurring operator. Now let $\{v_{mn}(i, j)\}$ denote the eigenvectors associated with the blurring matrix H and let the scalars $\{\lambda_{mn}\}$ represent the corresponding eigenvalues (see Appendix). By expanding $\hat{f}_k$ in terms of these eigenvectors we get

$$\hat{f}_k = \sum_{m,n} (\hat{f}_k, v_{mn}) v_{mn} \tag{63}$$

where $(\cdot, \cdot)$ denotes the inner product between two vectors. By also expanding g in terms of $\{v_{mn}\}$, and substituting these results into (57), we arrive at

$$\begin{aligned} \hat{f}_{k+1} &= \sum_{m,n} (\hat{f}_{k+1}, v_{mn}) v_{mn} \\ &= \sum_{m,n} \beta(g, v_{mn}) v_{mn} + (I - \beta H) \sum_{m,n} (\hat{f}_k, v_{mn}) v_{mn} \\ &= \sum_{m,n} [\beta(g, v_{mn}) + (1 - \beta\lambda_{mn})(\hat{f}_k, v_{mn})] v_{mn} \end{aligned} \tag{64}$$

or

$$(\hat{f}_{k+1}, v_{mn}) = \beta(g, v_{mn}) + (1 - \beta\lambda_{mn})(\hat{f}_k, v_{mn}). \quad (65)$$

Eq. (65) shows that once the eigenvectors of H have been obtained, the matrix iteration (57) can also be evaluated as a set of independent scalar iterations.

The restoration obtained after k Van Cittert iterations, $\hat{f}_k$, can be written in terms of the eigenvectors and eigenvalues as

$$\hat{f}_k = \beta \sum_{m,n} \left[\sum_{r=0}^{k} (1 - \beta\lambda_{mn})^r \right] (g, v_{mn}) v_{mn}$$

$$= \sum_{m,n} (1 - (1 - \beta\lambda_{mn})^{k+1}) \frac{(g, v_{mn})}{\lambda_{mn}} v_{mn}. \quad (66)$$

As $k \to \infty$ the sequence of iterates converges to

$$\hat{f}_\infty = \sum_{m,n} \frac{(g, v_{mn})}{\lambda_{mn}} v_{mn} = \sum_{m,n} \frac{(Hf, v_{mn})}{\lambda_{mn}} v_{mn}$$

$$= \sum_{m,n} \frac{\lambda_{mn}(f, v_{mn})}{\lambda_{mn}} v_{mn} = \sum_{m,n} (f, v_{mn}) v_{mn} = f \quad (67)$$

if

$$|1 - \beta\lambda_{mn}| \leq 1, \qquad \forall m, n. \quad (68)$$

This convergence condition is equivalent to that given in (60). As the eigenvalues are complex numbers, they must all lie in the shaded circle of the complex plane (Fig. 16).

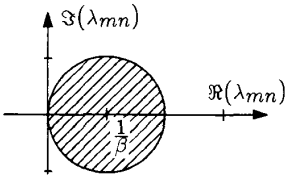


Fig. 16. Region of the complex plane in which all of the eigenvalues of the blurring operator must lie for the Van Cittert iteration to converge.

In the special case that the blur is space-invariant, the eigenvalues are the discrete Fourier transform coefficients $H(m\,n)$ and the eigenvectors are complex exponentials (see Appendix). In this case the inner products (g, v_{mn}) are values of the Fourier transform of the blurred image $G(u, v)$, and (67) is readily identified as the inverse filter solution.

The above analysis has assumed that measurement noise was not present. When noise has been added to the blurred image, (66) becomes

$$\hat{f}_k = \sum_{mn} \left\{ (1 - (1 - \beta\lambda_{mn})^{k+1}) \left[(f, v_{mn}) + \frac{(n, v_{mn})}{\lambda_{mn}} \right] v_{mn} \right\}. \quad (69)$$

When there is no noise, this converges to

$$\hat{f}_\infty = \sum_{m,n} (f, v_{mn}) v_{mn} = f \quad (70)$$

but when noise is present, the last term in (69) causes the limiting solution to deviate from the ideal. This deviation is given by

$$\|\hat{f}_k - f\| = \left\| \sum_{m,n} (1 - \beta\lambda_{mn})^{k+1} (f, v_{mn}) v_{mn} - \sum_{m,n} (1 - (1 - \beta\lambda_{mn})^{k+1}) \frac{(n, v_{mn})}{\lambda_{mn}} v_{mn} \right\|$$

$$\leq \sum_{m,n} |1 - \beta\lambda_{mn}|^{k+1} |(f, v_{mn})| + \sum_{m,n} |1 - (1 - \beta\lambda_{mn})^{k+1}| \frac{|(n, v_{mn})|}{|\lambda_{mn}|}$$

$$= E_1(k) + E_2(k). \quad (71)$$

This error bound has two terms. The first of these can be made arbitrarily small by letting $k \to \infty$. This term represents the degree of deblurring in the restored image. The second term in (71) approaches

$$\lim_{k \to \infty} E_2(k) = \sum_{m,n} \frac{|(n, v_{mn})|}{|\lambda_{mn}|}.$$

As the high-order eigenvalues of the blurring operator are typically infinitessimal or zero, this second term can become arbitrarily large. As $E_1(k)$ decreases with increasing k and $E_2(k)$ increases, their sum may attain its minimum after a finite number of iterations. Unfortunately, the optimal number of iterations is usually not known in advance.

To illustrate this effect, consider the example in Fig. 17. The left column of Fig. 17 shows the restoration of the noisy, defocused cameraman image in Fig. 11 using the Van Cittert scheme ($\beta = 1$). Results are shown after 15, 250, 4000, and ∞ iterations. The images in the center column show the error images caused by the partial deblurring and those in the right column show the error caused by the noise magnification. For a small number of iterations, the error caused by partial deblurring is clearly seen, whereas for a higher number of iterations the noise amplification is apparent. This effect is also seen in Fig. 18 in which the components $E_1(k)$ and $E_2(k)$ of the total error are plotted as a function of the number of iterations. For this example the optimum restoration occurred at approximately 250 iterations. The SNR improvement after 250 iterations was 5.7 dB. It is worthwhile noticing that the best visual result seems to occur for $k = 4000$. This indicates that the SNR measurement does not correlate well with the subjective judgment of the image quality. On the other hand, it can be considered as an additional advantage of the iterative schemes that they provide for the possibility of monitoring and terminating the iterations when a "visually optimal" solution has been reached.

Reblurring: In the previous section it was seen that a necessary and sufficient condition for the convergence of the Van Cittert iteration was that

$$|1 - \beta\lambda_{mn}| < 1, \qquad \forall m, n.$$

As β is a free parameter, this is equivalent to the condition

$$\Re(\beta\lambda_{mn}) > 0 \quad (72)$$

where $\Re(\cdot)$ denotes the real part operator. In the space-invariant case this implies that the blurring operator must have a transfer function with a positive real part for all frequencies. This condition is not satisfied for the two important blurs discussed earlier, linear motion blur, and out-of-focus blur.

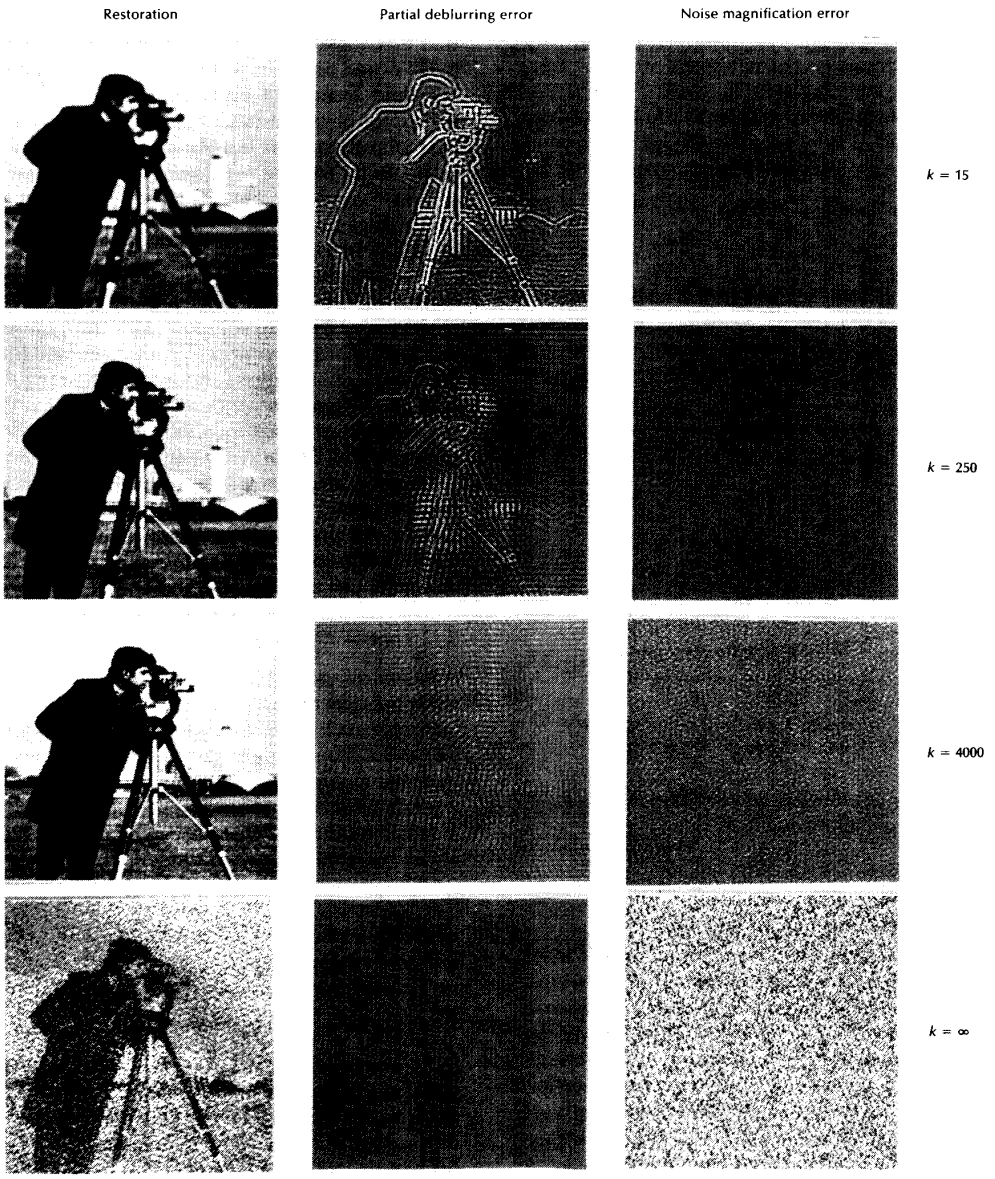


Fig. 17. Effect of limiting the number of iterations. Restorations, deblurring error, and noise magnification errors after 15, 250, 4000, and ∞ iterations.

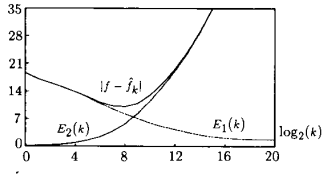


Fig. 18. Total error, partial deblurring error, and noise magnification error as a function of the number of iterations.

To overcome this problem, several authors have introduced the idea of using a "reblurring" operation in the iteration [1], [10], [11]. This is equivalent to applying the Van Cittert procedure to the identity

$$f = f + \beta H^*(g - Hf) \tag{73}$$

where H^* is the conjugate transpose of H. This yields the iteration

$$\begin{aligned}\hat{f}_{k+1} &= \beta H^* g + (I - \beta H^* H)\hat{f}_k \\ &= \hat{f}_k + \beta H^*(g - H\hat{f}_k).\end{aligned} \tag{74}$$

If a similar convergence analysis is applied to this iteration, convergence is seen to require

$$0 < |1 - \beta\lambda_{mn}^2| < 1. \tag{75}$$

This is equivalent to the requirement that $\lambda_{mn} \neq 0$, which is a weaker condition than the positive real property given in (72). This particular condition is not satisfied for blurring

operators having zero eigenvalues. However, in this case $\hat{f}_k$ converges to the pseudoinverse solution [53], [54].

