[Alejandro Ribés and Francis Schmitt]

LINEAR INVERSE PROBLEMS IN IMAGING

[An introductory survey]

I nverse problems are an old topic in mathematics and are currently so widely found in engineering that it is difficult to make a full list of their applications. In particular, inverse problems appear very often in image processing. There is a fundamental reason for this: electronic sensors acquire properties of the "reality" that are not the same as our natural sensors, the eyes. This fundamental reason even possesses a philosophical side, as Inmanuel Kant noted: "Things that we see are not by themselves what we see" [1]. This assertion is indeed the source of most inverse problems that appear in image processing. Physics give some insights to this point. A sensor involved in an imaging process can normally be modeled by an equation containing an integral of a function or the composition of several functions. For instance, an optical charge coupled devise (CCD) integrates light on a finite interval of the electromagnetic spectrum and gives as a result an electrical signal; this is further converted into a digital signal, the so-called gray level. The integral nature of the basic image acquisition floods the field of image processing with integral processes. In short, we could say the following: We do not see the objects by themselves

© ARTVILLE

because we see their integrals. This is the reason why inverse problems appear so often in image processing. When information about the "objects by themselves" is to be extracted from an image, the integral process should be inversed. This inversion, when working with a computer, is intimately related to linear algebra and matrix analysis.

The inversion of an integral or resolution of an integral equation can be difficult to solve; properly speaking, it can be an ill-posed problem. Let us try to explain why. An integral can be understood as an addition of elements in a continuous space. Thus, an integral is a process that loses information about the functions involved in the infinite addition. Functions are, in fact, reduced to a single number or several numbers which represent a more complex reality. Let us introduce here a very basic example: if we ask someone "How much is $1 + 2$?" he should answer 3; but if we ask the inverse question "Give me some numbers that add up to 3," the answer is not straightforward since infinite solutions exist. The mathematical study of this kind of problem started more than a century ago, especially with the definition of a well-posed problem [2] and continues to be an active field of research.

The aim of this survey is to show how linear inverse problems in imaging can be solved using basic techniques. For this, we provide a classification of the solutions into several families and we illustrate each sort of solution by examples found in current state-of-the-art imaging systems. Our classification imposes some order on the complexity that a nonexpert can find in the abundant literature. Moreover, it emphasizes the practical resolution of the problems.

The concepts used in this survey are illustrated by actual examples of inverse linear problems in imaging. We choose some of them for their historical importance, such as image restoration, and others for their critical issue in state-of-the-art imaging systems. For instance, the resolution of the so-called spectral reflectance reconstruction allows the acquisition of high-fidelity color digital images. Among others, the famous painting by Leonardo da Vinci, the "Mona Lisa" (Museum of Louvre, Paris, France), was scanned as part of the conservation restoration innovation systems for image capture and digital archiving to enhance training, education, and lifelong learning (CRISATEL) European project [3] in order to perform spectral reconstruction. Another inverse problem that illustrates this survey is associated with parallel magnetic resonance imaging (MRI), a rapid acquisition technique for medical imaging which is present in current MRI medical equipment. With these chosen examples, we will try to show the broad scope of applications where linear inverse problems are found.
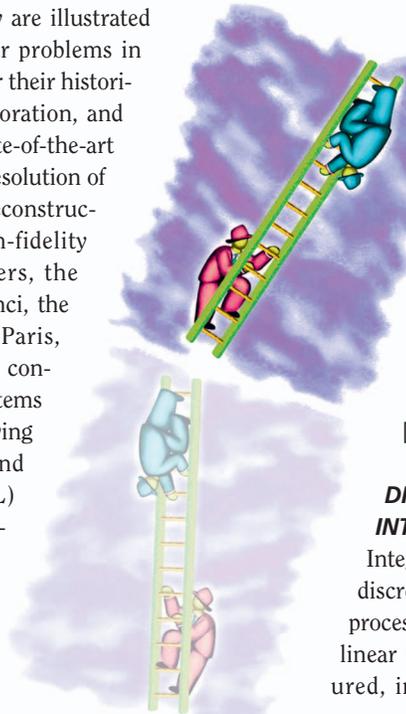
This survey is organized into two main parts. We first describe theoretical aspects of linear inverse problems and introduce basic concepts of linear inverse problems. We propose a new classification of methods for the resolution of linear inverse problems. This classification leads to the second part of the survey, which explores a set of examples illustrating the implementations of the various methods: image restoration, spectral reflectance reconstruction, and parallel MRI.
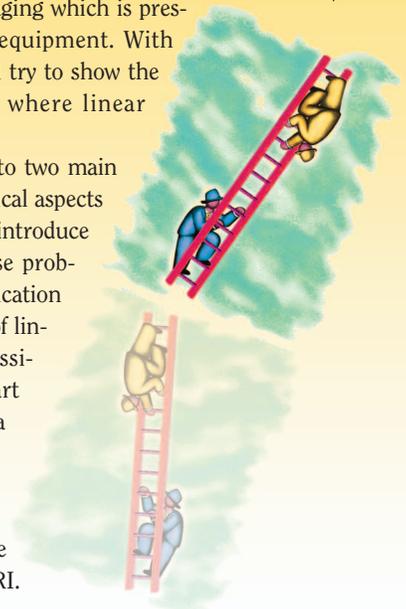
LINEAR INVERSE PROBLEMS: CONCEPTS

Most inverse problems in imaging result from integral processes involved in the acquisition of the image. This can be modeled by the following equation:

$$m(x) = \int_{\Lambda} \phi(x, \lambda) o(\lambda) d\lambda + n(x), \qquad (1)$$

where

- $m(x)$ is the information obtained from the imaging process, the measurements,
- $\phi(x, \lambda)$ is the kernel of the integral equation (this function has a different physical meaning depending on the application),
- $o(\lambda)$ is the property of the object we are indirectly measuring,
- $n(x)$ is the noise, unfortunately always present in actual inverse problems, and
- Λ is the space where λ is integrated.

The integral in (1) becomes an integral equation when we want to recover $o(\lambda)$, the object property. This sort of equation is a Fredholm integral equation of the first kind. Theoretically, they are known to be intimately related with ill-posed problems; see [4] for a mathematical treatment and [5] for some classic solutions.

