Image Reconstruction and Restoration: Overview of Common Estimation Structures and Problems

GUY DEMOMENT

Abstract—**Developments in the theory of image reconstruction and restoration in the past 20 or 30 years are outlined. Particular attention is paid to common estimation structures in the field and to practical problems not properly solved yet.**

I. Introduction and Outline

RESTORATION and reconstruction of images can be defined as the general problem of estimating a two-dimensional (2-D) object from a degraded version of this object. The mathematical form of the degradation process depends on the problem at hand. In image restoration, the unknown object and its degraded observed version, referred to as the image, are both scalar functions defined on $\mathbb{R}^2$ or $\mathbb{Z}^2$. In image reconstruction, observations result from the interaction between the unknown object and some scattering wave. The 2-D object must be reconstructed from a finite set of "projections," i.e., scalar functions defined on $\mathbb{R}$ or $\mathbb{Z}$.

This problem has been extensively studied for its obvious practical importance as well as its theoretical interest. Literature on the subject is abundant and highly varied since the problem arises in almost every branch of engineering and applied physics. It is, for instance, frequently encountered in various fields of application such as optics, X-ray or diffraction tomography, radioastronomy, biomedical engineering, machine vision, nondestructive evaluation, geophysics, etc. The variety of reference sources quoted in this paper is evidence for this fact.

Hence, the aim of the paper is not to give an exhaustive overview of the literature, which would be too ambitious, but to stress two major points.

First, most existing image reconstruction and restoration methods have a common estimation structure in spite of their apparent variety. This can be assumed from the literature [8], [12], [19], [24] and was clarified a few years ago in a series of seminal papers by Titterington [22], [23]. Everything was summed up in a single word: regularization. Since then, changes have occurred in the field. The common theoretical limitations presented by these methods have led to a broadening of their theoretical foundations, comparable to other recent changes occurring in the whole signal processing field. The same phenomenon occurred recently in the area of image processing, particularly with the introduction of random Markov fields and related estimation methods [80]–[89]. However, the same common structures are still present and can be used as a unifying framework for understanding this rather complex field, which is rapidly evolving through the influence of statisticians, electrical engineers, and machine vision experts.

Second, in addition to a common estimation structure, most image reconstruction and estimation methods present some common practical limitations. The most sophisticated methods require the use of several tuning parameters (referred to as hyperparameters) which should be estimated from the data. Methods are available to perform this task. Some of them are reputed to give correct results, others not, and a great effort should be made to clarify this aspect.

In order to set the framework and make these points, the paper is organized as follows.

In Section II, the problem of image reconstruction and restoration is formulated. Although it reduces in most practical situations to solving a system of linear equations, this special case is important because it provides an easy way of studying the ill-posed or ill-conditioned nature of these inversion problems.

Section III describes some of the current regularization approaches used to solve this problem. The important concepts of *a priori* information and compound criterion are introduced.

In Section IV, we give a Bayesian interpretation of these regularization techniques which clarifies the role of the tuning parameters and gives indications on how they could be estimated.

Sections V and VI are devoted to the practical aspects of computing the solution, first when the hyperparameters are known, and second when they must be estimated.

In Section VII, conclusions are drawn and points that still need to be investigated are outlined.

The statistical approach taken in the paper allows us to cover most probabilistic methods, and a number of deterministic ones as well. However, due to the wide range of the subject and space limitations, some of them, although useful in some situations, have been left out. Justice is done to them in the references which are organized under nine subheadings roughly corresponding to the sections of

Manuscript received March 17, 1989.

The author was with the Laboratoire des Signaux et Systèmes (CNRS/ESE/UPS), École Supérieure d'Électricité, 91192 Gif-sur-Yvette Cédex, France. He is now with the Department of Physics, Université de Paris-Sud, Orsay, France.

IEEE Log Number 8931343.

0096-3518/89/1200-2024$01.00 © 1989 IEEE

the paper. Although the division is somewhat arbitrary, we hope it will enable the reader to acquaint himself with results that could not be covered in this paper.

II. The Ill-Posed Nature of Image Reconstruction and Restoration

An image is generally defined as a real or complex-valued function of two space variables belonging to some support region. Although this support may be continuous, it is commonly sampled on a rectangular grid. This defines a set of pixels, and the image is represented by a vector $\boldsymbol{x}$ of N pixel intensity values, N being typically a power of two and also a huge number.

In image reconstruction and restoration problems, the original "object" cannot be directly measured and must be either reconstructed or restored from the observed data in order to remove the effects of the observation mechanism.

In image restoration, the cause of degradation is some distortion which can be modeled as

$$\boldsymbol{y} = A(\boldsymbol{x}) \bullet \boldsymbol{b} \tag{2.1}$$

with the following interpretation: $A(\cdot)$ is a degradation mechanism, $\boldsymbol{b}$ denotes a corruptive noise process, and the symbol $\bullet$ represents a pixel-by-pixel interaction. The distortion mechanism often involves convolution or blurring of $\boldsymbol{x}$ by some point spread function and the addition of an independent Gaussian white noise, with zero mean and variance σ_i^2, thus leading to the following simpler model:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}. \tag{2.2}$$

Sometimes, however, the distortion also involves a nonlinear pixel transformation and a multiplicative noise. The problem of restoring $\boldsymbol{x}$ is then much more difficult [1], [6], [99], [114], even though some methods developed in the previous case can still be formally applied. In model (2.2), matrix $\boldsymbol{A}$ and the statistical characteristics of the noise are implicitly assumed to be known. But this assumption is not always fulfilled, and these quantities may also have to be estimated [72], [95], [98], [102]. This is, for instance, a common situation in geophysics. In this paper, we will deal only with the linear, additive, and Gaussian case with a known distortion model. Although this is somewhat restrictive, the corresponding problem is generic in the sense that its solution is the basis of many other ones.

The problem of image reconstruction is a little more complicated since the true object $\boldsymbol{x}$ is no longer a measure of light intensity over some scene, but a mapping of some physical property, e.g., a density of matter in a physical object, an acoustical impedance, a complex permittivity [3], [5], etc. The true object must be reconstructed, i.e., decoded from wave interaction results that are not images. These data are commonly called "projections," even when, because of diffraction effects, for instance, they are not projections of the object in a geometrical sense, i.e., line integrals, or when they are simply object samples in the space or frequency domains. In most existing methods, the image reconstruction problem is also modeled as that of solving a system of linear equations. When the object is illuminated by a radiating source with a very short wavelength, as in X-ray tomography, the data $\boldsymbol{y}$ are explicit and linear functions of the object $\boldsymbol{x}$. But when diffraction effects cannot be neglected, this is only a rough approximation which corresponds to a first-order expansion of the solution to the corresponding wave propagation equations.

The whole set of data is obtained by rotating the object in the incident radiated field and by measuring its projections at various incidence angles belonging to a discrete finite set. All the data are then concatenated in a single vector $\boldsymbol{y}$, thus leading to a model similar to (2.2):

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}. \tag{2.2}$$

In both cases, our image processing problem is to obtain an estimate $\hat{\boldsymbol{x}}$ of the true object $\boldsymbol{x}$ from the data vector $\boldsymbol{y}$. Under our assumptions, a solution which emerges immediately is the least squares one with minimum norm, which is

$$\hat{\boldsymbol{x}}_0 = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{y} \tag{2.3}$$

when the normal matrix $\boldsymbol{A}^t\boldsymbol{A}$ is regular or else the generalized inverse solution

$$\hat{\boldsymbol{x}}_0 = \boldsymbol{A}^\dagger\boldsymbol{y}. \tag{2.4}$$

This seems to be a reasonable choice, from a statistical standpoint at least, since under our assumptions, $\hat{\boldsymbol{x}}_0$ is an unbiased and minimum variance solution.

Although the effective computation of $\hat{\boldsymbol{x}}_0$ can be made computationally tractable in actual problems with huge dimensions, this solution is usually unacceptable since $\boldsymbol{A}$ is most often ill conditioned. The noise component amplification exceeds any acceptable level.

The ill conditioning is a direct consequence of the ill posedness of the initial continuous-data problem which is approximated by (2.2). In the restoration problem, the image is a convolution integral

$$y(s, s') = \int\int_D h(s - r, s' - r')\, x(r, r')\, dr\, dr' \tag{2.5}$$

which is a particular case of a Fredholm equation of the first kind. The kernel of this integral equation h is the 2-D impulse response or point spread function (PSF) of the imaging system. Since the data are erroneous or noisy, we cannot expect to solve this equation exactly and the true solution must be approximated in some sense. The concept of distance between images is then a natural way of evaluating the quality of an approximation, which explains why x and y are often assumed to belong to Hilbert spaces. This continuous-data problem can then be rewritten as

$$y = Ax \tag{2.6}$$

where y and x are now elements of infinite dimension function spaces X and Y, respectively, and A is a linear

operator corresponding to an integral equation of the first kind. To be acceptable, the solution must satisfy three conditions: the usual mathematical conditions of existence and uniqueness, and the less usual physical condition of stability or continuity with respect to the data since these data are inevitably noisy. These are the so-called Hadamard conditions for a problem to be well posed [9]–[18].