Alternative derivations of the reblurred iteration have appeared in the literature. It can be shown [11] that this iteration corresponds to an iterative optimization procedure, based on the method of steepest descent [55], [56] for minimizing the norm of $g - H\hat{f}$ (the residual image). That is,

$$\min_{\hat{f}} \phi(\hat{f}) = \min_{\hat{f}} \|g - H\hat{f}\| = \min_{\hat{f}} [(g - H\hat{f})^*(g - H\hat{f})]^{1/2}. \tag{76}$$

The value of $\phi(\hat{f})$ can also be used to evaluate the degree of convergence of the iterative procedure because this number can be estimated *a priori* from the norm of the measurement noise [4]. The graph in Fig. 19 shows the value of

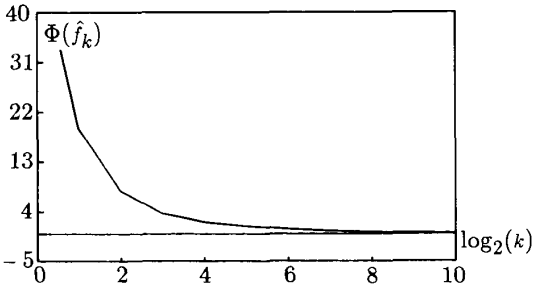


Fig. 19. A curve of $\phi(\hat{f})$ versus the number of reblurred Van Cittert iterations for the defocused image in Fig. 11(a).

the function $\phi(\hat{f})$ for the defocused cameraman, as a function of the number of iterations.

V. Regularization

A. Introduction

The previous section showed that the small eigenvalues of the blurring operator could cause the filtered observation noise to dominate the inverse filter solution. Because of this, many deblurring procedures that try to recover the high frequency components of an image are ill-conditioned. Methods to suppress this noise amplification include (constrained) least-squares solutions and methods which limit the number of iterations of the iterative implementation. This section will discuss the noise magnification problem in the more general context of ill-posed problems and regularization [11], [26]–[27], [57]–[65].

The dilemmas involved in estimating an original image f from a linearly distorted and noisy observation g which is now called an *inverse problem*, was first studied by Hadamard [66] in the early 1900s in the inversion of certain integral equations. He observed that the solution $\hat{f}$ could differ by an arbitrarily large amount from the true solution because of small errors in measuring the observed signal. Based on his investigations and that of later mathematicians, the term "ill-posed problem" was introduced to denote the class of inverse problems that behaved in a similar manner. At that point the main objective in solving ill-posed problems became [26], [65] "the construction of a physically acceptable and meaningful approximation of the true solution of an ill-posed problem which is sufficiently stable from the computational viewpoint." Regularization encompasses a class of solution techniques which entails the analysis of an associated well-posed problem, provided that this analysis yields physically meaningful answers to the ill-posed problem. Although Hadamard's arguments do not hold exactly in finite-dimensional spaces, in other words, for ill-conditioned matrices, many tools from regularization theory for infinite dimensional problems (such as methods for the inversion of certain integrals) have become popular and useful in finite-dimensional settings. We restrict ourselves here to the finite-dimensional formulation, in which H is a matrix operator.

Nearly all of the concepts used in regularization are based on incorporating knowledge about either the true solution or the noise into the solution algorithm. Observe that, in this sense, the procedures already discussed for truncating the number of iterations should be called regularization as well. In this section we describe the most widely used of the regularization methods, which is usually associated with the names of Tikhonov [26] and Miller [27]. Both the noniterative and iterative restorations based on Tikhonov-Miller regularization will be analyzed using the eigenvector expansions presented earlier. Another family of regularization methods based on restricting the space of feasible solutions will be discussed in Section VI.

B. Tikhonov-Miller Regularization

Tikhonov and Arsenin [26] were the first to study exclusively the concepts of regularization, although some important prior work had been performed by Phillips [67], Twomey [68] and a number of Russian mathematicians. The idea is to define a criterion to select an approximate solution from a *set* of admissible solutions. Based on (52), define a class of feasible solutions Q_ϵ as those images for which the norm of the residual image is bounded. That is,

$$\phi(\hat{f}) = \|g - H\hat{f}\| = \|n\| \leq \epsilon \tag{77}$$

where $\|\cdot\|$ denotes the Euclidean norm. This bound is related to the uncertainty in the observed image g and can be estimated from a smooth image region using

$$\epsilon \approx \|n\| \approx \sqrt{\text{var}\,(g)}|_{\text{smooth region}}. \tag{78}$$

Q_ϵ is primarily populated with unacceptable (very noisy) solutions because of the ill-conditioned nature of the restoration problem. Tikhonov defined the regularized solution as the one which minimizes a stabilizing functional $\Omega(\hat{f})$ on the set Q_ϵ.

Although a wide class of different stabilizing functionals is available (including, for example, maximum entropy measures [69], [70]), usually a stabilizing functional of the following form is chosen:

$$\Omega(\hat{f}) = \|C\hat{f}\| \tag{79}$$

where C is a matrix operator of size $N^2 \times N^2$, known as the *regularizing operator*. The properties of this operator will be described shortly. The computation of the regularized solution reduces to the minimization of (79) subject to (77). Using the method of undetermined Lagrange multipliers the problem reduces to the minimization of

$$\Phi(\hat{f}) = \|g - H\hat{f}\|^2 + \alpha\|C\hat{f}\|^2 \tag{80}$$

where α, the *regularization parameter*, is chosen so that (77) is satisfied with equality.

Another related approach was presented by Miller [27]. He replaced the minimization of $\Omega(\hat{f})$ by a constraint on the

value of $\Omega(\hat{f})$

$$\Omega(\hat{f}) = \|C\hat{f}\| \leq E. \tag{81}$$

Eqs. (81) and (77) can be combined into a single quadrature formula. The result is identical to the Tikhonov result with $\alpha = (\epsilon/E)^2$. Other ways to select α are discussed in [71], [72].

The minimization of $\phi(\hat{f})$ with respect to $\hat{f}$ is straightforward and leads to the normal equations

$$(H^*H + \alpha C^*C)\hat{f}_{tm} = H^*g. \tag{82}$$

The solution $\hat{f}_{tm}$ can be computed from g directly (if the operator $(H^*H + \alpha C^*C)$ is invertible) or iteratively. Both methods will be discussed and analyzed in terms of their eigenvector decomposition.

C. Direct Solution

From (82) the solution of the Tikhonov-Miller procedure is

$$\hat{f}_{tm} = (H^*H + \alpha C^*C)^{-1}H^*g \tag{83}$$

which is a more general description of the space-invariant constrained least-squares filter described in Section IV-B. The two solutions, in fact, are identical in the space-invariant case, if C is chosen appropriately, but the Tikhonov-Miller solution is valid in the space-varying case as well.

Assuming that both H, H^*, C, and C^* have the same set of eigenvectors $\{v_{mn}\}$,[1] and that $\{\lambda_{mn}\}$ and $\{\sigma_{mn}\}$ are the eigenvalues belonging to H and C, respectively, then the Tikhonov-Miller restoration $\hat{f}_{tm}$ is given by [11], [57], [64], [65], [73]

$$\hat{f}_{tm} = \sum_{m,n} \frac{\lambda^*_{mn}}{\lambda^2_{mn} + \alpha\sigma^2_{mn}} (g, v_{mn})v_{mn}. \tag{84}$$

Clearly, the effect of regularization is to modify the denominator of (67). The user chooses the regularization operator C and thus its eigenvalues. To decide what is a good choice for C it is appropriate to analyze the difference between the true and regularized solutions. This error can be bounded using a technique similar to the one presented earlier. The resulting bound is given by

$$\|f - \hat{f}_{tm}\| \leq \sum_{m,n} \left| \frac{\alpha\sigma^2_{mn}}{\lambda^2_{mn} + \alpha\sigma^2_{mn}} \right| |(f, v_{mn})| + \sum_{m,n} \left| \frac{\lambda^*_{mn}}{\lambda^2_{mn} + \alpha\sigma^2_{mn}} \right| |(n, v_{mn})|. \tag{85}$$

The first term on the right side of this expression denotes the error caused by the regularization. It can be minimized by setting $\alpha = 0$. The second term, which measures the noise magnification error, however, becomes infinite as $\alpha \to 0$ if any of the $\{\lambda_{mn}\}$ are zero. The choice of α requires a tradeoff between these two errors.

The user can also choose the regularizing operator C to trade off the two error terms. This is most conveniently done by selecting the eigenvalues σ_{mn}. As the original signal f should not be overly corrupted by the regularization, it is reasonable to choose $\sigma_{mn} \ll \lambda_{mn}$ when $|(f, v_{mn})| \gg |(n, v_{mn})|$. This means that there will be little regularization of components where on average the signal energy is much greater than the noise energy. On the other hand, in those components where the noise energy generally dominates the signal energy, there should be a great deal of regularization. In the space-invariant case this means that because (i) the signal energy is concentrated in the low frequency range, (ii) the noise is broad-band, and (iii) the blur acts like a form of low-pass filter, the regularizing operator C should act like a high-pass filter (such as a discrete approximation to a 2-D Laplacian filter.) The above qualitative discussion is thus in complete agreement with the motivation of the constrained least-squares filter in Section IV-B.

Observe that we can rewrite the image model (11), given by

$$f(i, j) = \sum_{p,q \in W} a(p, q)\, f(i - p, j - q) + u(i, j) \tag{86}$$

as follows:

$$f = Af + u$$
$$(I - A)f = u \tag{87}$$

where f and u are lexicographically ordered images and where A is the image model matrix which is defined by coefficients $a(p, q)$. By taking the norm of both sides of (87), we arrive at a relation similar to (81):

$$\|(I - A)f\| = \|u\| \leq E. \tag{88}$$

By setting $C = (I - A)$ it is clear that the regularizing operator and the 2-D recursive image model (11) are in fact related concepts. The restoration $\hat{f}_{tm}$ is fairly robust with respect to the choice of both α and C.

D. Iterative Solution

For a general linear operator (83) cannot be evaluated, because this requires the inversion of an $N^2 \times N^2$ matrix, but iterative solution methods can again be used. The following iteration is similar in form to the reblurred Van Cittert iteration. It can also be derived by minimizing (80) using a steepest descent algorithm [11], [55], [56]:

$$\hat{f}_{k+1,tm} = \beta H^*g + (I - \beta(H^*H + \alpha C^*C))\hat{f}_k$$
$$= (I - \alpha\beta C^*C)\hat{f}_k + \beta H^*(g - H\hat{f}_k). \tag{89}$$

The regularized solution after k iterations is given in terms of the eigenvalues and eigenvectors of the blurring and regularization operators as

$$\hat{f}_{k,tm} = \beta \sum_{mn} \left[\sum_{r=0}^{k} (1 - \beta(\lambda^2_{mn} + \alpha\sigma^2_{mn}))^r \right] \lambda^*_{mn}(g, v_{mn})v_{mn}$$
$$= \sum_{mn} \frac{\lambda^*_{mn}}{\lambda^2_{mn} + \alpha\sigma^2_{mn}} (1 - (1 - \beta(\lambda^2_{mn} + \alpha\sigma^2_{mn}))^{k+1}) \cdot (g, v_{mn})v_{mn}. \tag{90}$$

From this the convergence conditions follow directly

$$|1 - \beta(\lambda^2_{mn} + \alpha\sigma^2_{mn})| < 1, \quad \forall m, n. \tag{91}$$

If the iterations converge, the limiting solution is given by (84). Again, when the iteration is terminated after k iterations, there will be two sources of error, one because convergence has not been achieved and because the solution is regularized, and one caused by the filtered measurement

[1] This assumption is true, for example, if these operators are space-invariant. Similar expression can, however, be obtained for more general cases.

error:

$$\|f - \hat{f}_{k,tm}\| \leq \sum_{m,n} \left\{ \left| \frac{\alpha\sigma_{mn}^2 + \lambda_{mn}^2(1 - \beta(\lambda_{mn}^2 + \alpha\sigma_{mn}^2))^{k+1}}{\lambda_{mn}^2 + \alpha\sigma_{mn}^2} \right| |(f, v_{mn})| + \left| \frac{\lambda_{mn}(1 - (1 - \beta(\lambda_{mn}^2 + \alpha\sigma_{mn}^2))^{k+1})}{\lambda_{mn}^2 + \alpha\sigma_{mn}^2} \right| |(n, v_{mn})| \right\}. \quad (92)$$

This expression reduces to several of the ones already derived if the number of iterations is increased to ∞ or if the regularization parameter is set to zero. Observe that (89) reduces to the (reblurred) Van Cittert iteration if $\alpha = 0$ (no Tikhonov-Miller regularization).