DISCRETIZATION OF A FREDHOLM INTEGRAL EQUATION

Integral equations defined in continuous spaces must be discretized if we want to solve them numerically. In this process, the kernel $\phi(x, \lambda)$ in (1) becomes a matrix Θ in a linear system. In general, $\phi(x, \lambda)$ can be physically measured, indirectly estimated, mathematically modeled, or

simply not known. In any case, the resulting linear system takes the following form:

$$\mathbf{m} = \Theta \mathbf{o} + \mathbf{n}, \qquad (2)$$

where lowercase bold indicates a finite discrete vector. In (2), vector $\mathbf{o}$ is a discretization of the property of the object and $\mathbf{m}$ is a vector representing the function $m(x)$. Note that the discretization of $m(x)$ in the vector $\mathbf{m}$ is forced by the nature of the acquisition system; this will be clearly stated in the examples. Vector $\mathbf{n}$ represents the additive random noise.

ILL-POSED PROBLEM

The notion of a well-posed problem goes back to a famous paper by Jacques Hadamard published in 1902 [2]. A well-posed problem in the sense of Hadamard is a problem that fulfills the following three conditions:

1) The solution exists.
2) The solution is unique.
3) The solution depends continuously on the problem data.

If any of these conditions is not respected, the problem becomes ill-posed. Note that both the first and second conditions deal with the feasibility of the problem, and the last condition relates to the possible implementation of a stable numerical procedure for its resolution. The solution of a problem is always based on some data, typically obtained from experimentation. If the solution does not depend "smoothly" on the problem data, small variations on the data can create huge variations on the solutions, resulting in strong instability which is not acceptable.

REGULARIZATION

When solving ill-posed problems, the concept of regularization immediately appears. Regularization is used to well-pose a problem that is ill-posed. Once the problem is well-posed, we can solve it. Thus, all three conditions of Hadamard should be respected by the proposed solution. When working with linear algebra conditions, 1) and 2) are easily fulfilled by a pseudo-inverse [6]. However, it is in general more difficult to deal with the stability of the solution. Historically the so-called Tikhonov regularization is one of the oldest and most well-known techniques for stabilization, see [7] for broadest references. Also, the Wiener filter is a classical approach for ill-posed problems on signal and image processing [8].

RESOLUTION OF LINEAR INVERSE PROBLEMS: A CLASSIFICATION

Linear inverse problems have been solved using literally hundreds of methods and variations of these methods. When first confronted with a new inverse problem, the variety of solutions is indeed huge. This point is central for an engineer aiming to solve a new problem. Which is the best solution in a particular case? Why? This is not clearly stated when navigating in technical results of sometimes sophisticated mathematics. We decided to base this survey on a classification of the

methods. We have divided the solutions of linear inverse problems into four families: Fourier transform (FT) based, direct inversion, indirect inversion, and interpolation. They are presented in this section, and will be illustrated later using actual inverse problems in imaging.

FT SOLUTIONS

The FT was one of the most used mathematical tools in the twentieth century. Several introductions to image processing make extensive use of this concept, and a lot of methods based on the FT exist. In part, this is due to FT methods being very pedagogical and easy to understand. Moreover, its computer implementation, the fast Fourier transform (FFT), is efficient and quick. It is then not surprising that this technique has been intensively used in imaging. However, much care should be taken as the FT applied to inverse problems is always dependent on specific properties of the problem and it is then not general. We will further explain this dependency later with an example: the restoration of degraded images.

DIRECT RECONSTRUCTION PROBLEM

Direct reconstruction appears in the case where operator Θ in (2) is known. Then, the problem consists of finding vector $\mathbf{o}$ when $\mathbf{m}$ is given. This should be reached by inversing matrix Θ, in the absence of noise: $\mathbf{o} = \mathrm{inv}(\Theta)\,\mathbf{m}$. From this apparently simple linear system we remark that the matrix Θ is in general not a square matrix; consequently, the system itself is over- or underdetermined by definition. This means that either the system has no solution or it has many. Clearly, this does not respect conditions 1) or 2) of the Hadamard definition: the problem is ill-posed. The third condition is not as straightforward to see as the others, but modern numerical linear algebra presents enough resources for the analysis of the stability of a matrix. If the matrix is singular, its inverse will be unstable. The condition number, the rank of the matrix, or the Picard condition, among others, are good analytical tools to determine if we are dealing with an ill-posed problem; see [9] for a valuable reference on this subject.

CLASSICAL SOLUTIONS

Matrix Θ not being a square matrix, its inverse does not exist; but it can be replaced by the so-called pseudo-inverse, [6], [10]. The pseudo-inverse corresponds to the solution of the following minimization problem:

$$\min_{o} \|\mathbf{m} - \Theta \mathbf{o}\|^2. \qquad (3)$$

The solution is obtained from the normal equations by simply writing the derivatives of (3) equal to zero. At this point, care should be taken as the name pseudo-inverse is used for two different matrices depending on the number K of variables and on the number N of unknowns. When $K > N$, the problem is overdetermined and the pseudo-inverse is defined by

$$\Theta_{\mathrm{over}}^{-} = (\Theta^t \Theta)^{-1} \Theta^t. \qquad (4)$$

When $K < N$, the problem is underdetermined and the pseudo-inverse is defined by

$$\Theta^-_{\text{under}} = \Theta^t(\Theta\Theta^t)^{-1}. \qquad (5)$$

This dual definition of the pseudo-inverse can be a source of confusion in practical problems. Indeed, (4) and (5) seem to be equivalent but transposed, while corresponding to a different underlying problem. Further in this survey we emphasize this point by choosing actual examples which are under- and overdetermined. A pseudo-inverse respects Hadamard's conditions 1) and 2), but the stability of the solution is not guaranteed. Consequently, other techniques imposing stability were historically developed. In order to describe several classical techniques in the same mathematical framework, we introduce the following minimization problem:

$$\min_o J(\mathbf{o}) : J(\mathbf{o}) = (\mathbf{m} - \Theta\mathbf{o})^t \mathbf{N}(\mathbf{m} - \Theta\mathbf{o}) + \lambda^2 \|\mathbf{L}\mathbf{o}\|^2, \quad (6)$$

where $\mathbf{N}$ is a matrix that modifies the criterion $(\mathbf{m} - \Theta\mathbf{o})^t(\mathbf{m} - \Theta\mathbf{o})$ minimized in (3), $\lambda^2\|\mathbf{L}\mathbf{o}\|^2$ is an added regularization term reinforcing minimal $\mathbf{o}$ norm solutions, $\mathbf{L}$ is a positive definite matrix, and λ^2 a parameter that controls the influence of the regularization term. A closed form of the minimum of (6) can be found by simply differentiating $J(\mathbf{o})$ and setting its result to the zero vector:

$$\frac{\partial J(\mathbf{o})}{\partial \mathbf{o}} = -2\Theta^t \mathbf{N}(\mathbf{m} - \Theta\mathbf{o}) + 2\lambda^2 \mathbf{L}^t \mathbf{L}\mathbf{o} = 0. \qquad (7)$$