Let y and x belong to the same Hilbert space and let the PSF be square integrable. This assumption is fulfilled by most physical PSF's. The *direct problem*, i.e., the calculation of the imaging system response y to the true object x, is then well posed in the Hadamard sense: a small error δx on the data leads to a small error δy on the solution. However, this condition is not satisfied in the corresponding *inverse problem* where the true object x is to be computed from the response y:

$$x = A^{-1}y. \tag{2.7}$$

In fact, the necessary and sufficient conditions of existence, uniqueness, and stability of the solution are, respectively,

$$y \in \mathrm{Im}\,(A) \tag{2.8}$$

$$\mathrm{Ker}\,(A) = \{0\} \tag{2.9}$$

$$\mathrm{Im}\,(A) = \overline{\mathrm{Im}\,(A)} \tag{2.10}$$

where Im (A) and Ker (A) are the range and the null space of A, respectively, and where $\overline{\mathrm{Im}\,(A)}$ is the closure of Im (A). When the PSF h is square integrable, the Riesz–Frechet theorem indicates that the operator A is bounded and compact [12]. But the range of a compact operator is not closed (except in the degenerated case where its dimension is finite, which corresponds to a separable PSF). This means that the inverse operator A^{-1} is unbounded, its range Im (A) is unclosed, and the third Hadamard condition is not met for the inverse problem.

These difficulties remain unchanged in the symmetric problem

$$A^*y = A^*Ax \tag{2.11}$$

where A^* is the adjoint operator of A, i.e., $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for all $x \in X$ and $y \in Y$. $\langle \cdot, \cdot \rangle$ represents the scalar products used in the image and object spaces.

Additional light can be shed on these problems using a spectral approach. Equation (2.11) is used since A^*A is a self-adjoint, nonnegative definite operator, and hence has an eigensystem.

Let $\{\lambda_n\}$ denote the eigenvalues of A^*A (and of AA^*) ordered and counted with their multiplicity: $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq \cdots 0$. According to the Hilbert–Schmidt theorem [12], λ_n goes to zero as $n \to \infty$, and the alternative is either the limit is reached for $n = n_0$ or the limit is never reached for any finite value of n. Let E be the set of indexes n such that $\lambda_n \neq 0$, and let $\{u_n, v_n; \sigma_n\}$ be a singular system of operator A. The singular vectors u_n, v_n and the singular values σ_n satisfy the following conditions:

$$\begin{aligned} \sigma_n^2 &= \lambda_n & n &\in E \\ AA^*u_n &= \lambda_n u_n & A^*Av_n &= \lambda_n v_n \\ Av_n &= \sigma_n u_n & A^*u_n &= \sigma_n v_n. \end{aligned} \tag{2.12}$$

Equation (2.11) has a solution for a given image $y \in Y$ if and only if

$$\sum_{n\in E} \frac{1}{\sigma_n^2} \left|\langle y, u_n\rangle\right|^2 < +\infty. \tag{2.13}$$

Condition (2.13) is derived from Picard's criterion for solvability of an integral equation of the first kind [12]. In order for (2.13) to be fulfilled, components $\langle y, u_n\rangle$ of the image expansion on the set of eigenfunctions $\{u_n\}$ have to decrease faster than the eigenvalues σ_n^2 when $n \to \infty$. Since, in practice, these functions are rapidly oscillating when n increases, this condition can be intuitively interpreted as a limitation on the high-frequency content of the perfect image y.

In addition, when the existence and the uniqueness are assumed, a perturbation or noise δy on the data gives the perturbed or noisy solution [12]

$$x + \delta x = \sum_{n\in E} \frac{1}{\sigma_n} \langle y, u_n\rangle \left[1 + \frac{\langle \delta y, u_n\rangle}{\langle y, u_n\rangle}\right] v_n. \tag{2.14}$$

This shows that the error in each component of this expansion is proportional to the inverse signal-to-noise ratio $\langle \delta y, u_n\rangle / \langle y, u_n\rangle$. This is a first source of degradation since this ratio may often be much greater than unity when the noise δy and perfect image y have distinct spectral properties. However, even when the signal-to-noise ratio remains high over the whole spectrum $\{u_n\}$, the error components $\langle \delta y, u_n\rangle$ are divided by the singular values which go to zero as $n \to \infty$. Hence, small errors in the image induce large errors in the solution. This explains why usual trial-and-error methods cannot be applied to solve these integral equations of the first kind since a small observation error would not ensure a small error on the solution

$$\|\delta y\| \to 0 \not\Rightarrow \|\delta x\| \to 0. \tag{2.15}$$

This situation frequently occurs in inverse problems involving different function spaces for the object and the observed data, and careless application of inverse filters may lead to unacceptable results. In reconstruction, the solution is even more complicated since the uniqueness of the solution may also become a difficult problem [3].

In the discrete case, $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to finite dimensional spaces and the linear operator A is a matrix $\boldsymbol{A}$. Equation (2.2) has a unique solution of minimal norm $\hat{\boldsymbol{x}}_0 = \boldsymbol{A}^{\dagger}\boldsymbol{y}$ which depends continuously on $\boldsymbol{y}$ since the generalized inverse $\boldsymbol{A}^{\dagger}$ is always bounded [12]. The problem is then well posed in the least squares sense. The advantage of adopting this generalized inverse solution is that an inconsistent finite system of linear algebraic equations will not be ill posed in this sense, while it is ill posed in the sense of Hadamard.

But even in this setting, the inversion problem has, from a numerical viewpoint, instability properties akin to those of ill-posed problems. The spectral decomposition (2.14) is still valid, with the only difference being that the number of singular values of the normal matrix A^tA is now finite and equal to N. These singular values cannot be expressed in closed form, even when the normal matrix is Toeplitz. However, in the latter case, A^tA can be approximated by a circulant matrix whose eigenvectors are the columns of the discrete Fourier transform matrix [108]. The spectral decomposition (2.14) is then a Fourier spectrum with the usual spatial frequency interpretation. In the circulant case, the eigenvalues of A^tA are the PSF Fourier transform values computed at regularly spaced frequencies. When the degradation is a low-pass filter, σ_n take on very small values when $n \to N$. Even if we exclude the zero singular values, as in computing A^+, there will still be singular values close to zero. The matrix A is ill conditioned. The weight $\sigma_n^{-1} \langle \delta y, u_n \rangle$ in (2.14) can thus become very large for those σ_n close to zero, even when δy is small. We are in a somewhat paradoxical situation: the finer the pixel grid (or equivalently, the better the discrete approximation), the better the approximation of the singular values by the PSF, and the worse the conditioning of matrix A!

Problems of this kind arise not only in image processing, but in the whole applied physics field. A general principle for dealing with the instability of the problem is that of *regularization* [9]–[18]. It consists mainly of changing our idea of what a solution is. Obtaining the *true solution* from imperfect data is impossible. When regularizing the problem, this fact is acknowledged and the initial equation is considered to define, in fact, only a class of *admissible solutions* $\{\hat{x}\}$:

$$\{\hat{x} : \| y - Ax \| \leq \| b \|\} \tag{2.15}$$

among which an *acceptable solution* must be sought. But to do this, the problem must be stated more completely, which implies that some extra or *a priori information* be included.

As mentioned in the Introduction, some methods do not fall into the regularization framework (e.g., homomorphic filtering [1]). However, most of them do, and since regularization is conceptually very simple and intuitively natural, and since it can be given a conventional as well as a Bayesian statistical interpretation, it can be used as a unifying framework for this rather complex field of image processing.

III. Regularization of an Ill-Posed Problem

Numerous methods have been proposed for solving and regularizing various types of ill-posed problems [1]–[8], [12]. They can be separated into two categories: regularization in functional spaces, and control of dimensionality. The framework proposed here belongs to the first category. The basic feature is the introduction of a compromise between fidelity to the data and fidelity to some prior information about the solution. This compromise is measured with a single optimality criterion and can be physically justified as follows.