In [11], [73] Katsaggelos *et al.* recognize that the term $(I - \alpha\beta C^*C)$ in (89) behaves like a low-pass filter, suppressing the noise amplification in the iterates. As the characteristics of this *stabilizing term* are obviously related to the properties of the original image, they proposed to compress this term into one single low-pass operator C_s, which would reflect spectral knowledge about the original image. Eq. (89) then becomes

$$\hat{f}_{k+1} = C_s\hat{f}_k + \beta H^*(g - H\hat{f}_k). \quad (93)$$

One choice for C_s is the noise smoothing Wiener filter [74], [75], which assumes the form

$$C_s = S_{ff}(S_{nn} + S_{ff})^{-1} \quad (94)$$

where S_{nn} and S_{ff} are the autocorrelation matrices of the noise and the original image, respectively. It can be shown [73] that the limiting solution of the iteration in (93) using (94) assumes a form quite similar to the parametric Wiener filter [22]. The advantage of (93) over (89) is that the interpretation of (93) is more clear. In practice the construction of a suitable filter, C_s is sometimes easier than the selection of a regularizing operator C and the related bound E.

E. Example

This example demonstrates the effect of regularization and the different types of errors that are present in (71), (85), and (92). The restoration of the defocused cameraman image with an SNR of 40 dB in Fig. 11(a) is again used. A number of regularized restorations were formed using a Laplacian regularization operator and the resulting restoration, regularization error, and noise magnification error were displayed as a function of both the regularization parameter α and the iteration index k. These results are shown in Figs. 20–22. Each of these represents a montage of 20 images arranged in 5 rows of 4 images each. The four columns in each figure correspond to $k = 15$, 250, 4000, and ∞ iterations, and the five rows correspond to $\alpha = 0$, 10^{-4}, 10^{-3}, 10^{-2}, and 10^{-1}. Fig. 20 shows the noise magnification error, Fig. 21 shows the regularization error, and Fig. 22 shows the resulting restorations. Observe that the right-most columns show errors and restorations as a function of α only, and that the top-most rows show the two types of errors and restorations as a function of k only. (These, in fact, are identical to Fig. 17.) Further, the regularization error is seen to be considerable only near sharp intensity transitions in the image, while the noise magnification degrades the whole image. Again the tradeoff between the two types of errors is clear: for small k and/or large α the regularization error dominates, while for large k and small α the noise magnification dominates.

Finally, the noise magnification error, regularization error, and the total error $\|f - \hat{f}_k\|$ have been plotted in Fig. 23 as a function of both α and k.

VI. Deterministic Constraints

In many image restoration problems there is *a priori* knowledge available about the original image which cannot be expressed in the form of a stabilizing functional. This knowledge, however, often can be used to reduce the set of feasible solutions, in this way achieving another form of regularization [76]–[78]. (For example, it is known that image intensity can never be negative.) These *deterministic* properties can be incorporated into an image restoration algorithm if the set of solutions C_i which satisfy the constraint is a closed convex set [6]. A nonexpansive mapping P_i can be associated with each constraint set. It maps all images which violate the constraint onto C_i. That is,

$$P_i f = \begin{cases} f, & \text{if } f \in C_i \\ h, & \text{if } f \notin C_i, \end{cases} \quad (95)$$

where

$$\|h - f\| \leq \|x - f\|, \quad \forall x \in C_i. \quad (96)$$

Some examples of deterministic constraints which define closed convex sets are the nonnegativity of the intensity values or, more generally, a bounded range on the intensities, a maximum value for the signal energy, and finite support for the image. Other constraints are discussed in [4]–[8], [13], [63]. The power of these constraints is also discussed in these references.

There are two different, but related, methods for incorporating deterministic constraints into the process of image restoration. These two methods will be described in the following two subsections, but only the latter method is used in the later examples.

A. Projections onto Convex Sets

The theory of projections onto convex sets [6], [79] was developed to find an image in the intersection C_0 of m convex sets of images C_i, $i = 1, 2, \cdots, m$. 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 reconstructed image, then any image in the intersection should be reasonable. If the P_i denote the projections onto the convex sets C_i, then the iteration

$$\hat{f}_k = (P_1P_2 \cdots P_m)^k\hat{f}_0 \quad (97)$$

will converge to a point in the intersection C_0 for all initial estimates $\hat{f}_0$, unless C_0 is empty. If C_0 is empty the iterations will not stabilize [5], [6], [8], [12]. The exact properties of the limiting solution $\hat{f}_\infty$ will depend on the initial estimate, unless the intersection C_0 contains only a single element (which is extremely rare). The iterations (97) have found wide applications in various signal processing applications, such

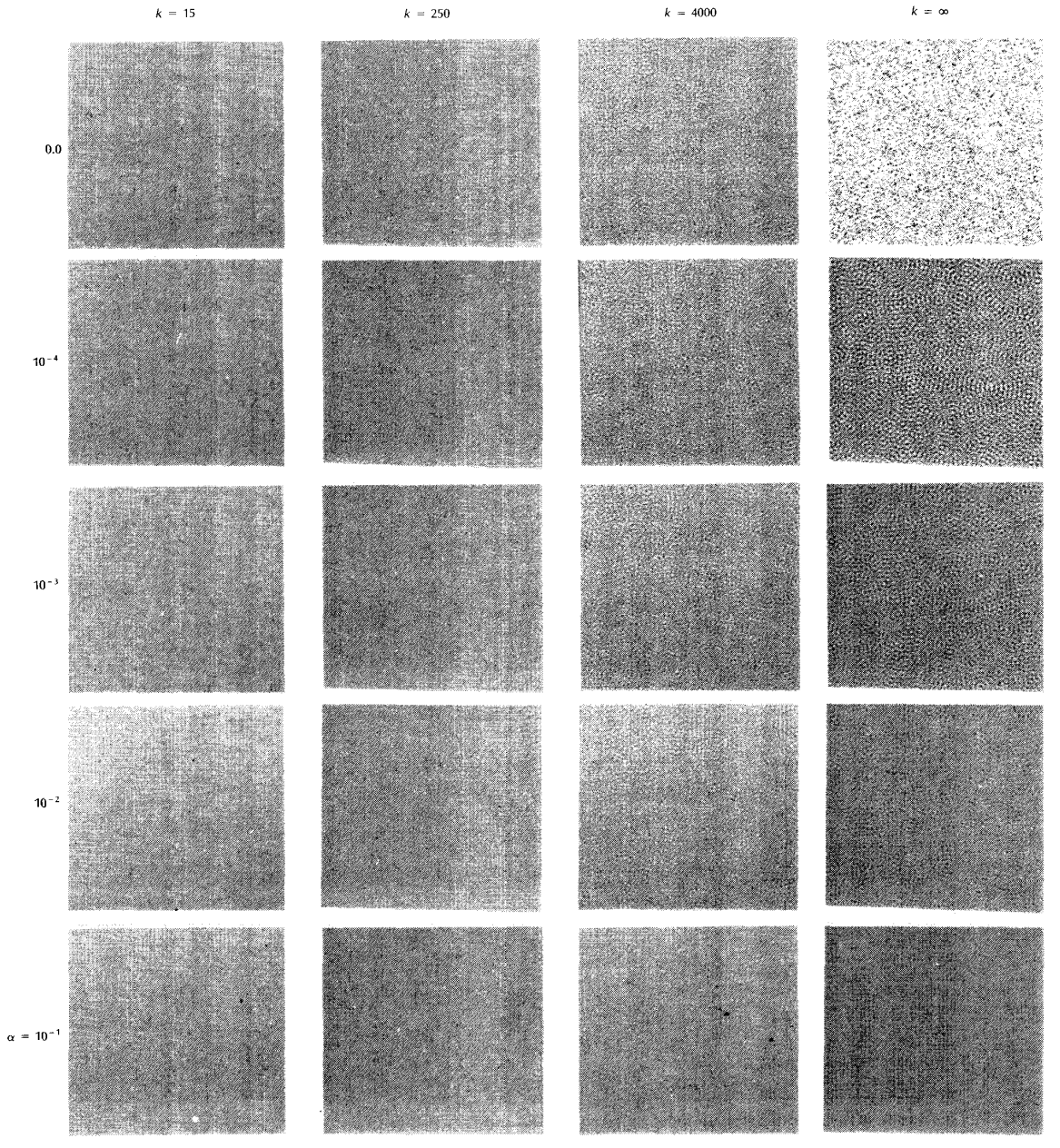


Fig. 20. Noise magnification error for the iterative regularized restoration of a defocused image with SNR = 40 dB.

as bandlimited extrapolation, space-limited extrapolation, and phase and magnitude retrieval [1], [8], [12], [63], [81]–[83].

In terms of the discussion in Section V, in which it was observed that the restoration problem becomes less ill-conditioned when more knowledge about the original image is incorporated into the solution method, better solutions will be obtained when more constraints are used, or when the constraints are made tighter. In both situations, the intersection C_0 is made smaller, thus reducing the deviation between the elements in the set. It should be observed that, because (77) and (81) define convex sets, Tikhonov-Miller regularization can also be used within the framework of projections onto convex sets [13], [84].

Recent research has led to the extension of the method of projections onto convex sets (POCS) to projections onto fuzzy sets [80]. In this method the "hard" boundaries defining a convex set 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.

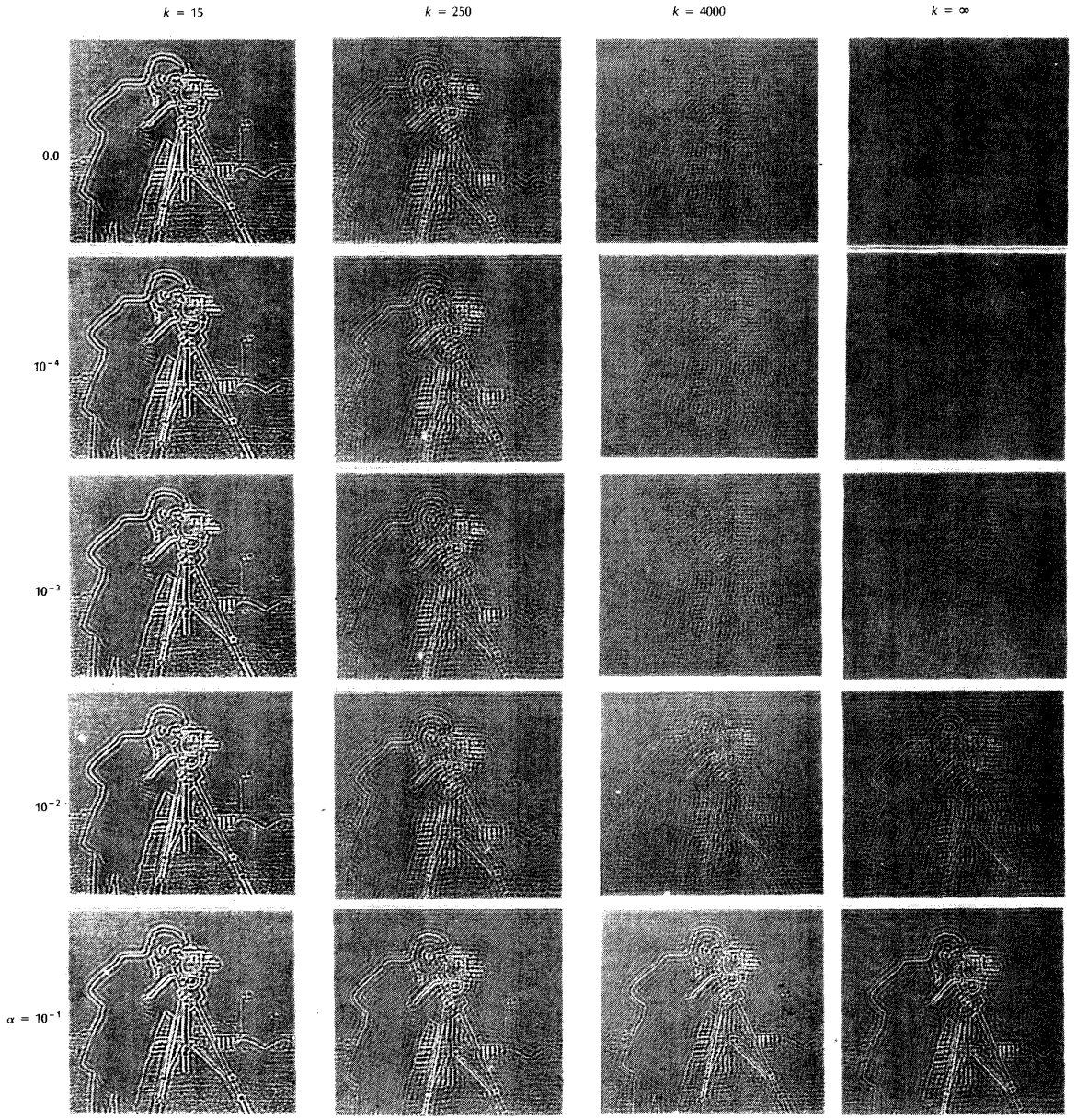


Fig. 21. Regularization error for different numbers of iterations and different values of the regularization parameter.