Solving (7) for $\mathbf{o}$ yields to the optimal least squares operator in the overdetermined case:

$$\Theta^-_{over} = (\Theta^t \mathbf{N}\Theta + \lambda^2 \mathbf{L}^t \mathbf{L})^{-1}\Theta^t \mathbf{N}. \qquad (8)$$

Its dual underdetermined operator is given by: $\Theta^-_{\text{under}} = \mathbf{N}\Theta^t(\Theta\mathbf{N}\Theta^t + \lambda^2 \mathbf{L}^t\mathbf{L})^{-1}$. In the rest of this section, three methods that are often used in solving linear inverse problems are presented. We choose to introduce them in a common mathematical framework; in fact, all are special cases of (8) presented above.

MODIFYING THE MINIMIZED NORM
Equation (6) is often simplified not using the regularization term $\lambda^2\|\mathbf{L}\mathbf{o}\|^2$. This leads to

$$\Theta^-_{\text{N-over}} = (\Theta^t \mathbf{N}\Theta)^{-1}\Theta^t \mathbf{N}, \qquad (9)$$

where $\mathbf{N}$ modifies the norm of a classical least squares problem; this approach is often called weighted least squares; see chapter 5 of [11]. Its dual underdetermined operator is given by $\Theta^-_{\text{N-under}} = \mathbf{N}\Theta^t(\Theta\mathbf{N}\Theta^t)^{-1}$. In fact, the case where $\mathbf{N}$ is the identity matrix minimizes the Euclidean norm; in this case (6) becomes (3). This matrix can take different forms depending on the application; for instance, it can be a smoothing operator. A

very popular choice is the inverse of a covariance matrix which is estimated from the noise affecting the measurements. Then, the operator defined in (9) becomes a special case of the Wiener filter.

WIENER FILTER
Most of the time, inverse operators are unstable due to the effect of noise. Knowledge of the model of noise or at least its covariance matrix is very useful as it can be integrated in the solution in order to stabilize it. When both covariance matrices over vectors $\mathbf{m}$ and $\mathbf{o}$ are known, the inversion operator takes the following form:

$$\Theta^-_{\text{Wiener-over}} = \left(\Theta^t \Psi_{\mathbf{m}}^{-1}\Theta + \Psi_{\mathbf{o}}^{-1}\right)^{-1}\Theta^t \Psi_{\mathbf{m}}^{-1}, \qquad (10)$$

where $\Psi_{\mathbf{o}}$ is the covariance matrix of the object characteristic we want to estimate and $\Psi_{\mathbf{m}}$ is the covariance matrix of the measurements. Its dual underdetermined operator is given by $\Theta^-_{\text{Wiener-under}} = \Psi_{\mathbf{o}}^{-1}\Theta^t(\Theta \Psi_{\mathbf{o}}^{-1}\Theta^t + \Psi_{\mathbf{m}}^{-1})^{-1}$. The relationship between (10) and (8) is clear when stating $\mathbf{N} = \Psi_{\mathbf{m}}^{-1}$ and $\lambda^2 \mathbf{L}^t \mathbf{L} = \Psi_{\mathbf{o}}^{-1}$. This solution is referred to as Wiener estimation; it requires the second-order statistics with respect to the signal and noise, in addition to the system matrix Θ.

TIKHONOV REGULARIZATION
Another classic technique to regularize this kind of problem is Tikhonov regularization. It consists of using the regularization term already presented in (6). In its classical form, matrix $\mathbf{L}$ is omitted and the norm is not modified by $\mathbf{N}$, leading to the following minimization problem:

$$\min_{\mathbf{o}} \|\mathbf{m} - \Theta\mathbf{o}\|^2 + \lambda^2 \|\mathbf{o}\|^2, \qquad (11)$$

and to the following solution:

$$\Theta^-_{\text{Tikhonov-over}} = (\Theta^t\Theta + \lambda^2 \mathbf{I})^{-1}\Theta^t, \qquad (12)$$

where $\mathbf{I}$ is the identity matrix. Its dual underdetermined operator is given by $\Theta^-_{\text{Tikhonov-under}} = \Theta^t(\Theta\Theta^t + \lambda^2 \mathbf{I})^{-1}$. This technique is widely used when treating the inversion of Θ. Care must be taken as it depends on the parameter λ; the choice of this parameter highly influences the estimated $\mathbf{o}$. If we look for a quick solution, this parameter can be manually adjusted. However, when it must be automatically chosen, λ is either calculated from specific characteristics of the problem, optimized by established techniques such as cross-validation [12], or optimized by the L-curve [9].

SINGULAR VALUE DECOMPOSITION
The computational inversion of a nonsquare and probably singular matrix is a central aspect when solving a linear inverse problem. To do so, a singular value decomposition (SVD) of the matrix is often used [13]:

$$\Theta = \mathbf{U}\Sigma\mathbf{V}^t = \sum_{i=1}^n \mathbf{u}_i \sigma_i \mathbf{v}_i^t, \qquad (13)$$

where $U = (u_1, u_{2,\ldots}, u_n)$ and $V = (v_1, v_{2,\ldots}, v_n)$ are complex matrices with orthonormal columns and $\Sigma = \text{diagonal}$ $(\sigma_1, \sigma_{2,\ldots}, \sigma_n)$ is a real matrix containing nonnegative diagonal elements called singular values which appear in nonincreasing order such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

A stable inverse Θ of can be calculated by

$$\Theta^-_{\text{f-SVD}} = V\Sigma_f^{-1}U^t = \sum_{i=1}^{n} v_i \frac{f_i}{\sigma_i} u_i^t,$$

where the filter factors $f_1, f_2, \ldots, f_n$ are used to control the singular values in the inversion. Depending on the choice of the filter factors, different regularization techniques can be implemented; for instance, choosing $f_i = \sigma_i^2/(\sigma_i^2 + \lambda^2)$ leads to Tikhonov regularization. The use of an SVD when analyzing and solving a linear inverse problem is common practice and it has been extensively studied [9]. However, in some cases the use of an SVD is not appropriate, i.e., when working with large matrices as the time required calculating the SVD becomes huge compared to inversion methods based on iterative algorithms. These iterative algorithms do not look for an exact solution, but solve either a minimization problem or the associated normal equations.