The least squares solution $\hat{x}_0$ is the minimizer of the total energy of residual error between the model Ax and the data y. In this sense, it provides maximum fidelity to the data. But for a wide-band observation noise, the energy of the restored or reconstructed object at high spatial frequencies is high, mainly due to the noise. In general, this solution $\hat{x}_0$, although unbiased, is rejected since the true image is expected to be significantly smoother. Thus, some infidelity to the data must be introduced in order to obtain a smoother solution. The question which immediately arises is, how can we get a smoother solution? In order to succeed in suppressing the noise without distorting the original image too much, we need information on the spectral content of the image. A simple and very natural way of performing such smoothing or regularization is to define two measures of "distance" $J_1(x, \hat{x}_0)$ and $J_2(x, \hat{x}_\infty)$ between x and two extreme pictures $\hat{x}_0$ and $\hat{x}_\infty$. $\hat{x}_0$ is the ultrarough least squares solution, and $\hat{x}_\infty$ corresponds to an *a priori* ultrasmooth object. We then balance the fidelity of the solution to the data, which is measured by J_1, and the fidelity to some prior information, which is measured by J_2. Usually, $\hat{x}_\infty$ turns out to be a picture of uniform intensity, but it may also be some reference map based on prior knowledge [1], [72]. A regularized solution $\hat{x}(\mu, y)$ is simply defined as the solution of the following problem:

$$\hat{x}(\mu, y) = \arg \min_{x \in X} \left\{ J_1(x, \hat{x}_0) + \mu J_2(x, \hat{x}_\infty) \right\}. \tag{3.1}$$

Note that $\hat{x}$ depends on y through $\hat{x}_0$ only. The choice of measures J_1 and J_2 defines a particular path between $\hat{x}_0$ and $\hat{x}_\infty$ in the set X of all possible objects [22]. This is a qualitative choice that determines the manner in which regularization is done. On the other hand, the choice of μ, which is the regularization parameter, is quantitative and allows the user to decide how far to go along this path to achieve an appropriate degree of smoothing. Perfect fidelity to the data is achieved if $\mu = 0$, whereas perfect fidelity to the priors is achieved if $\mu = \infty$.

Mathematical conditions on J_1 and J_2 are weak. Actually, J_1 and J_2 do not even need to be true distances in the usual mathematical sense. They simply need to be positive measures which vanish only if their arguments are identical. They may not be symmetric. Since J_1 and J_2 measure different properties of the solution, they need not be the same and, in fact, they are usually different. But the surface mapped out by the compound criterion $J_1(x, \hat{x}_0) + \mu J_2(x, \hat{x}_\infty)$ has to be a straightforward convex bowl with a unique lowest point to yield a unique solution $\hat{x}(\mu, y)$.

Although the scheme presented above allows us to cover most stochastic and deterministic restoration and reconstruction methods, two popular approaches do not fit well into this framework. They are briefly outlined below.

The first one belongs to the same functional regularization category and uses compactness and *a priori* bounds.

For some problems, the prior information available on the object can be expressed as a set of constraints such as *a priori* bounds, each defining a convex set C_i in the solution space X. An example of such a convex set is given by (2.16). It has long been known that restriction to a compact set ensures well posedness [12]. Problems of this type are solved using the method of projections onto convex sets (POCS) [117]–[119], [121]. However, this method does not yield a unique solution since all elements of the intersection of all the constraints sets, when nonempty, are equally acceptable. An iterative search for the solution depends on initial conditions, which is hardly satisfactory. To remedy this drawback, and also to exploit the statistical properties of the noise, an alternative was recently introduced [115] as a minimization of some distance to the data, such as J_1 under the set of previous constraints C_i:

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg\min_{x \in C} J_1(\boldsymbol{x}, \hat{\boldsymbol{x}}_0) \qquad C = \bigcap_i C_i. \quad (3.2)$$

This does yield a unique solution, but clearly differs from our scheme.

The second family of methods is the well-known truncated singular value decomposition (TSVD) [1] where regularization is performed through control of the dimensionality of the solution space. The summation in (2.14) is taken on a limited number of singular values λ_n, $n \geq p$, which significantly differ from zero, in order to avoid excessive noise contamination. This is equivalent to projecting the solution onto a "significant" subspace spanned by the remaining singular vectors v_n, $n = 1, 2, \cdots, p$. This method yields solutions that are numerically well conditioned. In most practical situations, the singular values decrease slowly when $n \to N$ and the truncation order p must be chosen by comparing the norms of the noise and of the residual errors [cf. (2.16)]. But the fundamental difference, and in some sense limitation, with other regularizing methods lies in the fact that the *a priori* information depends on the distorting system and not on the solution itself. TSVD regularization means that the solution is limited exclusively by the properties of the observation device, more precisely, by the width of the singular value spectrum. Such a conclusion clearly contradicts the principles of superresolution techniques. To remedy this, weighted singular value decompositions have been introduced [12], [18]. However, these have no simple functional interpretation and are far less popular than TSVD.

Despite the generality of (3.1), comparatively few distance measures have been used in image reconstruction and restoration. To our knowledge, J_1 and J_2 have always been chosen from the same limited set of candidates, which is listed below.

A. Quadratic (or Euclidean) and Weighted Quadratic Distances

A Euclidian distance between two pictures $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is, of course,

$$J_Q(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{i=1}^{N} (x_{1i} - x_{2i})^2 = (\boldsymbol{x}_1 - \boldsymbol{x}_2)^t(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \quad (3.3)$$

This is a particular case of weighted quadratic distances

$$J_{WQ}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_{1i} - x_{2i})(x_{1j} - x_{2j}) = (\boldsymbol{x}_1 - \boldsymbol{x}_2)^t \boldsymbol{W}(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_W^2 \quad (3.4)$$

where $\boldsymbol{W} = \text{diag}\{w_{ij}\}$ is a symmetric, positive definite matrix which is designed to outline certain special features of the desired proximity.

A distance of this type is the common choice for J_1 in most regularization schemes when noise $\boldsymbol{b}$ is assumed to be white, Gaussian, zero mean, and independent from $\boldsymbol{x}$:

$$J_{WQ}(\boldsymbol{x}, \hat{\boldsymbol{x}}_0) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[y_i - (\boldsymbol{A}\boldsymbol{x})_i\right]^2 = (\boldsymbol{x} - \hat{\boldsymbol{x}}_0)^t \boldsymbol{A}^t \Sigma^{-1} \boldsymbol{A}(\boldsymbol{x} - \hat{\boldsymbol{x}}_0) \quad (3.5)$$

where $\Sigma = \text{diag}\{\sigma_i^2\}$ and $\boldsymbol{W} = \boldsymbol{A}^t\Sigma^{-1}\boldsymbol{A}$.

B. Kullback Distances

A very important feature of X, the class of real world objects, is that in many picture processing problems, the pixel intensity is mainly nonnegative. It is highly desirable that this important physical property be preserved in the restored or reconstructed object. There is a variety of ways for achieving this, one of them being to consider that the object may be identified, after proper normalization, to a probability distribution, and then to use distance measures between such distributions. The Kullback distance is often used [64]–[79]:

$$J_K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{i=1}^{N} x_{1i} \log(x_{1i}/x_{2i}). \quad (3.6)$$

Note that this distance measure is nonsymmetric with respect to its arguments and, in this sense, pathological. This distance is the negative of the relative entropy of distribution $\boldsymbol{x}_1$ with respect to prior distribution $\boldsymbol{x}_2$.

C. Roughness Measures

A simple way of measuring the roughness of an image is to apply some finite difference operator to it and then take the Euclidian norm of the resulting differentiated image. Since such a differentiation operation is linear with respect to the original image, the resulting measure of roughness is quadratic:

$$\Omega(\boldsymbol{x}) = \|\nabla^k(\boldsymbol{x})\|^2 = \|\boldsymbol{D}_k\boldsymbol{x}\|^2 = \boldsymbol{x}^t\boldsymbol{D}_k^t\boldsymbol{D}_k\boldsymbol{x} \quad (3.7)$$

where k is the order of the difference operator $\nabla^k(\cdot)$. Usually, $k = 1$ or 2 [26]. $\Omega(\boldsymbol{x})$ is minimized if all the components of $\boldsymbol{x}$ are equal. Clearly, this is a special case

of weighted quadratic distance where the second argument x_2 is any uniform image and the weighting matrix is $W = D_k' D_k$.