B. Constrained Minimization

Another method for incorporating determinstic constraints into the restoration process is to extend the basic iterations given in (89) as follows: [1], [4], [14]

$$\hat{f}_{k+1} = P[(I - \alpha\beta C^*C)\hat{f}_k + \beta H^*(g - H\hat{f}_k)] \qquad (98)$$

where P is again a projection onto a convex set C. It can be shown that the convergence conditions for this iteration are still given by (91), and that iterative schemes of this type minimize a quadratic functional such as (80) subject to the nonlinear constraint related to the projection operator [14].

The difference in restoration performance between (97) and (98) is usually small. Their major differences lie in the number of constraints that they can handle, the convergence conditions, and the convergence speed. The remainder of this paper will consider only algorithms of the form of (98) because these can be extended to a more complicated observation equation (Section VIII) and can be replaced by alternative iterations which converge faster (Section IX).

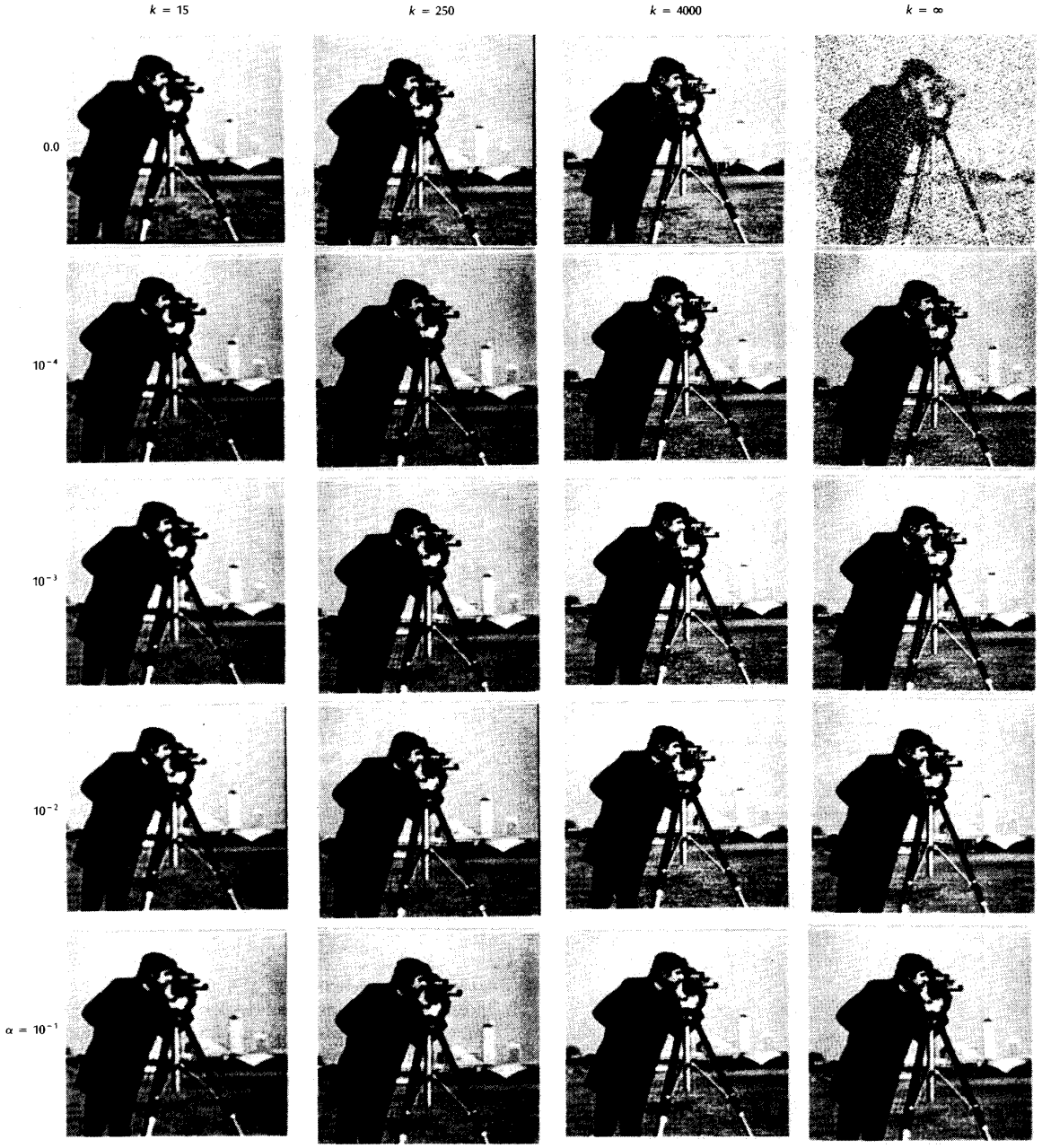


Fig. 22. Regularized restorations for different numbers of iterations and different values of the regularization parameter.

C. Constraint Tightness

In order to demonstrate the effect of deterministic constraints on the iterative restorations, consider the original text image in Fig. 24(a). This image is used for this demonstration because it has highly constrained intensity values, $25 \leq f(i,j) \leq 210$. Defocussing blur with $r = 7$ was simulated and noise with SNR = 30 dB was added to the result (Fig. 24(b)). In Fig. 25 two sequences of restorations are shown. The two upper results (1), (2) were obtained using the constrained least-squares filter, which did not make use of deterministic constraints. By using the iteration in (98) with different deterministic constraints which bound the intensities in the restored image, the results in (3)–(10) were obtained. It is clear that the tighter the constraints, the better the restoration.

The right sequence, obtained by nearly disabling the Tikhonov-Miller regularization ($\alpha = 0.00005$), shows that the

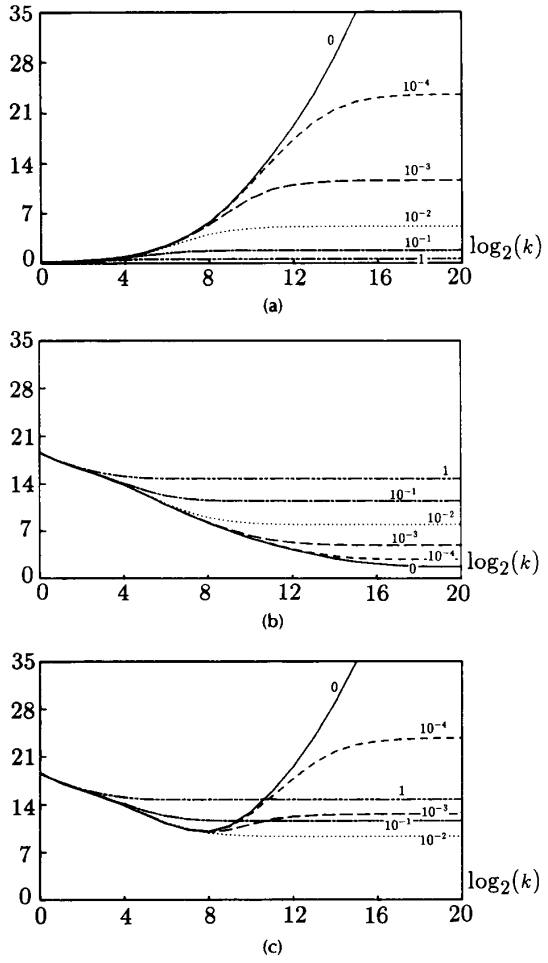


Fig. 23. (a) Noise magnification error, (b) regularization error, and (c) total error $\| f - \hat{f}_{k,tr} \|$ as a function of k and α.

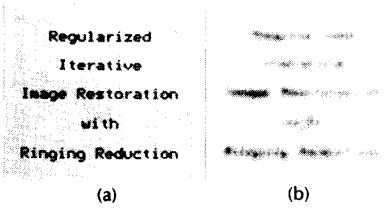


Fig. 24. (a) Original text image. (b) Defocused text image.

use of a deterministic constraint can reduce the noise magnification significantly. The left sequence, which uses much more regularization (α = 0.05), shows that deterministic constraints can also reduce the ringing artifacts which are visible in image (1) [14]. The issue of ringing reduction is discussed in greater detail in the next section.

VII. Spatially Adaptive Implementations

It has been widely observed that linear, shift-invariant restoration algorithms, such as the ones described in Section IV, often introduce ringing artifacts (superwhites,

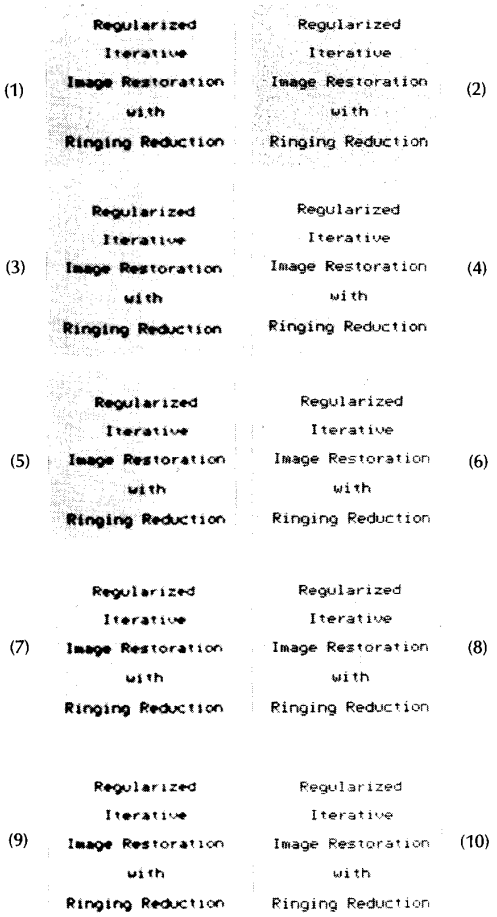


Fig. 25. Effect of tightness of a deterministic constraint. The column of restorations on the left have $\alpha = 0.05$ and the ones on the right have $\alpha = 0.00005$. From top to bottom the values for the bounds are: (1, 2): no bounds, (3, 4): [5,240], (5, 6): [15,220], (7, 8): [20, 215], (9, 10): [25, 210].

superblacks, overshoots, and undershoots) near sharp-intensity transitions. This ringing seriously reduces both the visual and measurable quality of the restoration. This section briefly considers the origin of ringing artifacts [14], and describes the spatially adaptive implementation of (98), by which ringing artifacts can be reduced.