DIFFICULTIES CHARACTERIZING THE DIRECT PROBLEM

All methods based on the direct inversion paradigm make a very strong hypothesis: the direct problem is characterized. This involves not only accurately knowing Θ but also having a priori information about the noise process disturbing (1). Knowing matrix Θ is not a trivial assumption. Its physical meaning varies from application to application and sometimes can be hard to obtain. In general, this matrix is either experimentally measured or modeled using physical information known about a particular problem. When Θ is measured, complex and time-expending experimental procedures are generally involved. When Θ is modeled, physical theoretical insight is usually required. Sometimes both procedures are mixed as some problems present great difficulty.

Noise is normally taken into account when performing a direct inversion. Its characterization typically requires a model that can be estimated by means of statistical analysis over a series of images or signals. Basic assumptions of this model should be carefully considered. Most of the time noise is assumed as Gaussian. In some applications, this is justified by the physics of the problem, but care should be taken as this assumption is sometimes driven by a "simplicity" consideration which is not necessarily appropriate. When possible, it is advised to measure noise histograms in order to have an idea of the underlying probability density function (PDF) of the noise affecting (1).

The difficulties characterizing the direct problem are better understood when illustrated by actual problems. Later in this survey, we describe these difficulties explicitly in a set of examples.

PROBABILISTIC POINT OF VIEW

A general theory about inverse problems is obtained when using a probabilistic point of view, see [14]. Inverse problems are seen as probabilistic inference problems where lack of information makes the inference impossible. Therefore, some kind of a priori knowledge must be used to compensate. In this section, we introduce a simplification of the general theory which applies to nonlinear problems and is outside the scope of this survey. In the simplified framework, (2) is considered in the absence of perturbations caused by noise: $m = \Theta o$. Now, vectors m and o are considered random variables. Consequently, their realizations are driven by PDFs $f_{\mathfrak{D}}(o)$ and $f_{\mathfrak{M}}(m)$ defined over two linear spaces: the model space, $\mathfrak{D}$, and data space, $\mathfrak{M}$. A priori information on the model parameters is represented by $f_{\mathfrak{D}}(o)$. In this context, by use of a Bayesian approach the prior PDF over the model space, $f_{\mathfrak{D}}(o)$, is transformed into the posterior PDF over the data space, $f_{\mathfrak{M}}(m)$. Let us illustrate this approach more precisely, defining the PDFs in the model and data space as multivariate Gaussian distributions:

- The a priori information indicates that the (unknown) model o is a sample of a known Gaussian PDF over the model space $\mathfrak{D}$ whose mean is o_{prior} and whose covariance matrix is Ψ_o:

$$f_{\mathfrak{D}}(o) = c_o\, e^{-\frac{1}{2}(o-o_{\text{prior}})^t \Psi_o^{-1}(o-o_{\text{prior}})}, \qquad (14)$$

where c_o is a constant of normalization not detailed here.

- Measurements of the observable parameters m are assumed to be represented by a Gaussian PDF centered at m_{obs} and with covariance matrix Ψ_m

$$f_{\mathfrak{M}}(m) = c_m\, e^{-\frac{1}{2}(m-m_{obs})^t \Psi_m^{-1}(m-m_{obs})}, \qquad (15)$$

where c_m is a constant.

From these two above PDFs (14) and (15), knowing that the relationship between m and o is linear, the Bayes theorem is applied to find the posterior distribution. We recall that in a Bayesian approach the product of the PDF of m conditional to o [called likelihood, in this case $f_{\mathfrak{M}}(\Theta o)$] and the PDF of o [$f_{\mathfrak{D}}(o)$, called prior] makes, after division by a normalizing constant, the following so-called posterior PDF:

$$f_{\text{posterior}}(o) = \frac{f_{\mathfrak{M}}(\Theta o)\, f_{\mathfrak{D}}(o)}{\int_{\mathfrak{D}} f_{\mathfrak{M}}(\Theta o')\, f_{\mathfrak{D}}(o')\, do'}. \qquad (16)$$

The posterior distribution is also Gaussian and takes the following form:

$$f_{\text{posterior}}(o) = $$
$$c_p\, e^{-\frac{1}{2}(\Theta o - m_{\text{obs}})^t \Psi_m^{-1}(\Theta o - m_{\text{obs}}) - \frac{1}{2}(o-o_{\text{prior}})^t \Psi_o^{-1}(o-o_{\text{prior}})},$$
$$(17)$$

where c_p is a constant. The mean value, $\widetilde{m}$, of this posterior distribution (17) can be calculated [14], leading to

$$\widetilde{m} = \left(\Theta^t \Psi_m^{-1} \Theta + \Psi_o^{-1}\right)^{-1} \left(\Theta^t \Psi_m^{-1} m_{obs} + \Psi_o^{-1} o_{prior}\right). \qquad (18)$$

Equation (18) is in fact a general form of the Wiener operator presented in (10) where o_{prior} is included. Tikhonov regularization can be understood in this context as a way of estimating matrix Ψ_o^{-1}.

From (18), it is easy to observe that the techniques presented in this survey are justified only when all uncertainties (modelization uncertainties, observational uncertainties, uncertainties in the a priori model) are Gaussian. Care should be taken applying these techniques to cases where this assumption is violated. Finally, this section intends to provide a glimpse of the probabilistic interpretation of linear inverse problems. The interested reader should be aware that the treatment of nonlinear problems and non-Gaussian PDFs necessitates the use of more complex mathematics than the ones presented above.