D. Local Energy Functions

All previous distances measure some global property of an image which is implicitly considered as a homogeneous field. But real images are often inhomogeneous, with nearly uniform regions separated by relatively sharp edges. To break off the homogeneity, local energy functions have been recently introduced [80]–[86]. In this approach, the original image x is regarded as a pair $x = (z, t)$ where z is the vector of observable pixel intensities and t denotes a dual vector of unobservable additional features such as edges, textures, etc. To express local properties of the image, a neighborhood system $\{d_i\}$ is defined where d_i is a collection of pixels that are assumed to interact directly with pixel i. As an elementary example, let us suppose that the intensity values of x are discrete. A basic characteristic of most images is that intensity values at nearby locations are likely to be nearly the same. To express this property, we define

$$\begin{aligned} V(z_i, z_j) &= -1 \quad z_i = z_j \quad j \in d_i \\ &= 1 \quad z_i \neq z_j \quad j \in d_i \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{3.8}$$

These bond energies, computed among neighboring pixels, are added to define the energy of the image:

$$J(z) = \sum_i \sum_j V(z_i, z_j) = \sum_i V_i(x) \tag{3.9}$$

where V_i is the potential of the ith set of neighbor pixels and where i ranges over the set of all sites which are pairs of neighbors. The idea is that $J(z)$ is small for pictures consistent with the properties we wish to use to define our priors, in this example, pictures, for which neighboring pixels tend to have the same intensity value.

Of course, the property of locally constant intensities is not sufficient, and we have to introduce additional properties of other features such as edges. For this, we define a global energy of the form

$$J_G(x) = J_1(z) + J_2(t) + J_3(z, t) \tag{3.10}$$

composed of three terms: an intensity energy $J_1(z)$, an edge energy $J_2(t)$, and a third term $J_3(z, t)$ that describes the interaction between edges and pixels [81]–[83], [86]. These local energy functions are also in the form $\Sigma_i V_i$. They have close connections to statistical physics and Gibbs distributions. This point will be discussed in the next section. The design of these potentials is generally arduous and still empirical.

Having chosen the appropriate distances J_1 and J_2, the regularized solution $\hat{x}(\mu, y)$ is obtained for a given value of μ by minimizing the corresponding criterion. However, this brings up some practical questions about the complexity of the calculations, the complication which arises from the positive character of the solution, and the possible existence of local minima of the criterion.

The literature demonstrates that there is no satisfactory answer to these three questions simultaneously. As a rule, the calculations are easiest when J_1 and J_2 are quadratic:

$$\hat{x} = \arg\min \left\{ \| x - \hat{x}_0 \|_W^2 + \mu \| x \|_K^2 \right\}. \tag{3.11}$$

This corresponds to the classical regularization methods of Phillips, Twomey, and Tykhonov [13], [15]–[16], who have dominated the image restoration literature for the past 15 years [24]–[29], [44]–[63]. The solution to (3.11) is unique, linear with respect to the data, and can be calculated explicitly:

$$\hat{x} = (A'\Sigma^{-1}A + \mu K)^{-1} A'\Sigma^{-1} y. \tag{3.12}$$

But apart from the computational problems, which will be considered in Section V, this solution is not guaranteed to be positive, a point which is unacceptable in many applications. To remedy this deficiency, we can either directly impose positivity on the solution by nonlinear programming or we can use another distance J_2, e.g., Kullback distance. As the minimization of $\{J_1(x, \hat{x}_0) + \mu J_2(x, \hat{x}_\infty)\}$ is equivalent to that of $\{J_2(x, \hat{x}_\infty) + (1/\mu) J_1(x, \hat{x}_0)\}$, this yields the methods of reconstruction or restoration which are called "maximum entropy" methods [64]–[79]. Unfortunately, it is no longer possible to obtain a closed-form solution, and minimization of the criterion must be carried out iteratively.

When we wish to improve the description of the image by using contours and textures as well as the pixel intensity values, the difficulties in obtaining the solution are considerably increased. The criterion no longer presents a unique minimum and the global minimum must be sought by rather unwieldy techniques [80]–[86].

In addition to these computational difficulties, other difficulties arise such as the choice of the value of the regularization coefficient μ and possibly of other parameters which characterize the distances J_1 and J_2. These are the *hyperparameters* of the problem. There is a large variety of methods for determining them, but before considering them in Section VI, it will be useful to study the Bayesian interpretation of regularization methods.

IV. Bayesian Interpretation of Regularization Problems

An important part of statistical inference methods is based on the use of *a priori* information on the parameters to be estimated, which is added to the information provided by the data. It is thus not surprising, taking into account the profound nature of the regularization principle described above, that there is a close relationship between regularization and Bayesian estimation [30]–[43]. In a statistical context, the *a priori* information on the object x is expressed in the form of an *a priori* probability distribution $p(x)$. Bayes' rule allows us to combine it with the information contained in the data to obtain the *a posteriori* distribution

$$p(x/y) = p(y/x)\, p(x)/p(y). \tag{4.1}$$

In this relationship, $p(\boldsymbol{y}/\boldsymbol{x})$ denotes the probability distribution of the data conditioned on the real solution $\boldsymbol{x}$. It is completely determined by the knowledge of models (2.1) or (2.2) and the noise distribution. The last term

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}/\boldsymbol{x})\, p(\boldsymbol{x}) \tag{4.2}$$

ensures the normalization of the *a posteriori* distribution.

In a strict Bayesian sense, (4.1) is the solution to the inversion since it summarizes all the information on $\boldsymbol{x}$. However, manipulation of probability density functions is cumbersome and very often intractable, and a decision must be made to attribute a value to each pixel. A popular choice is that of attributing to $\boldsymbol{x}$ the value which maximizes the *a posteriori* (MAP) distribution [30]–[43]

$$\hat{\boldsymbol{x}} = \arg\max p(\boldsymbol{x}/\boldsymbol{y}). \tag{4.3}$$

But this is only a special case since MAP estimation corresponds to minimizing an average cost with a zero–one loss function. Recently, other cost functionals have gained interest in the context of Markov random fields modeling and lead to maximization of marginal probabilities [80]–[86].

In the case of interest here, it is clear that regularizing according to the general scheme given in the previous section is equivalent to choosing the optimal image maximizing the *a posteriori* distribution under the condition that this distribution is expressed in the following way:

$$p(\boldsymbol{x}/\boldsymbol{y}) \propto \exp\left\{-\tfrac{1}{2}\left[J_1(\boldsymbol{x}, \hat{\boldsymbol{x}}_0) + \mu J_2(\boldsymbol{x}, \hat{\boldsymbol{x}}_\infty)\right]\right\}. \tag{4.4}$$

The probability distribution expressed in (4.4) is only one possible choice. Any strictly monotone function would be suitable. However, choice of an exponential function is adequate here since with a linear model (2.2) and with the usual hypotheses of normality and independence on the noise, the conditional distribution $p(\boldsymbol{y}/\boldsymbol{x})$ is truly

$$p(\boldsymbol{y}/\boldsymbol{x}) \propto \exp\left\{-\tfrac{1}{2} J_1(\boldsymbol{x}, \hat{\boldsymbol{x}}_0)\right\} \tag{4.5}$$

as long as $J_1 = J_{WQ}$ with $\boldsymbol{W} = \boldsymbol{A}'\Sigma^{-1}\boldsymbol{A}$ are chosen. To complete the analogy, the *a priori* distribution must also be written as follows:

$$p(\boldsymbol{x}) \propto \exp\left\{-\frac{\mu}{2} J_2(\boldsymbol{x}, \hat{\boldsymbol{x}}_\infty)\right\}. \tag{4.6}$$

Again, this is only one Bayesian interpretation of regularization methods. It is also often considered as a justification of regularization methods, although this is a controversial point which we shall not discuss here. Suffice it to say that, on the one hand, the Bayesian approach is, in fact, the framework in which the most recent restoration methods have been introduced, and on the other hand, that it allows an appreciable extension of the range of hyperparameter estimation methods.

In the case of regularization methods of the Phillips, Twomey, and Tykhonov type, the distance J_2 is a quadratic, definite positive form, and (4.6) shows that this reverts to modeling the real object as a homogeneous random Gaussian field with mean $\hat{\boldsymbol{x}}_\infty$ and with covariance matrix $\boldsymbol{W}^{-1}$. The matrix $\boldsymbol{W}^{-1}$ can be chosen in the form $(\boldsymbol{D}'\boldsymbol{D})^{-1}$ if the distance J_2 is a roughness measure of the type (3.7). More generally, it can be any positive definite matrix, but its structure must be specified beforehand. If this matrix is itself parameterized, these hyperparameters must be determined in the same manner as the regularization coefficient μ, which is analogous to the inverse of a signal-to-noise ratio.