A. Ringing Artifacts

Consider a linear space-invariant deblurring filter with the frequency response $L(m, n)$. The deviation of this filter from the inverse filter $H^{-1}(m, n)$ can be measured by the error spectrum $E(m, n)$, defined by

$$E(m, n) = 1 - L(m, n)\, H(m, n). \tag{99}$$

Through some straightforward mathematical manipulations it can be shown that the restoration error $\hat{F}(m, n) - F(m, n)$ consists of two terms:

$$F(m, n) - \hat{F}(m, n) = -E(m, n)\, F(m, n) + \frac{1 - E(m, n)}{H(m, n)}\, N(m, n). \tag{100}$$

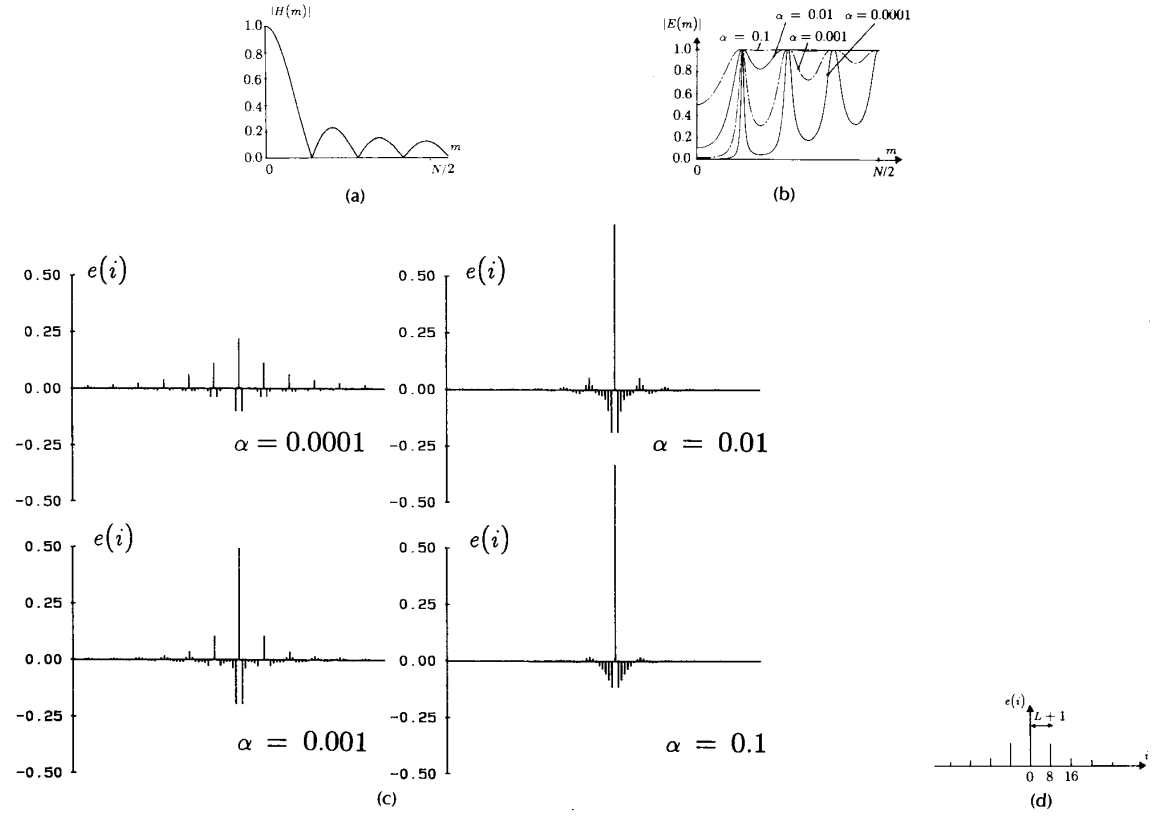


Fig. 26. Properties of the regularization error of the constrained least-squares filter for linear motion blur over 8 pixels ($L = 7$). (a) Modulus of the transfer function. (b) Modulus of the error spectrum. (c) Error sequences for different values of α. (d) Typical behavior of the dominant impulses in an error sequence.

The *noise magnification error* (second term in (100)) is independent of the original image. The more the filter $L(m, n)$ resembles the inverse filter $H^{-1}(m, n)$, the larger this error will be. The *regularization error* (first term on the right side of (100)) introduces data-dependent degradations related to the local structure in the image. Ringing is therefore attributable to the regularization error.

The relationship between the regularization error and ringing artifacts can be illustrated by considering a general deblurring filter, such as the constrained least squares filter. The error spectrum $E(m, n)$ for this filter is shown in Fig. 26 for the case of linear motion with $L = 7$ in the horizontal direction [14]. The error sequence $e(i, j)$, which is defined as the inverse Fourier transform of $E(m, n)$, is also shown in Fig. 26 for this example. The regularization error in the spatial domain is given by the convolution of $f(i, j)$ with $-e(i, j)$. Owing to the peaks in $E(u, v)$, $e(i, j)$ is dominated by positive impulses at integer multiples of the blurring distance $L + 1$. This, in turn, leads to negative echos of the intensity transitions in the restored images, that is, ringing artifacts.

One way of reducing ringing artifacts is through the use of *a priori* knowledge of the original image (Section VI). If the image data, for example, consists of blurred bright point sources against a black background (such as occurs in astronomical imaging), the ringing manifests itself as negative intensity values. A positivity constraint on the restoration can thus prevent this ringing from happening. The iterative restoration procedures that were presented in Section VI are particularly effective for ringing reduction when the constraints can be made tight. However, in the restoration of more complicated signals, such as images of natural scenes, the use of deterministic constraints alone is usually insufficient to significantly reduce ringing.

Another technique for reducing ringing locally regulates the noise magnification and regularization errors. This adaptation depends upon the local edge content of the image. 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. Fortunately, however, it is known from psychophysical experiments that, although the response of the human visual system is very complex, the visibility of the noise is greatly masked near sharp-intensity transitions (noise-masking effect), whereas blurring generally appears to be unacceptable in this context [85]. As blurring is acceptable in nearly constant portions of an image, but noise magnification is not, the restoration filter should use considerably more regularization in these parts of an image.

Restoration filters which are implemented in the frequency domain, such as the Wiener and constrained least-squares filters, are unsuitable for such an adaptive

approach. In [86], Tekalp *et al.* describe a multiple image model Kalman restoration filter in which a number of image models are used to filter an image in agreement with the local edge orientations. In this way they achieve adaptive regularization. Although real images cannot be adequately characterized using only a limited number of image models, this scheme outperforms the nonadaptive ones. The next section describes an iterative restoration method in which both ringing reduction methods—adaptive processing and the use of constraints—are incorporated.

B. Iterative Restoration in a Weighted Space

Adaptivity can be incorporated into the restoration algorithms by defining a different regularization operator C for every pixel. In our implementation this will be done implicitly by considering a single global regularization operator and by varying the noise smoothing/regularization tradeoff through the use of weighted norms.

To define the adaptive regularized iterative procedure (77) and (81) are replaced by

$$\|g - H\hat{f}\|_R = [(g - H\hat{f})^* R(g - H\hat{f})]^{1/2} \leq \epsilon \quad (101)$$

and

$$\|C\hat{f}\|_S = [(C\hat{f})^* S(C\hat{f})]^{1/2} \leq E. \quad (102)$$

Here R is an $N^2 \times N^2$ diagonal matrix containing positive weighting coefficients r_{ij} associated with the pixels at locations (i, j). These locally regulate the restoration process. In the vicinity of steep-intensity transitions, r_{ij} is assigned a relatively large value. Owing to the fixed upper bound ϵ, the residual has to be small at these points, which enforces (pseudo) inverse filtering and avoids edge blurring. In image regions where the intensities vary more gradually, the weights are assigned small values to permit a larger residual. As a consequence, there is little deblurring in these regions and little noise magnification either. The weighting matrix R may also be used to account for the nonstationarity of the noise variance or for missing image data [14]. The ultimate restoration depends upon the specific properties of the restoration algorithm employed.

The smoothness requirement imposed on the restored image by (102) is locally adapted by the diagonal weighting matrix S in order to prevent ringing. Near sharp-intensity transitions the regularization is nearly disabled by assigning small values to the corresponding coefficients in the weighting matrix S. In smooth regions, these coefficients are made larger to guarantee noise suppression.

Following the same line of reasoning as in Section V-B, (101) and (102) can be combined into a single quadrature formula

$$\Phi(\hat{f}) = \|g - H\hat{f}\|_R^2 + \alpha\|C\hat{f}\|_S^2 \leq 2\epsilon^2. \quad (103)$$

$\Phi(\hat{f})$ can be minimized, subject to the condition that the solution has to be a member of the closed convex subset which is defined by the projection operator P by using an adaptive version of (98) [14].

$$\hat{f}_{k+1} = P[(I - \alpha\beta C^* SC)\hat{f}_k + \beta H^* R(g - H\hat{f}_k)]. \quad (104)$$

A sufficient condition to guarantee the convergence of this iteration is

$$0 < \beta < \frac{2}{\lambda_{\max}} \quad (105)$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix $(H^*RH + \alpha C^*SC)$. The adaptivity introduced by the weighting matrix R is seen to regulate the size of the *restoration term* $\beta H^* R(g - H\hat{f}_k)$. This concept was first introduced by Ichioka and Nakajima [3] who locally adapted the restoration process by changing the parameter β. Near edges β was made large to maximize restoration, while no restoration was achieved where β was made small. The weighting matrix R achieves the same effect.

The *stabilizing term* $(I - \alpha\beta C^* SC)\hat{f}_k$ represents a locally adaptive low-pass filter. For small values of s_{ij} (near edges) no filtering is done on $\hat{f}_k$, while heavy low-pass filtering is done where the coefficients in the S matrix are large. A related adaptive smoothing approach was proposed by Katsaggelos *et al.* [11], [15], [73]. Using ideas similar to those outlined in Section V-D, the stabilizing term is replaced by a single space-varying (edge-dependent) noise-smoothing filter $C_S(u, v; i, j)$. A measure of the local variance, $\sigma_f^2(i, j)$ is computed either from g or from a preliminary restoration. It is used to regulate this adaptive filter according to a noise-masking principle. For example, the adaptive version of the noise-smoothing Wiener filter can be used in this context (compare Eq. (94)) [11], [15].

$$C_S(m, n; i, j) = \frac{1}{1 + \left[\frac{1}{1 + b\sigma_f^2(i, j)}\right]^a \frac{S_{nn}(m, n)}{S_{ff}(m, n)}} \quad (106)$$

C. Examples

The first example considers the constrained adaptive restoration of the defocused cameraman image in Fig. 11(a). The weighting coefficients s_{ij} were tuned as described in [14], and the intensity values in the restored image were constrained to the intensity interval [10], [240]. The result of iteration (104) (Fig. 27) has an SNR improvement of 8.1 dB,

Fig. 27. Constrained adaptive restoration of the defocused cameraman image with noise added (SNR = 40 dB). SNR improvement is 8.1 dB.

compared with a maximum of 6.2 dB for the space-invariant restorations of Section IV.

The second example shows the combined effect of the two weighting matrices. The original cameraman image in Fig. 10 was blurred by horizontal motion with $L = 8$ and noise was added with SNR = 30 dB. Next, 50% of the pixels of this noisy blurred image were randomly discarded to simulate severe corruption of the image data. The resulting corrupted image is shown in Fig. 28(a), where all of the discarded pixels are given an intensity value of 0 (black). In the restoration process, using iteration (104), the corresponding coefficients r_{ij} were set to zero to exclude this erroneous data. The result of the restoration is presented in Fig. 28(b).

(a) (b)

Fig. 28. (a) Noisy motion blurred cameraman image with 50% of the data discarded. (b) Restoration.