INDIRECT RECONSTRUCTION PROBLEM

The inverse operator can be constructed without characterizing the direct problem. This can be surprising for a beginner, but it is indeed possible. Let us take another look at the discrete equation (2), where the noise term is removed for simplicity

$$\mathbf{m} = \Theta \mathbf{o}. \tag{19}$$

If Θ is unknown, the only available information concerns the other elements of the equation. In some problems, corresponding pairs of vectors $(\mathbf{m}_i, \mathbf{o}_i)$, for $i = 1, \ldots, P$, can be obtained. These corresponding pairs $(\mathbf{m}_i, \mathbf{o}_i)$ depend strongly on the application; their number, precision, or nature of acquisition can be significantly different. This approach can also be called calibration.

Let's now insert in the columns of an $N \times P$ matrix $\mathbf{O}$ all the $\mathbf{o}_i$s and in the columns of a $K \times P$ matrix $\mathbf{M}$ all their corresponding $\mathbf{m}_i$s. The construction of $\mathbf{O}$ and $\mathbf{M}$ allows us to rewrite a set of P equations as (19) in a single matrix expression:

$$\mathbf{O} = \Theta_{\text{Indirect}}^{-} \mathbf{M}, \tag{20}$$

where $\Theta_{\text{Indirect}}^{-}$ is a $N \times K$ matrix representing the inversion of the unknown matrix Θ. A straightforward solution of this linear system would be

$$\Theta_{\text{Indirect}}^{-} = \mathbf{O}\,\mathbf{M}^{-1}, \tag{21}$$

if $\mathbf{M}$ is a full rank square matrix, but usually $P \gg K$. Moreover, the stability of the solution would not be assured because of the presence of noise in $\mathbf{O}$ and $\mathbf{M}$. The problem is ill-posed in the Hadamard sense. However, $\mathbf{M}$ can be inversed using a pseudo-inverse as follows:

$$\Theta_{\text{Indirect}}^{-} = \mathbf{O}\,\mathbf{M}^t (\mathbf{M}\,\mathbf{M}^t)^{-1}. \tag{22}$$

Tikhonov regularization can also be used in (22), as already explained in a preceding section.

MULTIVARIATE LINEAR REGRESSION

If we take a closer look at (22), we observe that this equation represents a multivariate linear regression model, a well-known expression that can be found in any multivariate statistics textbook; see [15] for reference. Multivariate linear regression is the generalization to multiple response variables of the familiar linear regression model; see [16] for an introduction. In fact, the least-squares normal equations of the multivariate linear regression are $\mathbf{Q}\,\mathbf{M}\,\mathbf{M}^t = \mathbf{O}\,\mathbf{M}^t$, where $\mathbf{M}$ is the matrix containing observational data in its columns and $\mathbf{O}$ is a matrix containing the observations of the dependant variable. A basic algebra manipulation gives $\mathbf{Q}$, the so-called least squares estimator of the linear model parameters: $\mathbf{Q} = \mathbf{O}\,\mathbf{M}^t (\mathbf{M}\,\mathbf{M}^t)^{-1}$; obviously this equation is equivalent to equation (22) and $\mathbf{Q} = \Theta_{\text{Indirect}}^{-}$. Indirect reconstruction techniques establish a multivariate linear regression. Thus, the use of regression is mathematically well founded for the resolution of integral equations; indeed, a linear underlying relationship between $\mathbf{m}$ and $\mathbf{o}$ exists. Finally, the use of regression and Tikhonov regularization is common practice in statistics and leads to the so-called Ride regression; see, for instance, [12].

RECONSTRUCTION AS INTERPOLATION

There exists another paradigm for solving an integral equation of an imaging problem. This is a particular case, but it appears in some imaging applications. If we consider an integral equation (1) as a sampling process, the kernel ϕ would become a sampling function, typically a delta Dirac, and $o(\lambda)$ would be the signal to be sampled. This is conceptually different from the paradigms presented previously. The reconstruction of the original image becomes then intimately related to the well-known sampling theory.

When dealing with a linear inverse problem as a sampling problem, we should be aware that a strong assumption is being made about the kernel of the integral equation. Whether the functions composing the kernel can be considered as approximations of sampling functions is completely dependent on the specific application.

EXAMPLE: IMAGE RESTORATION

Restoration of degraded images is a classical ill-posed linear inverse problem in image processing, see [17] or [18]. A degraded image can have different causes, for instance defocusing, motion, noise, or parasite signals. The degradation is normally considered invariant of the position within the image and it is then modeled by a convolution integral:

$$g(x, y) = \iint \text{image}(\alpha, \beta)\, d(x - \alpha, y - \beta)\, d\alpha\, d\beta + n(x, y), \tag{23}$$

where g is the degraded image, x and y index the pixels of the degraded image, d is the cause of the degradation and α and β index the original image, and n represents the noise. Convolution is important in image processing and is also the base for filtering and enhancement of digital images. Note that (23) is a particular case of (1). The fact that the integral in (23) is defined in a bidimensional space does not change any fundamental aspect of our discussion.

In image filtering, enhancement, or reconstruction, knowing operator Θ means that the cause of the degradation of the image is a priori known and characterized. This can be studied in the case where the imaging acquisition tool presenting the default is available for experimentation. In general, when not present or simply when the degradation is of unknown origin, estimating Θ can be difficult. The convolution in (23) is a linear operation normally represented by the operator * as follows:

$$g = image\,^*d + n. \tag{24}$$

This operation becomes a multiplication on the Fourier domain. Consequently, when looking for the nondegraded image from a degraded one (either for filtering, enhancement, or restoration purposes), a deconvolution becomes a division in the Fourier domain:

$$I = G/D, \tag{25}$$

where I, G, and D are, respectively, the Fourier transforms of the image to be restored, the degraded image, and the degradation process. For an actual solution, in (25), noise must be taken into account and zero division must be avoided. This is done by the Wiener filter in its Fourier domain formulation [8], which is one of the oldest solutions to this problem, and continues to be the base of numerous modern image restoration systems. Details about this subject can be found in standard textbooks; see chapter 5 of [17] for an introduction. Care should be taken using Fourier transform techniques as they depend on properties of the specific inverse problem. Well-known established solutions exist based on the fact that a deconvolution is a division in the Fourier domain. However, when confronted with a new inverse problem that is not a convolution, the application of this kind of approach can simply not be possible.