In maximum entropy methods, the proximity measure J_2 is of the Kullback distance type. Therefore, the real object is modeled as a random field with *a priori* distribution

$$p(\boldsymbol{x}) \propto \exp\left\{\frac{\mu}{2} S(\boldsymbol{x}, \hat{\boldsymbol{x}}_\infty)\right\} \tag{4.7}$$

where S is the negative of the Kullback distance or the relative entropy of the image. This choice is often given as the only one which is statistically consistent with a positive object [71]–[74]. We should nonetheless note that in this fully Bayesian approach, the constant μ should be specified by a "hyperprior" density. But in practice, this dimensional constant is not set *a priori*. Its choice is mainly empirical and depends on the quality of the data. However, apart from these more or less philosophical problems, the main drawback of these methods arises from the proximity measure which is used. The entropy of the whole image as defined in (3.6) is the sum of the entropies of each pixel, which implicitly assumes an *a priori* independence between the pixels since their joint *a priori* distribution (4.7) is the product of their marginal distributions:

$$p(\boldsymbol{x}) = \prod_i p(x_{1i}). \tag{4.8}$$

This no doubt explains why they give very good results in problems where the objects have supports which are of very limited dimensions and almost pinpoint, and why they differ only slightly from quadratic methods (with or without a positivity constraint) in the other cases.

In the case of local energy functions, the *a priori* distribution can be expressed in the following form:

$$p(\boldsymbol{x}) \propto \exp\left\{-J_G(\boldsymbol{x})\right\} \tag{4.9}$$

which is a Gibbs distribution. A physical system with such an associated distribution is such that the lower energy configurations have higher expectation. However, it must be noted that because of the strong nonlinearity of the Gibbs potentials and because of the large number of possible configurations for an image (M^N, M being the possible number of states, levels of gray, for example), it is impossible to compute all the distributions which are defined by (4.9) [80]–[86]. The marginal distributions $p(x_i)$, for example, cannot be calculated. Only conditional distributions $p(x_i/x_j, j \in d_i)$ where d_i indicates a neighborhood of pixel i can be simply and easily com-

puted from the local description leading to the choice of local energies. This difficulty has a strong influence on the computation of the solution, as we shall see in the following section. The assignment of these energy functions $J_G(x)$ also defines a Markov random field (MRF) image model since an MRF has the property that the probability distribution of the configuration of the field can always be expressed in the form of a Gibbs distribution [82]. The MRF–Gibbs equivalence is exploited for computing the solution. In order to partly overcome the computational problems, the fact that the *a posteriori* distribution is again Gibbsian with approximately the same neighborhood system as that of the original object is exploited and a sampling method called "Gibbs sampler" is used. However, the computational complexity increases rapidly with the neighborhood dimension, which is a major drawback in restoration problems with large PSF.

V. Practical Problems in Computing the Solution

Practical problems in computation of the reconstructed or restored image play an important role in the final choice of a particular method because of the very high dimensions of the real images. These problems arise even when the two distances J_1 and J_2 are quadratic and when a closed-form solution can be obtained since the use of the relation (3.12) requires the inversion of a very large matrix.

When the object has a limited support and when the corresponding hypotheses of uniformity of the image edges and circulant structure of the covariance or weighting matrix can be made, the matrix to be inverted is circulant in the case of a space-invariant degradation. It can then be rapidly diagonalized by FFT, which considerably simplifies its inversion [26]–[27].

Inversion can also be performed recursively using Kalman filtering techniques inspired by optimal statistical signal processing. A significant part of the literature on image restoration in recent years has been devoted to the development of these techniques [44]–[63]. Computational savings are generally obtained by introducing a state vector of reduced dimension compared to that of the real image. Although Kalman filtering techniques have been successful in 1-D problems, their extension to 2-D is not trivial. Derivation of suitable 2-D recursive models, compatible with a reasonable computational load, is a major difficulty. Image modeling with stationary 2-D random fields is an active research topic. But the study of these fields is not a simple extension of time series properties [8]. The lack of any Archimedean ordering in the plane raises new problems: existence of favored directions in lexicographically ordered ARMA processes, instability of finite memory inverse prediction filters, and impossibility to extend the 1-D stochastic realization theory to 2-D problems. The only state-space models for which a complete realization theory exists are those of Attasi [45] and Roesser [54]. Usually, additional assumptions, such as separability of covariance, are made in order to facilitate the identification of the image model. It should be noted that these ad hoc assumptions hardly reflect the true nature of the object.

When the distance J_2 is an entropy measure, things become more complicated since we no longer have a closed-form solution and we must proceed by iteration. The literature on this subject most often deals with iterative methods of the first order, of the conjugate-gradient type, but they must be modified in order to ensure the positivity of the solution at each iteration [64]–[79]. Experience shows that the computational cost is 10–20 times greater than with former linear methods using quadratic distances.

When Markovian representations associated with Gibbs potentials are used, the main difficulty in computing the solution arises from the impossibility of calculating the *a posteriori* distribution. This difficulty was resolved by using stochastic techniques of the Monte Carlo type. The first idea consists of introducing an auxiliary distribution

$$p_T(x/y) \propto \left[p(y/x)\, p(x)\right]^{1/T} \tag{5.1}$$

where $T > 0$ is an additional parameter called "temperature of the system." Note that if $T \to \infty$, (5.1) tends to a uniform distribution, while if $T \to 0$, the distribution (5.1) is concentrated on the maximum *a posteriori* estimate. The second idea consists of creating, in a stochastic way, a set of images by making random runs following a Gibbs distribution $p(x)$. Starting from an initial image $x^{(0)}$, the change from $x^{(n)}$ to $x^{(n+1)}$ only concerns one pixel, at the most, and is done randomly according to the local conditional distributions $p(x_i/x_j, j \in d_i)$ which can easily be calculated from $p(x)$. When $n \to \infty$, $x^{(n)}$ resembles a random configuration with distribution $p(x)$ whose empiric mean is the requested expectation. This is the "Gibbs sampler" [82]. When this technique is used with a distribution like the one defined in (5.1), and when the temperature T is brought very slowly to zero, convergence takes place towards the minimum *a posteriori* energy. This is "simulated annealing" [85]–[87]. The conditions to be fulfilled by the random set of images $x^{(0)}$, $x^{(1)}, \cdots$, in order to achieve convergence are very flexible and do not depend on how the object pixels are explored. The only requirement is that all pixels be explored. This flexibility and the particular form of the probability distribution p_T allow parallelization of the processing by using independent, and even asynchronous, processors, and thus should compensate for the impressive computational complexity of the method. Note that, in practice, the solution may not be unique and depends on the "cooling" schedule.

VI. Practical Problems in Computing Hyperparameters

All methods described above require the knowledge of the value of the regularization parameter μ, and more generally, of all the hyperparameters defining the distance measures J_1 and J_2: noise variances, correlation parame-

ters of the object, and parameters of local energy functions. Let θ denote the set of all hyperparameters. Estimation of θ is the most delicate part of image reconstruction and restoration methods, and it must be admitted that this problem is not properly solved yet.

It is obvious that vector θ, as the object $\boldsymbol{x}$, should be calculated from the observed data $\boldsymbol{y}$. When θ contains only parameter μ, the most intuitive and oldest idea [13], [16], [26], [28] is to consider μ as a Lagrange multiplier in the equivalent problem

$$\hat{\boldsymbol{x}}(\mu, \boldsymbol{y}) = \arg \min_{x \in X} J_2(\boldsymbol{x}, \hat{\boldsymbol{x}}_\infty), \quad \text{subject to } J_1(\boldsymbol{x}, \hat{\boldsymbol{x}}_0) \leq c. \tag{6.1}$$

The degree of regularization is set by the value of c which can be considered as a statistics whose probability distribution can be deduced from $p(\boldsymbol{y}/\boldsymbol{x})$. Since c now follows a known distribution, its value can be easily chosen, and a common choice is that of its expected value. For example, in the frequent case when $J_1(\boldsymbol{x}, \hat{\boldsymbol{x}}_0)$ is a quadratic form, c follows a χ^2 distribution with N degrees of freedom and $c = N$ is often chosen.

Such a choice has been reported to lead to overregularization of the solution. The residuals are given by $[\, y_i - (\boldsymbol{A}\hat{\boldsymbol{x}})_i \,]/\sigma_i$, $i = 1, 2, \cdots, N$. With $\hat{\boldsymbol{x}}$ equal to the true $\boldsymbol{x}$, these residuals would be a sample of N standard normal variables. But as $\hat{\boldsymbol{x}}(\mu, \boldsymbol{y})$ is inevitably biased, the residual errors used in the computation of $J_1(\boldsymbol{x}, \hat{\boldsymbol{x}}_0)$ do not exactly follow any known distribution.