VIII. Nonlinear Restoration

The deconvolution methods presented so far have all used a linear image formation and recording model in which the effects of any pointwise nonlinearities have been ignored. This section looks at the restoration of blurred images described by the more complicated nonlinear model which incorporated this nonlinearity. For both space-invariant and space-varying blurs, this model will be summarized as

$$g = s\{Hf\} + n. \tag{107}$$

Although the inversion of this relation is a nonlinear problem, it can still be tackled by the regularization methods described in the previous sections. To define the nonlinear (adaptive) restoration algorithm that makes use of this model, (101) is replaced by

$$\|g - s\{H\hat{f}\}\|_R \leq \epsilon. \tag{108}$$

Combining this with (102) into a single quadrature formula yields

$$\Phi(\hat{f}) = \|g - s(H\hat{f})\|_R^2 + \alpha\|C\hat{f}\|_S^2 \leq 2\epsilon^2. \tag{109}$$

As in Section VII-B, $\Phi(\hat{f})$ is minimized subject to the condition that the solution be a member of the convex set of feasible solutions. This yields the iteration

$$\hat{f}_{k+1} = P[(I - \alpha\beta C^*SC)\hat{f}_k + \beta H^* N_s\{H\hat{f}_k\}\, R(g - s\{H\hat{f}_k\})]. \tag{110}$$

Here $N_s(\cdot)$ is the diagonal Jacobian matrix defined by

$$N_s(h) = \operatorname{diag}\left(\left.\frac{\partial s}{\partial x}\right|_{x=h(1)}, \left.\frac{\partial s}{\partial x}\right|_{x=h(2)}, \cdots, \left.\frac{\partial s}{\partial x}\right|_{x=h(N^2)}\right). \tag{111}$$

The relaxation parameter β is chosen to ensure convergence. This iteration can be shown theoretically to converge only to a suboptimal solution, a property which is inherent to the nonlinear formulation of the problem. If the sensor nonlinearity $s\{\cdot\}$ is replaced or approximated by a linear function, which is justifiable in situations of low noise or low contrast [87], this iteration reduces to that in (104).

The estimation of f given (107) was first studied by Hunt [87] and later improved by Trussell and Hunt [88] using a probabilistic approach to the estimation problem. Using Gaussian processes to model the image and noise, a maximum *a posteriori* (Bayesian) estimate of the image, f was derived. As this estimator required solving a nonlinear equation, an iterative scheme was used. The following iteration solves (107) for the MAP estimator [87]

$$\hat{f}_{k+1} = (I - \beta R_f^{-1})\hat{f}_k + \beta R_f^{-1}\bar{f} + \beta H^t N_s\{H\hat{f}_k\} \cdot R_n^{-1}(g - s\{H\hat{f}_k\}). \tag{112}$$

Here $\bar{f}$ is the local (nonstationary) image mean and R_f and R_n are the covariance matrices of the image and noise, respectively. The form of these two iterations ((110) and (112)) are quite similar, and their major difference lies in the form of the two stabilizing terms, $(I - \alpha\beta C^*SC)\hat{f}_k$ and $(I - \beta R_f^{-1})\hat{f}_k$, respectively. To use (112) directly requires that the two covariance matrices be inverted. As these are nondiagonal, Fourier methods are used instead, which restricts the MAP estimator to space-invariant restoration. Because this restriction does not hold for (110), it is somewhat more general than (112). If the sensor nonlinearity $s\{\cdot\}$ is replaced by a linear function in (112), it can be shown that the limiting solution is given by the Wiener filter estimate with the *a priori* mean included [87].

To illustrate the nonlinear iteration (110), consider the restoration of noisy blurred *density* images, that is, the recorded data represent the densities of silver grains on developed photographic film. The relation between light intensities and silver density is governed by the Hurter-Driffield curve [35] (see Fig. 29), which can be approximated in

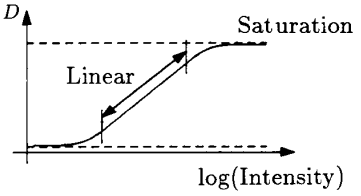


Fig. 29. The Hurter-Driffield curve.

the "linear" region by

$$D = \gamma \log E - D_0. \tag{113}$$

Here E is the total exposure, γ is the gamma of the film, and D_0 is an offset value. In addition to the above relation there is a region of fog where, even for very low exposures, some small amount of silver is deposited, and a region of saturation where all available silver has been deposited because of intense light.

Figure 30(a) shows such a density image of the defocussed cameraman. The image formation and recording process was simulated by first defocusing the original, next passing the defocussed image through a Hurter-Driffield curve, and finally adding noise to the result with an SNR of 40 dB. It is obvious that this image cannot be restored using any linear filter. The resulting restoration obtained by the nonlinear iteration in (110) is shown in Fig. 30(b). Although the iterations given in (110) and (112) are suitable for such complicated degraded images, nonlinear restoration remains a complex problem. For this reason algorithms of this type are not yet widely used.

IX. Algorithms with Faster Convergence

All of the iterative restoration procedures that we have considered to this point can be interpreted as the minimization of a functional using the iterative method of steepest descent [55], [56]. With this approach, if $\Phi(\hat{f})$ is the functional to be minimized, the algorithm proceeds by repeatedly moving in the direction of the negative of the

(a)

(b)

Fig. 30. (a) Nonlinearly degraded and defocused cameraman image. (b) Restoration.

gradient of $\Phi(\hat{f})$, denoted by $-\nabla\Phi(\hat{f})$. Thus

$$r_k = -\tfrac{1}{2}\nabla\Phi(\hat{f}_k) \quad \text{(steepest descent direction)} \tag{114}$$

$$\hat{f}_{k+1} = \hat{f}_k + \beta_k r_k. \tag{115}$$

The parameter β_k is known as the step size. It is well known, however, that methods of steepest descent converge very slowly. For this reason, the use of more efficient iterative methods, which converge more quickly, have been studied for both the 1-D [89]–[94] and 2-D [95]–[96] cases. This section examines several of these more efficient iterations.

A. Optimized Method of Steepest Descent

In (89), (98), (104), and (110) the convergence of the iterations was controlled by the parameter β, which was assumed to have a fixed value. However, because β also controls the rate of convergence, it is desirable to optimize its value at every iteration.

The optimal value of β_k for each iteration can be derived by minimizing $\Phi(\hat{f}_{k+1}(\beta_k)) = \Phi(\hat{f}_k + \beta_k r_k)$ at each iteration, that is, to choose that β_k which minimizes $\Phi(\hat{f}_{k+1})$ along the path $\hat{f}_{k+1} = \hat{f}_k + \beta_k r_k$. For the iterations that have been considered to this point with a *linear* projection operator P, this yields

$$\beta_k = \frac{(r_k, Pr_k)}{\|HPr_k\|_R^2 + \alpha\|CPr_k\|_S^2}. \tag{116}$$

For a nonlinear projection operator P and for the nonlinear restoration problem when the pointwise nonlinearity is included, such an explicit relation cannot be obtained. Then the common approach is to use a line search method to find a value for β which approximately minimizes $\Phi(\hat{f}_{k+1}(\beta_k))$ [55]. Although these optimization procedures for β_k obviously increase the convergence speed of the iterations, the improvements are usually moderate and may not justify the effort involved.

B. 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 [55]. 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. The expansion is generated by making use of information from previous iteration steps. This section focuses on the use of the conjugate gradients method in image deconvolution [89], [90], [92], [96]. This is the most important of the conjugate direction methods. One of the advantages of this method is its convergence in a finite number of iterations when exact arithmetic is assumed (superlinear convergence). When nonexact arithmetic is used or the problem is nonquadratic, this method will not converge in a finite number of steps because the conjugacy condition will no longer hold. It has been experimentally shown, however, that the conjugate gradients method exhibits a considerably higher convergence speed than the method of steepest descent.

The use of nonlinear constraints to represent *a priori* knowledge about the original solution is not consistent with the conjugate direction methods, but they have nonetheless been used in conjunction with them. The most simple and computationally efficient approach is to project the conjugate gradient iterates themselves after each iteration step [89], [96]. However, alternative, more complicated methods to incorporate nonlinear constraints, such as the gradient projection method [55], could be considered as well.

The (extended) conjugate gradients iteration, which thus represents an alternative to the iterations given in (89), (98), (104), and (110) is defined by

$$\begin{aligned} r_k &= -\tfrac{1}{2}\nabla\Phi(\hat{f}_k), \\ p_k &= r_k + \gamma_k p_{k-1}, \\ \hat{f}_{k+1} &= P[\hat{f}_k + \beta_k p_k]. \end{aligned} \tag{117}$$

Here p_k is called the *direction vector*, which depends upon the current steepest descent vector r_k and the preceeding vector p_{k-1}. The parameter γ_k regulates the conjugacy of the subsequent directions p_k. Observe that for $\gamma_k \to 0$, (117) reduces to (115). For the unconstrained minimization of the quadratic functionals (that is, P = Identity), it can be shown that the parameters β_k and γ_k are given by [55], [96]

$$\gamma_k = \frac{(r_k, r_k)}{(r_{k-1}, r_{k-1})} \tag{118}$$

$$\beta_k = \frac{(p_k, r_k)}{\|Hp_k\|_R^2 + \alpha\|Cp_k\|_S^2}. \tag{119}$$

Using the above equations, (117) reduces to the original formulation of a conjugate gradients algorithm. Clearly, by incorporating a projection operator P into the algorithm, the concept of an orthogonal solution decomposition can no longer hold. It has been shown, however, that in the practice of image restoration, the use of the previous direction vector p_{k-1} is useful in determining the current direction vector when the modifications made by the projection operator are relatively small. The choice for the values of γ_k and β_k becomes more difficult in this case as well. Usually a suitable choice for γ_k is still given by (118) or by a slightly

modified form [56]. The optimal value for β_k which minimizes $\Phi(\hat{f}_{k+1}(\beta_k))$ for a linear projection P is given by (119) with p_k replaced by its projected version Pp_k, while for all other situations, a line search method must be employed.

C. Iteration Methods 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 converge linearly [55]. It is, however, well known that there exist iterative methods with a convergence rate which is quadratic or higher. Recently, a number of papers have described iterative restoration schemes which exhibit Qth order convergence speed ($Q \geq 2$) [93]–[96]. These algorithms, which are suitable only for the unconstrained minimization of (80) and (103), are based on a Taylor expansion of the minimization problem, and are closely related to Newton-Raphson-like iteration methods. The basic form of these algorithms, which require a "double iteration," is given by

$$\begin{aligned} \hat{f}_0 &= \beta H^* R g, \\ B_0 &= I - \beta(H^*RH + \alpha C^*SC), \\ \hat{f}_{k+1} &= \sum_{j=0}^{Q_k - 1} B_k^j \hat{f}_k, \qquad (Q_k \geq 2), \\ B_{k+1} &= B_k^{Q_k} = \underbrace{B_k \cdot B_k \cdot B_k \cdots B_k}_{Q_k}. \end{aligned} \tag{120}$$

Here Q_k determines the convergence order in each iteration step. Sufficient conditions for this iteration are again given by (105).

The performance of this iteration can be derived by comparing the explicit expressions for the $(k+1)$st iterate of (120) with the one obtained from the standard unconstrained steepest descent iteration with a fixed value of β. For the $(k+1)$st iterate of (120), we can write ($Q_k = Q$)

$$\hat{f}_{k+1} = \sum_{i=0}^{Q^{k+1}-1} (I - \beta(H^*RH + \alpha C^*SC))^i \beta H^* R g, \tag{121}$$

while for the $(k+1)$st iterate of the steepest descent iteration, we get

$$\hat{f}_{k+1} = \sum_{i=0}^{k+1} (I - \beta(H^*RH + \alpha C^*SC))^i \beta H^* R g. \tag{122}$$

From (121) and (122) it can be seen that the two procedures compute exactly the same solution. However, the steepest descent algorithm requires $Q^{k+1} - 1$ iterations to obtain the same solution that (120) reaches after only $k+1$ iterations. The extra expense for the enormous reduction in the required number of iterations is more computations in a single iteration step, and extra memory required to store B_k. The efficiency of the iterations therefore depends strongly on the choice of the convergence order parameter Q and the way in which the algorithm has been implemented.

D. Examples

We consider the linear, space-invariant restoration of the defocused cameraman image in Fig. 11(a) using iterations (89), (117), and (120). Fig. 31 shows the restoration results, one for each algorithm, which differ very little from one another. The result in Fig. 31(a) was obtained after 4000 iterations of the basic steepest descent algorithm, the one in Fig. 31(b) was obtained after 80 iterations of the conjugate gradient algorithm, and Fig. 31(c) was obtained after only 12 iterations of iteration (120) with $Q = 2$.