EXAMPLE: MULTISPECTRAL IMAGING AND SPECTRAL REFLECTANCE RECONSTRUCTION

Conventional color digital cameras producing three-band images appear to be limited when high-fidelity color reproduction has to be performed. While high-fidelity sound systems are commercially available devices, high-fidelity color digital cameras are only found, at the moment, in research laboratories. A way of obtaining color high fidelity is to use a camera with more than three bands, leading to so-called multispectral color imaging.

The main components involved in an image acquisition process are depicted in Figure 1. We denote the spectral radiance of the illuminant by $l_R(\lambda)$, the spectral reflectance of the object surface imaged in a pixel by $r(\lambda)$, the spectral transmittance of the optical systems in front of the detector array by $t(\lambda)$, the spectral transmittance of the kth optical color filter by $f_k(\lambda)$ and the spectral sensitivity of the CCD array by $\alpha(\lambda)$. Note that only one optical color filter is represented in Figure 1. In a multispectral system, a set of filters is often set up in a

barrel which rotates to automatically change filters between acquisitions. There also exist systems that do not need any mechanical displacement in order to change the filter transmittance. For instance, liquid crystal tunable filters (LCTFs) provide this capability. They are basically an accumulation of different layers, each layer containing linear parallel polarizers sandwiching a liquid crystal retarder element; see [19] for an example of its use.

Supposing a linear optoelectronic transfer function of the acquisition system, the camera response c_k for an image pixel is then equal to

$$c_k = \int_\Lambda l_R(\lambda) r(\lambda)\, t(\lambda)\, f_k(\lambda)\, \alpha(\lambda)\, d\lambda + n_k$$
$$= \int_\Lambda \phi_k(\lambda)\, r(\lambda)\, d\lambda + n_k, \tag{26}$$

where $\phi_k(\lambda) = l_R(\lambda)\, t(\lambda)\, f_k(\lambda)\quad \alpha(\lambda)$ denotes the spectral sensitivity of the kth channel, n_k is the additive noise, and Λ is the range of the spectrum where the camera is sensible. The assumption of system linearity comes from the fact that the CCD sensor is inherently a linear device. However, for real acquisition systems, this assumption may not hold—for example, due to electronic amplification nonlinearities or stray light in the camera [20], [21]. Stray light may be strongly reduced by appropriate black anodized walls inside the camera. Electronic nonlinearities may be corrected by an appropriate calibration of the amplifiers.

The aim of spectral reflectance reconstruction is to obtain a reflectance curve for each pixel of the image. This is illustrated in Figure 2, which contains a color image of "Saint Jacques le mineur," a painting by Georges de la Tour (Musée Toulouse Lautrec, Albi, France) scanned by a 10-filter multispectral camera [3]. We show below for this painting image two examples of reconstructed reflectances for two different pixels.

DISCRETIZATION OF THE INTEGRAL EQUATION

By uniformly sampling the spectra at N equal wavelength intervals, we can rewrite (26) as a scalar product in matrix notation:

$$c_k = \phi_k^t \mathbf{r} + n_k, \tag{27}$$

where $\mathbf{r} = [r(\lambda_1)\, r(\lambda_2) \ldots r(\lambda_N)]^t$ and $\phi_k = [\phi_k(\lambda_1)\phi_k(\lambda_2) \ldots \phi_k(\lambda_N)]^t$ are vectors containing the sampled spectral reflectance function and the sampled spectral sensitivity of the kth channel of the acquisition system, respectively. The vector $\mathbf{c}_K = [c_1 c_2 \ldots c_K]^t$ representing the responses of all K channels may then be described using matrix notation as

$$\mathbf{c}_K = \Theta\, \mathbf{r} + \mathbf{n}, \tag{28}$$

where $\mathbf{n} = [n_1\, n_2 \ldots n_K]^t$ and Θ is the K-line, N-column matrix defined as $\Theta = [\phi_k(\lambda_n)]$, where $\phi_k(\lambda_n)$ is the spectral sensitivity of the kth channel at the nth sampled wavelength.

SOLUTIONS BASED ON DIRECT INVERSION

DIFFICULTIES CHARACTERIZING THE DIRECT PROBLEM

In spectral reflectance reconstruction, knowing Θ means that a physical characterization of the acquisition system has been performed. This characterization requires at least the measurement of the CCD sensitivity, filter transmittances, and transmittance of the optics. This is illustrated graphically in Figure 3. This characterization involves the realization of physical experiments in which, typically, a monochromator is used for measuring the CCD sensitivity and a spectroradiometrer is used for measuring the spectral transmittances of the filters and of the other optical systems of the camera. For a CCD, the noise model can be considered Gaussian; this assumption is justified by the physics of the problem. To study the noise, a series of images is acquired with the camera lens being occluded with a lens cap or with the whole equipment placed in a dark room.

NORM MINIMIZATION

In [22], the following operator:

$$\Theta_{\text{Hardeberg}}^{-} = \mathbf{R}\mathbf{R}^{t}\Theta^{t}(\Theta\,\mathbf{R}\mathbf{R}^{t}\Theta^{t})^{-1}, \qquad (29)$$

is proposed to regularize a pseudo-inverse. The matrix $\mathbf{R}$ contains a selected set of sampled spectral reflectances which represent well the spectral properties of the objects to be imaged. This approach assumes that any reflectance can be a linear combination of these representative reflectances. For an oil painting digitization, a set of 64 pure pigments provided by the National Gallery in London [21] is used. When comparing Hardeberg's operator with the dual underdetermined operator $\Theta_{\text{N-under}}^{-} = \mathbf{N}\Theta^{t}(\Theta\mathbf{N}\Theta^{t})^{-1}$ [see (9)],