To overcome this difficulty, we can consider that distances such as J_1 and J_2 are, in fact, measures of loss of the form $J\{\hat{\boldsymbol{x}}(\mu, \boldsymbol{y}), \boldsymbol{x}\}$. The expectation conditioned on the data

$$E\{J\} = \int J\{\hat{\boldsymbol{x}}(\mu, \boldsymbol{y}), \boldsymbol{x}\}\, p(\boldsymbol{y}/\boldsymbol{x})\, d\boldsymbol{y} \tag{6.2}$$

defines an average risk which depends on μ, and the choice of the value of μ which minimizes this criterion is surely acceptable. Unfortunately, this average risk also depends on the actual object which is unknown! As this risk is not calculable, we can try to estimate it, and this is done by cross-validation methods in the case of quadratic J distances [91], [96]–[97]. The basic principle is very simple, and consists of removing a datum y_i and of predicting it with the help of a regularized solution computed from the remaining data. The value of μ which ensures the best average prediction over all the removed data is retained. Although this criterion presents good asymptotic properties and is currently used in one-dimensional problems, it seems to have been seldom used in image processing [90].

The two methods presented above are deterministic in nature, but it is obvious that if an *a priori* distribution could be attributed to $\boldsymbol{x}$, a criterion depending on μ alone could be obtained by averaging (6.2), which corresponds to a Bayes' risk. The main interest of a Bayesian interpretation of regularization probably lies at this level. In fact, we have seen in the previous section that, save in the very special case of an *a priori* Gaussian distribution, it is impossible to calculate explicitly the maximum *a posteriori* solution and that the solution has to be computed iteratively. The main benefit to be drawn from a Bayesian interpretation is essentially that of methodology.

Another approach could be to estimate the hyperparameters and the object in the same way, by maximizing a joint *a posteriori* distribution. Since some of these quantities are random while others are deterministic, we may define a generalized likelihood function

$$p(\boldsymbol{x}, \boldsymbol{y}/\theta) = p(\boldsymbol{y}/\boldsymbol{x}, \theta)\, p(\boldsymbol{x}/\theta) \tag{6.3}$$

and maximize it over X and Θ:

$$(\hat{\boldsymbol{x}}, \hat{\theta}) = \arg \max_{x \in X, \theta \in \Theta} p(\boldsymbol{x}, \boldsymbol{y}/\theta). \tag{6.4}$$

Other functionals can be chosen (e.g., $p(\boldsymbol{x}/\boldsymbol{y}, \theta)$ which is the *a posteriori* distribution of the object), but their evaluation generally involves computation of a normalization factor $p(\boldsymbol{y})$ which is also a function of θ. This makes the estimation of θ impracticable. On the other hand, $p(\boldsymbol{x}, \boldsymbol{y}/\theta)$ is easier to compute. No convergence result can be established for $\boldsymbol{x}$ since the dimensions of the object and image grow accordingly. However, even though no such limitation applies to θ, the properties of this estimator have not been established yet. This is due to the complexity of the mathematical derivations involved in the general case. Some results [92] indicate that the MAP estimator does not always exhibit the desirable property of convergence, which makes its use questionable. Alternative approaches have to be sought.

The most commonly employed method consists of maximizing a marginal likelihood which is obtained by integrating the object out of the problem:

$$p(\boldsymbol{y}/\theta) = \int p(\boldsymbol{y}/\boldsymbol{x}, \theta)\, p(\boldsymbol{x}/\theta)\, d\boldsymbol{x}. \tag{6.5}$$

This can be considered as a special case of the EM algorithm which is a broadly applicable method for computing maximum likelihood estimates when the observations $\boldsymbol{y}$ are viewed as incomplete data [107]. The term "incomplete data" implies the existence of two spaces X and Y and a mapping from X to Y. The observed data $\boldsymbol{y}$ are a realization from Y and the corresponding $\boldsymbol{x}$ in X is not observed directly, but only indirectly through $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{x})$. The corresponding distributions $p(\boldsymbol{x}/\theta)$ and $p(\boldsymbol{y}/\theta)$ depend on parameters θ that have to be estimated. The *EM* algorithm is an iterative process which consists of an expectation (E) step followed by a maximization (M) step at each iteration. Assume that $\theta^{(n)}$ denotes the current value of θ after n iterations of the algorithm. The next iteration can be described in two steps, as follows.

E Step: Compute $E\{\log p(\boldsymbol{x}/\theta^{(n)})/\boldsymbol{y}, \theta\} = Q(\theta^{(n)}/\theta)$.

M Step: Choose $\theta^{(n+1)} = \arg \max_{\theta \in \Theta} Q(\theta/\theta^{(p)})$.

Equation (6.5) is the E step: conditional expectation given $\boldsymbol{y}$ and $\theta = \theta^{(n)}$. For the M step, we maximize this expectation or marginal log likelihood over Θ. The calculation of the likelihood itself is simplified when the restoration is made by a Kalman filter which performs a decomposition of the observed data into uncorrelated

variables, the innovations. Nevertheless, the maximization of (6.5) with respect to θ cannot be made explicitly and must be sought iteratively.

In the case of Markovian models, the problem becomes more complicated since, as we saw before, it is inconceivable to calculate the exact likelihood. We then use a pseudolikelihood drawn from local conditional distributions which are computable [80]–[83]:

$$q(y/\theta) = p(y/x, \theta)\, q(x/\theta) \tag{6.6}$$

where $q(x/\theta)$ is a product of local conditional distributions:

$$q(x/\theta) = \prod_i p(x_i/x_j, \theta; j \in v_i). \tag{6.7}$$

The maximization of the pseudolikelihood is performed by using an *EM* iterative algorithm, for example [81],

$$\theta^{(n+1)} = \arg\max_{\theta\in\Theta} E\left\{\log q(y/\theta^{(n)})\right\} \tag{6.8}$$

and calculation of the expectation in this expression is carried out using a Gibbs sampler.

Finally, we could also, in a fully Bayesian framework, attribute an *a priori* distribution to the hyperparameters θ and estimate them as the image x itself using an MAP technique. This solution was suggested recently for maximum entropy methods, but it is still too soon to estimate its usefulness in these image restoration and reconstruction problems.

In general, regardless of the particular expression of J_1 and J_2, and whatever interpretation is adopted, the important practical problem of the choice of regularization parameter μ, and more generally of the hyperparameters θ, remains open.

VII. Conclusions

In the usual vision model, image restoration and reconstruction are low-level information processing tasks. Under the pressure of recent spectacular technological developments that allow more and more complex data processing to be done, their traditional mathematical and statistical bases are being widened to rise toward higher information processing levels.

In this rapid and inevitably superficial survey, we have tried to show that, in spite of their great diversity, image restoration and reconstruction methods have a common estimation structure which is not basically undermined by the recent appearance of methods based on Markov modeling. These methods are interesting, as they allow an improvement in the description of an image, but they also present an impressive computational complexity. It thus becomes more necessary than ever to have at one's disposal results of comparative studies of the performances of these different methods by using different models and different types of image.

Computational and methodological complexity has to be balanced against the quality of results. Of course, this may depend on the later use of the processed image, but many papers report different restorations of the same standard images with, indeed, few visual differences. It will be invaluable to carry out meaningful comparative studies of the performances of a variety of reconstruction and restoration methods, using different regularization techniques on different types of images, to provide the user with objective elements of choice in a given practical situation.

Acknowledgment

The author wishes to express his gratitude to Dr. Y. Goussard for his expert editing assistance.

References

A. Textbooks on Image Reconstruction and Restoration

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

[2] B. R. Frieden, "Image enhancement and restoration," in *Picture Processing and Digital Processing*, T. S. Huang, Ed. Berlin: Springer-Verlag, 1975, pp. 177–248.

[3] G. T. Herman, H. K. Tuy, H. Langenberg, and P. C. Sabatier, *Basic Methods of Tomography and Inverse Problems*. Adams Hilger, 1987.

[4] J. H. Justice, *Maximum-Entropy and Bayesian Methods in Applied Statistics*. Cambridge, England: Cambridge University Press, 1986.

[5] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1988.

[6] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982.

[7] C. R. Smith and W. T. Grandy, Jr., *Maximum-Entropy and Bayesian Methods in Inverse Problems*. Dordrecht, The Netherlands: Reidel, 1985.

[8] A. S. Willsky, *Digital Signal Processing and Control and Estimation Theory*. Cambridge, MA: M.I.T. Press, 1979.

B. Basic Regularization Techniques

[9] M. Bertero, C. de Mol, and G. A. Viano, "The stability of inverse problems," in *Inverse Scattering Problems in Optics*, H. P. Baltes, Ed. Berlin: Springer-Verlag, 1980.

[10] J. W. Hilgers, "On the equivalence of regularization and certain reproducing kernel Hilbert space approaches for solving first kind problems," *SIAM J. Numer. Anal.*, vol. 13, pp. 172–184, 1976. ("Erratum," vol. 15, p. 1301, 1978.)