X. Summary and Further Questions

This tutorial paper has discussed many recent developments in the field of iterative image deblurring. It has been shown that these iterative procedures are well suited to the image restoration problem. They can be used in a variety of problems, ranging from the most simple linear deconvolution of noiseless images to the problems of constrained, adaptive, and nonlinear restoration. The paper discussed the relationship between the basic Van Cittert iteration and its extensions with various other methods, including frequency domain filters, Kalman filters, and iterative methods which have a higher convergence speed.

While (iterative) image restoration has received maturity, the related blur identification problem is still open for further research. In order to make image restoration applicable to practical situations of interest, i.e., restoring images which have been subject to real blurs, the unknown blurs have to be estimated from the noisy blurred images themselves. Some new initiatives have appeared recently which tackle this problem in its most realistic form [42], [43], [97], [98]. It can be expected that a shift will occur from the pure image restoration problem toward the combined image identification and restoration problem in the coming decade.

(a) (b) (c)

Fig. 31. Three results obtained using different iterative algorithms. (a) 4000 iterations of the steepest descent algorithm. (b) 80 iterations of the conjugate gradient algorithm. (c) 12 iterations of the quadratically converging algorithm.

Appendix

A. Eigenvalue Analysis for 2-D Systems

In this appendix, we will briefly review some 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 [28], and the texts by Andrews and Hunt [22] and Gonzalez and Wintz [99].

Lexicographic Ordering: Let the 2-D array $\mathcal{F}$ represent a discrete image with M rows and N columns.

$$\mathcal{F} = \{f(i, j)\}, \quad 0 \le i \le M - 1, 0 \le j \le N - 1. \tag{122}$$

This data can be written in terms of a 1-D mapping that is known as lexicographic ordering. The $M \times N$ 2-D array is converted to a $MN \times 1$ vector f by concatenating its rows. Thus,

$$f^T = [f(0, 0), f(0, 1), \cdots, f(0, N - 1), f(1, 0), f(1, 1), \cdots, f(M - 1, N - 1)]. \tag{123}$$

Block-Toeplitz and Block-Circulant Matrices: If the pointwise nonlinearity and additive noise of the blurred image are ignored, we have the following superposition relation:

$$g(i, j) = \sum_{\forall(k,l)} h(i, j; k, l) f(k, l). \tag{124}$$

By lexicographically ordering both $f(i, j)$ and $g(i, j)$, we have

$$g = Hf, \tag{125}$$

where H is the blurring matrix, which is of size $MN \times MN$. (Observe that H will always be square if f and g are of the same size.) Although H may have a very complex structure, it is usually quite sparse and structured. Some relevant cases are discussed in [28]. Here we consider the case where the point-spread function $h(i, j; k, l)$ is space-invariant. Then (124) can be written as a 2-D linear convolution.

$$g(i, j) = \sum_{\forall(k,l)} h(i - k, j - l) f(k, l) = h(i, j) * f(i, j). \tag{126}$$

The matrix H now 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. Furthermore, the submatrices are arranged in a Toeplitz pattern. This is illustrated in Fig. 32.

The block-Toeplitz matrix H 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 two-dimensional discrete Fourier transforms. The errors introduced by the approximation are usually small. With a block-circulant matrix $H_j = H_{j-M}$ and the elements $h(j, k)$ of H_j are replaced by $h(j, k - N)$. In signal processing terms the approximation of the block-Toeplitz matrix by a block-circulant one converts a two-dimensional linear convolution into a two-dimensional circular convolution.

Eigensystem of a Matrix: A powerful tool in both the analysis and implementation of linear equations such as (124) is the set of eigenvalues and eigenvectors (eigensystem) of the matrices involved. Let $\{v_{mn}(i, j)\}$ denote the eigenvectors associated with the blurring matrix H and let the scalars

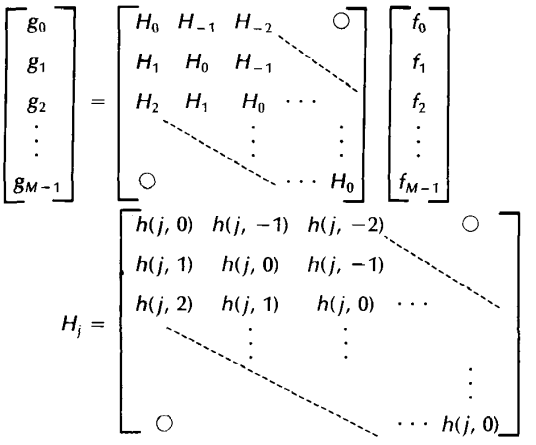


Fig. 32. Block Toeplitz structure of a space-invariant H. g_i and f_k refer to the jth row of g and the kth row of f, respectively.

$\{\lambda_{mn}\}$ represent the corresponding eigenvalues. That is,

$$Hv_{mn} = \lambda_{mn} v_{mn}, \quad 0 \le m \le M - 1, 0 \le n \le N - 1. \tag{127}$$

The double subscript notation is used because the eigenvectors, when unstacked, are in fact images. They are actually the set of images which are unaffected by the blurring operator except for a change of scale. Writing out the eigenvector equations

$$\sum_{\forall(k,l)} h(i, j; k, l) v_{mn}(k, l) = \lambda_{mn} v_{mn}(i, j), \quad 0 \le i, m \le M - 1 \quad 0 \le j, n \le N - 1. \tag{128}$$

If H and H^* have the same set of eigenvectors, the matrix H can be expressed in terms of its eigenvectors as

$$H = \sum_{m,n} \lambda_{mn} v_{mn} v^*_{mn} \tag{129}$$

which can also be written as

$$h(i, j; k, l) = \sum_{m,n} \lambda_{mn} v_{mn}(i, j) v_{mn}(k, l). \tag{130}$$

The advantage of using the eigensystem of a matrix is that linear relations, such as (124), can be evaluated as a set of independent equations. To this end we first decompose f in terms of the eigenvectors of H, that is,

$$f = \sum_{mn} (f, v_{mn}) v_{mn} \tag{131}$$

where (f, v_{mn}) denotes the inner product between f and the mnth eigenvector. Substituting (131) into (125) yields

$$g = Hf = H \sum_{m,n} (f, v_{mn}) v_{mn}, = \sum_{m,n} (f, v_{mn}) H v_{mn} = \sum_{m,n} (f, v_{mn}) \lambda_{mn} v_{mn}. \tag{132}$$

By also expanding g in terms of the eigenvectors of H, we get

$$\sum_{m,n} (g, v_{mn}) v_{mn} = \sum_{m,n} \lambda_{mn} (f, v_{mn}) v_{mn} \tag{133}$$

or

$$(g, v_{mn}) = \lambda_{mn}(f, v_{mn}), \qquad \forall m, n. \tag{134}$$

Thus, by virtue of using the eigensystem of H, we have replaced the superposition summation (124) by a far more simple set of scalar equations (134).

In general, it is hardly ever possible to find the eigenvectors $\{v_{mn}\}$ of an arbitrary matrix H of size $MN \times MN$ because of its size. This becomes possible only when H is highly structured. In the particular case where H is block-circulant (that is, $h(i, j; k, l)$ is spatially invariant), the eigenvectors are the complex exponentials

$$v_{mn}(k, l) = e^{-j2\pi((mk/M) + (nl/N))} \tag{135}$$

and the eigenvalues are equal to the DFT samples

$$\lambda_{mn} = \sum_{k,l} h(k, l)\, v_{mn}(k, l) \tag{136}$$

which can be evaluated efficiently using a 2-D FFT algorithm. Furthermore, the dot products (f, v_{mn}) and (g, v_{mn}) are samples of the 2-D DFTs of $f(i, j)$ and $g(i, j)$ respectively.

References

[1] R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained iterative signal restoration algorithms," *Proc. IEEE*, vol. 69, pp. 432–450, Apr. 1981.

[2] 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, pp. 1455–1465, Nov. 1983.

[3] Y. Ichioka and N. Nakajima, "Iterative image restoration considering visibility," *J. Opt. Soc. Amer.*, vol. 71, pp. 983–988, Aug. 1981.

[4] H. J. Trussell, "Convergence criteria for iterative restoration methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, pp. 129–136, Feb. 1983.

[5] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 201–212, Apr. 1984.

[6] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1—Theory," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 81–94, Oct. 1982.

[7] 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. MI-1, pp. 95–101, Oct. 1982.

[8] V. T. Tom, T. F. Quatieri, M. H. Hayes, and J. H. McClellan, "Convergence of iterative non-expansive signal restoration algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, pp. 1052–1058, Oct. 1981.

[9] 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, pp. 139–160, 1984.

[10] 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, July 1980.

[11] A. K. Katsaggelos, *Constrained Iterative Restoration Algorithms*, Ph.D. Thesis, Georgia Institute of Technology, 1985.

[12] A. Levi and H. Stark, "Image restoration by the method of generalized projections with applications to restoration from magnitude," *J. Opt. Soc. Am.*, vol. 74, pp. 932–943, Sept. 1984.

[13] M. I. Sezan, A. M. Tekalp, and C. T. Chen, "Regularized signal restoration using the theory of convex projections," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1565–1568, 1987.

[14] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Regularized iterative image restoration with ringing reduction," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-36, pp. 1874–1888, Dec. 1988.

[15] A. K. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "Nonstationary iterative image restoration," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1985, pp. 696–699.

[16] Y. Ichioka, Y. Takabo, K. Matsuoka, and T. Suzuki, "Iterative image restoration by a method of steepest descent," *J. Optics (Paris)*, vol. 12, pp. 35–41, 1981.

[17] G. Ross, "Iterative methods in information processing for object restoration," *Optica Acta*, vol. 29, no. 11, pp. 1523–1542, 1982.

[18] A. K. Katsaggelos and R. W. Schafer, "Iterative deconvolution using different distorted versions of an unknown signal," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1983, pp. 659–662.

[19] A. K. Katsaggelos, "A general formulation of adaptive iterative image restoration algorithms," *Proc. Princeton Conf. on Info. Sciences and Systems*, 1986, pp. 42–46.

[20] J. Maeda and K. Murata, "Restoration of band-limited images by an iterative regularized pseudoinverse method," *J. Opt. Soc. Amer. A*, vol. 1, pp. 28–34, Jan. 1984.

[21] R. J. Mammone and R. J. Rothacker, "General iterative method of restoring linearly degraded images," *J. Opt. Soc. Amer. A*, vol. 4, no. 1, pp. 208–215, Jan. 1987.

[22] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[23] J. Biemond, "Stochastic linear image restoration," in *Advances in Computer Vision and Image Processing*, T. S. Huang, Ed. London: JAI Press Inc., ch. 5, vol. II, pp. 502–528, 1986.

[24] J. W. Woods and V. K. Ingle, "Kalman filtering in two dimensions—Further results," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-29, pp. 188–197, Apr. 1981.

[25] B. R. Hunt, "The application of constrained least-squares estimation to image restoration by digital computer," *IEEE Trans. Computers*, vol. C-22, pp. 805–812, Sept. 1973.

[26] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. New York: Wiley, 1977.

[27] K. Miller, "Least squares methods for ill-posed problems with a prescribed bound," *SIAM J. Math. Anal.*, vol. 1, pp. 52–74, Feb. 1970.

[28] W. K. Pratt, "Vector space formulation of two-dimensional signal processing operations," *Computer Graphics and Image Processing*, vol. 4, pp. 1–24, 1975.

[29] A. K. Jain, "Advances in mathematical models for image processing," *Proc. IEEE*, vol. 69, pp. 502–528, May 1981.

[30] M. P. Ekstrom and J. W. Woods, "Two-dimensional spectral factorization with applications in recursive digital filtering," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-24, pp. 115–128, Apr. 1976.

[31] A. Habibi, "Two-dimensional Bayesian estimate of images," *Proc. IEEE*, vol. 60, pp. 878–883, July 1972.

[32] R. Chellappa, "Two-dimensional discrete Gaussian Markov random field models for image processing," in *Progress in Pattern Recognition 2*, L. N. Kanal and A. Rosenfeld, Eds. Dordrecht, The Netherlands: Elsevier Science Publishers, 1985.