$$c_k = \int_{\Lambda} I_R(\lambda)\ r(\lambda)\ t(\lambda)\ f_k(\lambda)\ \alpha(\lambda)\ d\lambda + n_k$$
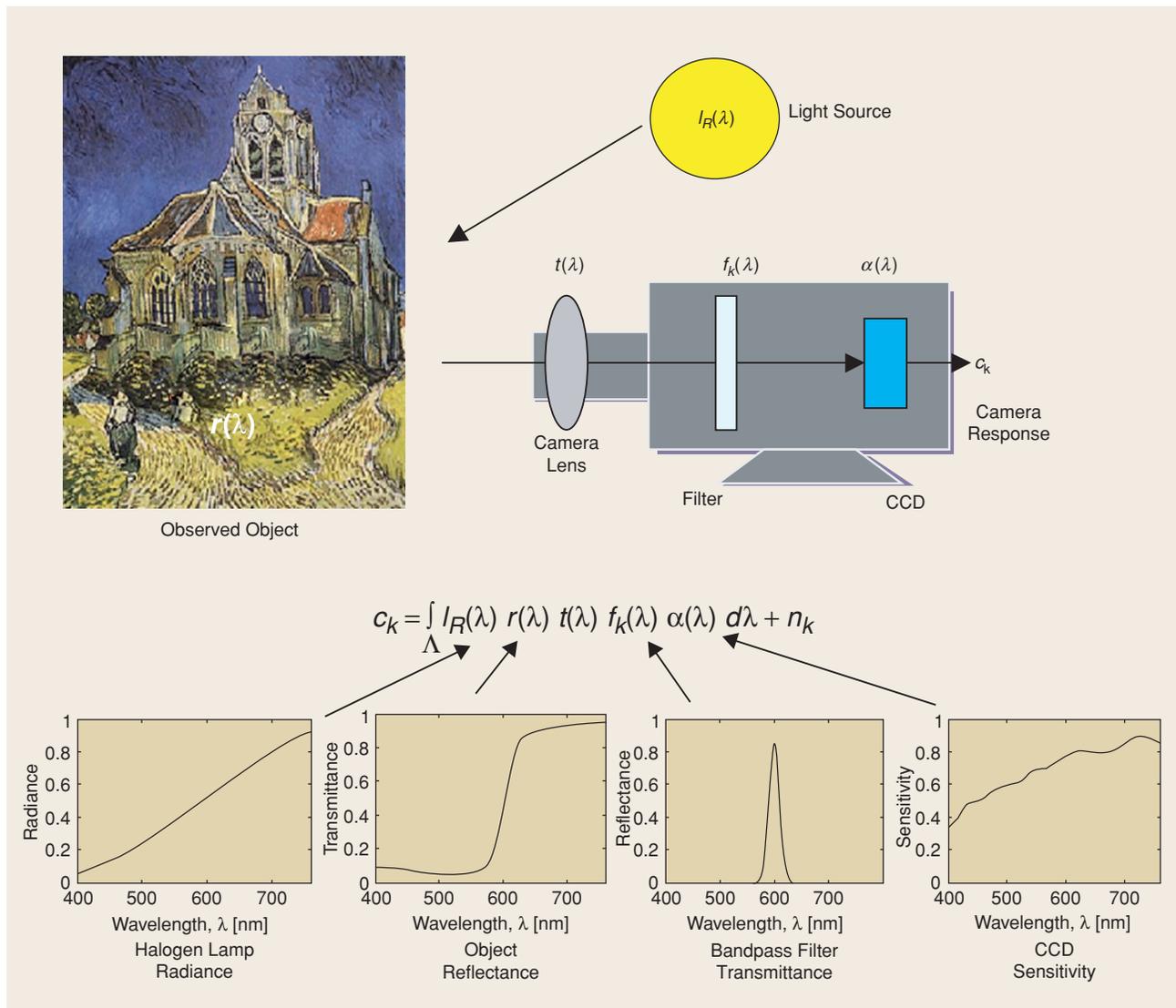
[FIG1] Schematic view of the image acquisition process. The camera response depends on the spectral radiance of the light source, the spectral reflectance of the objects in the scene, the spectral transmittance of the camera lens (not illustrated in the equation), the spectral transmittance of the color filter, and the spectral sensitivity of the sensor.

SMOOTHING INVERSE

From our knowledge, this technique was introduced by Mancill and Pratt, see [23] or Chapter 16 of [18] where the technique is applied to the similar problem of spectral radiance estimation, where the radiance $l_R(\lambda)$ instead of the reflectance $r(\lambda)$ is estimated in (26). The approach used by [24] is directly inspired from both the above cited references. This technique is basically the application of the generalization of the pseudo-inverse to the non-Euclidian distance. This is seen in the dual underdetermined of (9), where $N = N_\Delta$ is a $N \times N$ matrix such that the built operator $\Theta^-_{N_\Delta\text{-under}}$ minimizes the average squared second differences $\Delta = \sum \Delta_i$, where $\Delta = [(r(\lambda_{i+1}) - r(\lambda_i)) - (r(\lambda_i) - r(\lambda_{i-1}))]^2$. Note that Δ_i is a measure of the curvature of the reflectance functions. Unfortunately, N_Δ is a singular matrix and, consequently, it cannot be inverted. The method uses then the following nonsingular matrix:

$$N'_\Delta = N_\Delta + \varepsilon\, I, \tag{30}$$

where I is the identity matrix and ε is a small positive constant ($\varepsilon \ll 1$). The method for determining this parameter is normally not specified. It is, in general, fixed a priori or optimized manually. Automatic optimization is, in general, not needed. Please note the similarities of this stabilization method with the Tikhonov regularization (12).

The smoothing inverse method normally obtains much better results than a normal pseudo-inverse. In fact, the concept of smoothing is natural when considering linear spectral reconstruction. In Figure 4, we show two measured spectral curves (red) along with their reconstructed counterparts using a simple pseudo-inverse (green) and a smoothing inverse (blue), the smoothing matrix being N'_Δ, where ε is 0.01. Both curves are from the GretagMacbeth digital camera (DC) color chart, a well-known test chart for color calibration control. We clearly see that the pseudo-inverse reconstructed curves oscillate around its true value. Smoothing in this case appears well-adapted as part of the reconstruction technique.