[11] M. Z. Nashed and G. Wahba, "Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operators equations," *SIAM J. Math. Anal.*, vol. 5, pp. 974–987, 1974.

[12] 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, 1981.

[13] 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.

[14] 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, 1983.

[15] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.

[16] S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," *J. Ass. Comput. Mach.*, vol. 10, pp. 97–101, 1962.

[17] G. Wahba, "Ill-posed problems: Numerical and statistical methods for mildly, moderately and severely ill-posed problems with noisy data," in *Proc. Int. Conf. Ill-Posed Problems*, Newark, NJ, Oct. 1979.

[18] ——, "Constrained regularization for ill-posed linear operator equations, with applications in meteorology and medicine," in *Statistical Decision Theory and Related Topics III*, S. S. Gupta and J. O. Berger, Ed. New York: Academic, 1982, pp. 383–417.

C. Choice of a Regularization Prescription

[19] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," M.I.T., Cambridge, Art. Intell. Lab. Memo 924, 1986.

[20] J. Cullum, "The effective choice of the smoothing norm in regularization," *Math. Comput.*, vol. 33, pp. 149-170, 1979.
[21] P. Hall and D. M. Titterington, "Common structure of techniques for choosing smoothing parameters in regression problems," *J. Roy. Statist. Soc. B*, vol. 49, pp. 184-198, 1987.
[22] D. M. Titterington, "General structure of regularization procedures in image restoration," *Astron. Astrophys.*, vol. 144, pp. 381-387, 1985.
[23] ——, "Common structure of smoothing techniques in statistics," *Int. Statist. Rev.*, vol. 53, pp. 141-170, 1985.

D. Standard Regularization Procedures

[24] M. Bertero, C. de Mol, and G. A. Viano, "On the problems of object restoration and image extrapolation in optics," *J. Math. Phys.*, vol. 20, pp. 509-521, 1979.
[25] C. W. Helstrom, "Image restoration by the method of least squares," *J. Opt. Soc. Amer.*, vol. 57, pp. 297-303, 1967.
[26] B. R. Hunt, "The inverse problem of radiography," *Math. Biosci.*, vol. 8, pp. 161-179, 1970.
[27] ——, "Biased estimation for nonparametric identification of linear systems," *Math. Biosci.*, vol. 10, pp. 215-237, 1971.
[28] ——, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Trans. Comput.*, vol. C-22, pp. 805-812, 1973.
[29] A. K. Katsaggelos, J. Biemond, R. M. Mersereau, and R. W. Schafer, "A general formulation of constrained iterative restoration algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1985, pp. 700-703.

E. Bayesian Approach

[30] J. F. Abramatic and L. M. Silverman, "Non-stationary linear restoration of noisy images," in *Proc. IEEE Conf. Decision Contr.*, Fort Lauderdale, FL, Dec. 1979, pp. 92-99.
[31] B. P. Agrawal and V. K. Jain, "Digital restoration of images in the presence of correlated noise," *Comput. Elec. Eng.*, vol. 3, pp. 65-74, 1976.
[32] M. H. Buonocore, W. R. Brody, and A. Makovsky, "Fast minimum variance estimator for limited angle CT image reconstruction," *Med. Phys.*, vol. 8, pp. 695-702, 1981.
[33] A. M. Darling, T. J. Hall, and M. A. Fiddy, "Stable noniterative object reconstruction from incomplete data using a priori knowledge," *J. Opt. Soc. Amer.*, vol. 73, pp. 1466-1469, 1983.
[34] J. N. Franklin, "Well-posed stochastic extensions of ill-posed linear problems," *J. Math. Anal. Appl.*, vol. 31, pp. 682-716, 1970.
[35] B. R. Frieden, "Statistical estimates of bounded optical scenes by the method of prior probabilites," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 118-119, 1973.
[36] K. M. Hanson, "Bayesian and related methods in image reconstruction from incomplete data," in *Image Recovery: Theory and Application*, H. Stark, Ed. New York: Academic, 1987, pp. 79-125.
[37] G. T. Herman and A. Lent, "A computer implementation of a Bayesian analysis of image reconstruction," *Inform. Contr.*, vol. 31, pp. 364-384, 1976.
[38] G. T. Herman, H. Hurwitz, A. Lent, and H. P. Lung, "On the Bayesian approach to image reconstruction," *Inform. Contr.*, vol. 42, pp. 60-71, 1979.
[39] B. R. Hunt and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 876-882, 1976.
[40] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Comput.*, vol. C-26, pp. 219-229, 1977.
[41] H. J. Trussel, "Improved methods of maximum *a posteriori* restoration," *IEEE Trans. Comput.*, vol. C-27, pp. 57-62, 1979.
[42] ——, "The relationship between image restoration by the maximum *a posteriori* method and a maximum entropy method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 114-117, 1980.
[43] S. L. Wood and M. Morf, "A fast implementation of a minimum variance estimator for computerized tomography image reconstruction," *IEEE Trans. Biomed. Eng.*, vol. BME-28, pp. 56-68, 1981.

F. Kalman Filtering Techniques

[44] A. O. Aboutalib, M. S. Murphy, and L. M. Silverman, "Digital restoration of images degraded by general motion blurs," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 294-302, 1977.
[45] S. Attasi, "Modelling and recursive estimation for double indexed sequences," in *System Identification: Advances and Studies*, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976, pp. 289-348.
[46] M. R. Azimi-Sadjadi and P. W. Wong, "Two-dimensional block Kalman filtering for image restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1736-1749, 1987.
[47] J. Biemond, J. Rieske, and J. J. Gerbrands, "A fast Kalman filter for images degraded by both blur and noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1248-1256, 1983.
[48] S. S. Dikshit, "A recursive Kalman window approach to image restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 125-140, 1982.
[49] D. X. Cheng, D. Saint-Félix, and G. Demoment, "Comparison between a factorization method and a partitioning method to derive invariant Kalman filters for fast image restoration," in *Mathematics in Signal Processing*, T. S. Durrani *et al.*, Eds. Oxford, England: Clarendon, 1987, pp. 349-362.
[50] A. Habibi, "Two-dimensional Bayesian estimate of images," *Proc. IEEE*, vol. 60, pp. 878-883, 1972.
[51] M. S. Murphy and L. M. Silverman, "Image model representation and line-by-line recursive restoration," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 809-816, 1978.
[52] N. E. Nahi, "Role of recursive estimation in statistical image enhancement," *Proc. IEEE*, vol. 60, pp. 872-877, 1972.
[53] S. A. Rajala and R. J. P. de Figueiredo, "Adaptive nonlinear image restoration by a modified Kalman filtering approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 1033-1042, 1981.
[54] R. Roesser, "A discrete state-space model for linear image processing," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 1-10, 1975.
[55] D. Rohler and P. S. Krishnaprasad, "Radon inversion and Kalman reconstructions: A comparison," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 483-487, 1981.
[56] D. Saint-Félix, D. X. Cheng, and G. Demoment, "Image restoration using a non-causal state space model and a fast 2D Kalman filter," in *Mathematics in Signal Processing*, T. S. Durrani *et al.*, Eds. Oxford, England: Clarendon, 1987, pp. 363-378.
[57] B. R. Suresh and B. A. Shenoi, "New results in two-dimensional Kalman filtering with applications to image restoration," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 307-319, 1981.
[58] S. L. Wood, A. Makovski, and M. Morf, "Reconstructions with limited data using estimation theory," in *Computer Aided Tomography and Ultrasonics in Medicine*, J. Raviv *et al.*, Eds. Amsterdam: North-Holland, 1979, pp. 219-231.
[59] J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 473-482, 1977 (and "Correction," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 628-629, 1979).
[60] J. W. Woods and V. K. Ingle, "Kalman filtering in two dimensions: Further results," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 188-197, 1981.
[61] J. W. Woods, J. Biemond, and A. M. Tekalp, "Boundary value problem in image restoration,"in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1985, pp. 18.11.1-18.11.4.
[62] J. W. Woods, S. Dravida, and R. Mediavilla, "Image estimation using doubly stochastic Gaussian random field models,"*IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 245-253, 1987.
[63] W. Zhe, "Multidimensional state-space model Kalman filtering with application to image restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1576-1592, 1986.