[33] R. L. Kashyap, "Image models," in *Handbook of Pattern Recognition and Image Processing*, T. Y. Young and K. S. Fu, Eds. Orlando, FL: Academic Press, 1986.

[34] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Th.*, vol. IT-18, pp. 232–240, Mar. 1972.

[35] B. R. Hunt, "Digital image processing," *Proc. IEEE*, vol. 63, pp. 693–708, Apr. 1975.

[36] T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image processing," *Proc. IEEE*, vol. 59, pp. 1586–1609, Nov. 1971.

[37] M. Potmesil and I. Chakravarty, "Synthetic image generation with a lens and aperture camera model," *ACM Trans. Graph.*, vol. 1, pp. 25–108, Apr. 1982.

[38] P. A. Stokseth, "Properties of a defocussed optical system," *J. Opt. Soc. Am.*, vol. 59, pp. 1314–1321, Oct. 1969.

[39] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw-Hill, 1968.

[40] H. H. Hopkins, "The frequency response of a defocussed optical system," *Proc. Roy. Soc. Ser. A*, vol. 231, pp. 91–103, Feb. 1955.

[41] J. Bescos, I. Glaser, and A. A. Sawchuk, "Restoration of color images degraded by chromatic aberrations," *Applied Optics*, vol. 19, pp. 3869–3876, Nov. 1980.

[42] 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, Signal Processing*, vol. ASSP-34, pp. 963–972, Aug. 1986.

[43] J. Biemond, F. G. van der Putten, and J. W. Woods, "Identification and restoration of images with symmetric noncausal blurs," *IEEE Trans. Circ. Syst.*, vol. CAS-35, 385-393, Apr. 1988.
[44] T. G. Stockham, Jr., T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE*, vol. 63, pp. 678-692, Apr. 1975.
[45] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, ch. 10. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[46] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: M.I.T. Press, 1983.
[47] D. Graupe, D. J. Krause, and J. B. Moore, "Identification of autoregressive moving average parameters of time series," *IEEE Trans. Automatic Control*, pp. 104-106, Feb. 1975.
[48] A. Rosenfeld and A. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic Press, 1982.
[49] K. A. Dines and A. C. Kak, "Constrained least-squares filtering," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-25, pp. 346-350, Aug. 1977.
[50] J. Biemond, J. Rieske, and J. J. Gerbrands, "A fast Kalman filter for images degraded by both blur and noise," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-31, pp. 1248-1256, Oct. 1983.
[51] P. H. Van Cittert, "Zum Einfluss der Spaltbreit auf die Intensitätsverteilung in Spektrallinien II," *Z. Physik*, vol. 69, pp. 298-308, 1931.
[52] H. Bially, "Iterative Behandlung linearer Funktionalgleichungen," *Arch. Ration. Mech. Anal.*, vol. 4, pp. 166-176, July 1959.
[53] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," *Am. J. Math.*, vol. 73, pp. 615-624, 1951.
[54] 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, pp. 798-825, Sept. 1974.
[55] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.
[56] R. Fletcher, *Practical Methods of Optimization, Vol. 1: Unconstrained Optimization*. Chichester: Wiley, 1980.
[57] M. Z. Nashed, "Operator theoretic and computational approaches to ill-posed problems with applications to antenna theory," *IEEE Trans. Antennas Propagat.*, vol. AP-29, pp. 220-231, Mar. 1981.
[58] 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, pp. 107-124, Nov. 1983.
[59] E. R. Pike, J. G. McWhirter, M. Bertero, and C. DeMol, "Generalized information theory for inverse problems in signal processing," *Proc. Inst. Elec. Eng., Pt. F*, vol. 131, pp. 660-667, Oct. 1984.
[60] 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. Berlin: Springer-Verlag, 1978.
[61] A. Lannes, S. Roques, and M. J. Casonove, "Resolution and robustness in image processing: a new regularization principle," *J. Opt. Soc. Amer.*, vol. 4, pp. 189-199, Jan. 1987.
[62] N. B. Karayiannes and A. N. Venetsanopoulos, "Image restoration: a regularization approach," *Proc. 2nd Int. Conf. on Image Processing and its Applications*, 1986, pp. 1-5.
[63] H. Stark, Ed., *Image Recovery: Theory and Application*. Orlando: Academic Press, 1987.
[64] M. Z. Nashed, "Aspects of generalized inverses in analysis and regularization," in *Generalized Inverses and Applications*. New York: Academic Press, pp. 193-244, 1976.
[65] M. Z. Nashed, Ed., *Ill-posed Problems: Theory and Practice*. Dordrecht: Reidel, 1981.
[66] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. New Haven: Yale University Press, 1923.
[67] 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.
[68] 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.
[69] S. F. Gull and J. Skilling, "Maximum entropy method in image processing," *Proc. Inst. Elec. Eng. F.*, vol. 131, pp. 646-659, Oct. 1984.
[70] C. AuYeung, R. M. Mersereau, and R. W. Schafer, "Maximum entropy deconvolution," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1986, pp. 273-276.
[71] G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Math. Anal.*, vol. 14, pp. 651-667, Sept. 1977.
[72] D. Girard, "Practical optimal regularization of large linear systems," *Mathematical Modeling and Numerical Analysis*, vol. 20, pp. 75-87, 1986.
[73] A. K. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "A general formulation of constrained iterative restoration algorithms," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1985, pp. 700-703.
[74] J. F. Abramatic and L. M. Silverman, "Nonlinear restoration of noisy images," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. PAMI-4, pp. 141-149, Mar. 1982.
[75] H. E. Knutsson, R. Wilson, and G. H. Granlund, "Anisotropic nonstationary image estimation and its applications: Part I—Restoration of noisy images," *IEEE Trans. Commun.*, vol. COM-31, pp. 388-397, May 1983.
[76] K. Stewart and T. S. Duranni, "Constrained signal reconstructin—a unified approach," *Signal Processing III: Theories and Applications*, I. T. Young *et al.* Eds. Amsterdam: Elsevier North Holland, 1986.
[77] D. B. Sherman, K. A. Stewart, and T. S. Durrani, "The use of context in image restoration," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Proces.*, 1986, pp. 1473-1476.
[78] H. J. Trussell and M. R. Civanlar, "The initial estimate in constrained iterative restoration," *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Proc.*, Feb. 1983, pp. 643-646.
[79] L. M. Bregman, "The method of successive projection for finding a common point of convex sets," *Soviet Math. Doklady*, vol. 6, pp. 688-692, 1965.
[80] M. R. Civanlar and H. J. Trussell, "Digital image restoration using fuzzy sets," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-34, pp. 919-936, Aug. 1986.
[81] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta*, vol. 21, pp. 709-720, 1974.
[82] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circ. Syst.*, vol. CAS-22, pp. 735-742, Sept. 1975.
[83] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-28, pp. 672-680, Dec. 1980.
[84] M. I. Sezan and A. M. Tekalp, "Iterative image restoration with ringing suppression using POCS," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1988, pp. 1300-1303.
[85] G. L. Anderson and A. N. Netravali, "Image restoration based on a subjective criterion," *IEEE Trans. Systems, Man, Cybernetics*, vol. SMC-6, pp. 845-853, Dec. 1976.
[86] A. M. Tekalp, J. W. Woods, and H. Kaufman, "A multiple model algorithm for the adaptive restoration of images," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1983, pp. 832-835.
[87] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Computers*, vol. C-26, pp. 219-229, Mar. 1977.
[88] H. J. Trussell and B. R. Hunt, "Improved methods of maximum *a posteriori* restoration," *IEEE Trans. Computers*, vol. C-27, pp. 57-62, Jan. 1979.
[89] R. Marucci, R. M. Mersereau, and R. W. Schafer, "Constrained iterative deconvolution using a conjugate gradient algorithm," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1982, pp. 1845-1848.
[90] R. Prost and R. Goutte, "Discrete constrained iterative deconvolution with optimized rate of convergence," *Signal Processing*, vol. 7, pp. 209-230, Dec. 1984.
[91] C. L. Yeh and R. T. Chin, "Constrained optimization for image restoration using nonlinear programming," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1985, pp. 676-679.

[92] T. K. Sarkar, F. I. Tseng, S. A. Dianat, and B. Z. Hollmann, "Deconvolution by the conjugate gradient method," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Mar. 1985, pp. 445–488.
[93] S. Singh, S. N. Tandon, and H. M. Gupta, "An iterative restoration technique," *Signal Processing*, vol. 11, pp. 1–11, July 1986.
[94] C. E. Morris, M. A. Richards, and M. H. Hayes, "Iterative deconvolution algorithm with quadratic convergence," *J. Opt. Soc. Amer. A*, vol. 4, pp. 200–207, 1987.
[95] C. E. Morris, M. A. Richards, and M. H. Hayes, "Fast reconstruction of linearly distorted signals," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-36, pp. 1017–1025, July 1988.
[96] R. L. Lagendijk, R. M. Mersereau, and J. Biemond, "On increasing the convergence rate of regularized iterative image restoration algorithms," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1987, pp. 1183–1186.
[97] R. L. Lagendijk, D. L. Angwin, H. Kaufman, and J. Biemond, "Recursive and iterative methods for image identification and restoration," *Proc. 4th European Signal Processing Conf. EUSIPCO-88*, pp. 235–238.
[98] R. L. Lagendijk, J. Biemond, and D. E. Boekee, "Blur identification using the expectation-maximization algorithm," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1989, pp. 1397–1400.
[99] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1987.

Jan Biemond (Senior Member, IEEE) was born in De Kaag, The Netherlands, on March 27, 1947. He received the M.S. and Ph.D. degrees from Delft University of Technology, Delft, The Netherlands, in 1973 and 1982, respectively.

He is currently Professor in the Laboratory for Information Theory of the Department of Electrical Engineering at the Delft University of Technology. His research interests include multidimensional signal processing, image enhancement and restoration, data compression of images, and motion estimation with application in image coding and computer vision. He has authored and co-authored more than 40 papers in these fields. In 1983 he was Visiting Researcher at Rensselaer Polytechnic Institute, Troy, NY, and at the Georgia Institute of Technology, Atlanta.

Dr. Biemond is a member of the IEEE-ASSP Technical Committee on Multidimensional Signal Processing and of the IEEE-CAS Technical Committee on Visual Signal Processing and Communication. He served as the General Chairman of the Fifth ASSP/EURASIP Workshop on Multidimensional Signal Processing, held at Noordwijkerhout, The Netherlands, in September 1987. He is Co-Editor of the *Int. Journal on Multidimensional Systems and Signal Processing*.

Reginald L. Lagendijk (Student Member, IEEE) was born in Leiden, The Netherlands, on April 22, 1962. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Delft University of Technology, The Netherlands, in 1985 and 1990, respectively.

He is currently on the Faculty of the Department of Electrical Engineering at the Delft University of Technology. His research interests include information theory and multidimensional signal processing, with emphasis on image identification and restoration, and data compression of image sequences.

Russell M. Mersereau (Fellow, IEEE) received the S.B. and S.M. degrees in 1969 and the Sc.D. in 1973 from the Massachusetts Institute of Technology.

He was a Research Associate at M.I.T. before joining the School of Electrical Engineering at Georgia Tech in 1975.

Professor Mersereau's current research interests are in the development of algorithms for the enhancement, modeling, and coding of computerized images, and in computer vision. In the past this research has been directed to problems of digital filter design, the reconstruction and estimation of distorted signals from partial information of those signals, computer image processing and coding, and effect of image coders on human perception of images, and applications of digital signal processing methods in speech processing, digital communications, and pattern recognition.

He has served on the Editorial Board of the *Proceedings of the IEEE* and as Associate Editor for signal processing of the *IEEE Transactions on Acoustics, Speech, and Signal Processing*, and as a member of the Steering Committee of the *IEEE Transactions on Medical Imaging*. He has published extensively and is the coauthor of the text *Multidimensional Digital Signal Processing*. He is a co-recipient of the 1976 Browder J. Thompson Memorial Prize of the IEEE for the best technical paper by an author under the age of 30, a recipient of the 1977 Research Unit Award of the Southeastern Section of the ASEE, and departmental teaching awards, and the 1990 Signal Processing Society Award.