G. Maximum Entropy

[64] R. K. Bryan and J. Skilling, "Deconvolution by maximum entropy, as illustrated by application to the jet of M87," *Mon. Not. R. Astr. Soc.*, vol. 191, pp. 69-79, 1980.
[65] S. F. Burch, S. F. Gull, and J. Skilling, "Image restoration by a powerful maximum entropy method," *Comput. Vision, Graphics, Image Processing*, vol. 23, pp. 113-128, 1983.
[66] G. J. Daniell and S. F. Gull, "Maximum entropy algorithm applied to image enhancement," *Proc. IEE*, vol. 127-E, pp. 170-172, 1980.
[67] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.*, vol. 62, pp. 511-518, 1972.
[68] ——, "Restoring with maximum entropy, II: Superresolution of photographs of diffraction-blurred impulses," *J. Opt. Soc. Amer.*, vol. 62, pp. 1202-1210, 1972.

[69] ——, "Restoring with maximum entropy. III. Poisson sources and backgrounds," *J. Opt. Soc. Amer.*, vol. 68, pp. 93-103, 1978.
[70] ——, "Statistical models for the image restoration problem," *Comput. Graphics Image Processing*, vol. 12, pp. 40-59, 1980.
[71] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686-690, 1978.
[72] S. F. Gull and J. Skilling, "Maximum entropy method in image processing," *Proc. IEE*, vol. 131-F, pp. 646-659, 1984.
[73] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, pp. 227-241, 1968.
[74] ——, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, pp. 939-952, 1982.
[75] R. Kikuchi and B. H. Soffer, "Maximum entropy image restoration. I. The entropy expression," *J. Opt. Soc. Amer.*, vol. 67, pp. 1656-1665, 1977.
[76] A. Mohammad-Djafari and G. Demoment, "Maximum-entropy diffraction tomography," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, Apr. 1986, pp. 1749-1752.
[77] ——, "Maximum-entropy Fourier synthesis with application to diffraction tomography," *Appl. Opt.*, vol. 26, pp. 1745-1754, 1987.
[78] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26-37, 1980. ("Comments and correction," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 942-943, 1983.)
[79] S. Sibisi, "Two-dimensional reconstructions from one-dimensional data by maximum entropy," *Nature*, vol. 301, pp. 134-136, 1983.

H. Markov Random Fields and Related Estimation Techniques

[80] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, pp. 259-302, 1986.
[81] B. Chalmond, "Image restoration using an estimated Markov model," *Signal Processing*, vol. 15, pp. 115-129, 1988.
[82] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesien restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721-741, 1984.
[83] S. Geman, "Stochastic relaxation methods for image restoration and expert systems," in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, G. J. Erikson and C. R. Smith, Eds. Norwell, MA: Kluwer Academic, 1988, pp. 265-311.
[84] H. Jinchi, T. Simchony, and R. Chellapa, "Stochastic relaxation for MAP restoration of gray level images with multiplicative noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Dallas, TX, Apr. 1987, pp. 1236-1239.
[85] S. Kirkpatrick, C. D. Gellatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.
[86] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Ass.*, vol. 82, pp. 76-89, 1987.
[87] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087-1091, 1953.
[88] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 232-240, 1972.
[89] ——, "Markov image modeling," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 846-850, 1978.

I. Hyperparameter Estimation

[90] D. Girard, "Les méthodes de régularisation optimale et leurs applications en tomographie," thèse de Docteur-Ingénieur, Institut National Polytechnique de Grenoble, France, 1984.
[91] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215-223, 1979.
[92] Y. Goussard, G. Demoment, and F. Monfront, "Maximum *a posteriori* detection-estimation of Bernoulli-Gaussian processes," in *Proc. 2nd IMA Conf. Math. Signal Processing*, Warwick, England, Dec. 1988, 19 pp.
[93] J. W. Hilgers, "A note on estimating the optimal regularization parameter," *SIAM J. Numer. Anal.*, vol. 17, pp. 472-473, 1980.
[94] H. Kaufman, J. W. Woods, S. Dravida, and A. M. Tekalp, "Estimation and identification of two-dimensional images," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 745-756, 1983.
[95] F. van der Putten, J. Biemond, and J. W. Woods, "A hybrid identification scheme for image and blur parameters," in *Signal Processing III*, I. T. Young *et al.*, Eds. New York: Elsevier, 1986, pp. 765-768.
[96] F. O'Sullivan, B. S. Yandell, and W. J. Raynor, Jr., "Automatic smoothing of regression functions in generalized linear models," *J. Amer. Statist. Ass.*, vol. 81, pp. 96-103, 1986.
[97] B. W. Silverman, "A fast and efficient cross-validation method for smoothing parameter choice in spline regression," *J. Amer. Statist. Ass.*, vol. 79, pp. 584-589, 1984.
[98] A. M. Tekalp, H. Kaufman, and J. W. Woods, "Identification of image and blur parameters for the restoration of noncausal blurs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 963-972, 1986.

J. Miscellaneous

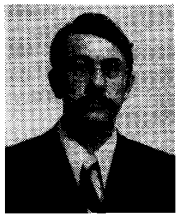
[99] J. F. Abramatic and L. M. Silverman, "Nonlinear restoration of noisy images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-4, pp. 141-149, 1982.
[100] G. L. Anderson and A. N. Netravali, "Image restoration based on a subjective criterion," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 845-853, 1976.
[101] C. L. Byrne, R. M. Fitzgerald, M. A. Fiddy, T. J. Hall, and A. M. Darling, "Image restoration and resolution enhancement," *J. Opt. Soc. Amer.*, vol. 73, pp. 1481-1487, 1983.
[102] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 58-63, 1976.
[103] M. Cannon, H. J. Trussell, and B. R. Hunt, "Comparison of image restoration methods," *Appl. Opt.*, vol. 17, pp. 3384-3390, 1978.
[104] P. Chan and J. S. Lim, "One-dimensional processing for adaptive image restoration," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 117-126, 1985.
[105] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 461-472, 1982.
[106] M. E. Davison, "The ill-conditioned nature of the limited angle tomography problem," *SIAM J. Appl. Math.*, vol. 43, pp. 428-448, 1983.
[107] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.
[108] M. P. Ekstrom and R. L. Roads, "On the application of eigenvector expansions to numerical deconvolution," *J. Comput. Phys.*, vol. 14, pp. 319-340, 1974.
[109] G. T. Herman, A. Lent, and H. Hurwitz, "A storage-efficient algorithm for finding the regularized solution of a large, inconsistent system of equations," *J. Inst. Math. Appl.*, vol. 25, pp. 361-366, 1980.
[110] A. K. Jain and E. Angel, "Image restoration, modeling, and reduction of dimensionality," *IEEE Trans. Comput.*, vol. C-23, pp. 470-476, 1974.
[111] A. K. Jain, "A semicausal model for recursive filtering of two-dimensional images," *IEEE Trans. Comput.*, vol. C-26, pp. 343-350, 1977.
[112] A. K. Jain and J. R. Jain, "Partial differential equations and finite difference methods for image processing. Part II: Image restoration," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 817-834, 1978.
[113] H. E. Knutsson, R. Wilson, and G. H. Granlund, "Anisotropic nonstationary image estimation and its application. Part I: Restoration of noisy images," *IEEE Trans. Commun.*, vol. COM-31, pp. 388-397, 1983.
[114] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 165-177, 1985.
[115] R. M. Leahy and C. E. Goutis, "An optimal technique for constraint-based image restoration and reconstruction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1629-1642, 1986.
[116] N. D. A. Mascarenhas and W. K. Pratt, "Digital image restoration under a regression model," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 252-266, 1975.
[117] M. I. Sezan and H. Stark, "Image restoration by the method of convex projections," *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 95-101, 1982.
[118] ——, "Tomographic image reconstruction from incomplete view data

by convex projections and direct Fourier inversion," *IEEE Trans. Med. Imaging*, vol. MI-3, pp. 91–98, 1984.

[119] A. M. Tekalp and H. J. Trussell, "Comparative study of some recent statistical and set-theoretic methods for image restoration," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, NY, Apr. 1988, pp. 988–991.

[120] T. J. Uhl, "Linear and nonlinear image restoration methods in comparison," in *Signal Processing III*, I. T. Young, Ed. New York: Elsevier Science, 1986, pp. 739–742.

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



Guy Demoment was born in France in 1948. He graduated from the École Supérieure d'Électricité in 1970, and received the "Doctorat d'État" degree in physics from the Université de Paris-Sud, Orsay, in 1977.

From 1971 to 1977 he was with the Service des Mesures de l'École Supérieure d'Électricité. From 1977 to 1988 he was with the Centre National de la Recherche Scientifique, assigned to the Laboratoire des Signaux et Systèmes. He is presently a Professor in the Department of Physics, Université de Paris-Sud, Orsay. After some work on biological systems modeling, his interests moved toward inverse problems in signal and image processing.

Dr. Demoment is a member of TCO and SRV.