WIENER FILTER

As we have already seen, Wiener estimation requires second-order statistics, in this case the covariance matrices of the spectral reflectance curves Ψ_r and of the camera responses Ψ_c. We must estimate the covariance matrices accurately. This entails some experimental work: the covariance matrix Ψ_r is usually estimated from the reflectances of a set of color patches and the
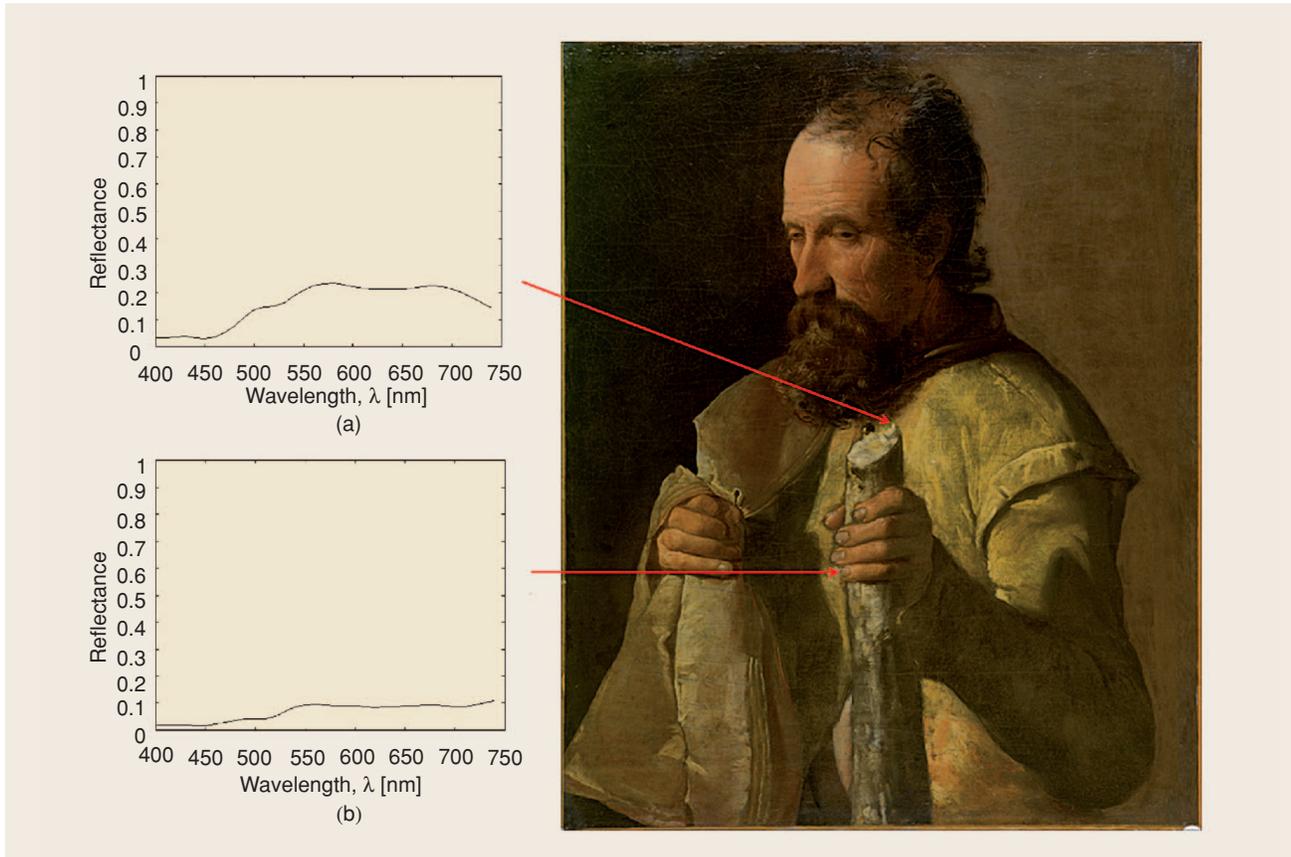
[FIG2] (a) Spectral reconstruction of the reflectance curves on two pixels of the multispectral image. The spectral reflectance curves have been estimated using an indirect reconstruction method. (b) Painting "Saint Jacques le mineur" by Georges de la Tour (Musée Toulouse Lautrec, Albi, France).

covariance matrix Ψ_c is estimated by measuring the noise properties of the CCD camera actually used. This leads to

$$\Theta_{\text{Wiener}}^- = \Psi_r^{-1}\Theta^t \left(\Theta\,\Psi_r^{-1}\Theta^t + \Psi_c^{-1}\right)^{-1}, \qquad (31)$$

which is a direct application of the dual underdetermined of expression (10). In the early work of [23], they propose some approximations of the covariance matrices. They model Ψ_r as a first-order Markov process covariance matrix of the form $\Psi_r = \sigma_r^2\,M_\rho$, where ρ, $0 \le \rho \le 1$ is the adjacent element correlation factor, M_ρ describes the structure of the Markov process, and σ_r^2 represents the variance of the reflectances. A white noise model is chosen with a covariance matrix $\Psi_c = \sigma_n^2\,I$, where σ_n^2 is the noise variance and I is the identity matrix. This example illustrates how assumptions about the structure of the covariance matrices can be made when not enough physical information can be obtained from the imaging system. The less information we have, the more assumptions we make. In fact, the choice of this noise model is not adequately justified when applied to spectral reconstruction. When carefully studying noise sources in multispectral images, this noise model can be much more complex. See, for instance, [25] where a much more complete approach for modeling Ψ_c is used.

SOLUTIONS BASED ON INDIRECT INVERSION

Indirect reconstruction is possible when spectral reflectance curves of a set of P color patches are known and an image of these patches is acquired by a multispectral camera. From this data, a set of corresponding pairs (c_p, r_p) for $p = 1, \ldots, P$, is obtained, where c_p is a vector of dimension K containing the camera responses and r_p is a vector of dimension N representing the spectral reflectance of the pth patch. Corresponding pairs (c_p, r_p) are easy to obtain; professional calibrated color charts such as the GretagMacbeth DC are sold with the measurements of the reflectances of their patches. In addition, if a spectroradiometer is available, performing the measure is a fairly simple experiment. Obtaining the camera responses from the known spectral curves of the color chart is just a matter of taking a multispectral image. In Figure 5, we illustrate the indirect reconstruction approach in the case of spectral reflectance reconstruction.

A straightforward solution is given by (22), when applied to this case is

$$\Theta_{\text{Indirect}}^- = R\,C^t(C\,C^t)^{-1}, \qquad (32)$$

where R is a $N \times P$ matrix with columns containing all the r_ps and C is a $K \times P$ matrix with columns containing their corresponding c_ps.

PRINCIPAL COMPONENT ANALYSIS AND REGRESSION

This method introduced by Burns [26] is basically a multivariate regression, but instead of looking for the operator that matches matrices R and C, we look for an operator that matches another matrix A and C. This matrix A is calculated from R by principal component analysis (PCA), which is a statistical technique for data dimensionality reduction [11]. SVD described in (13) is intimately related to PCA. In fact, A is a reduced diagonal matrix
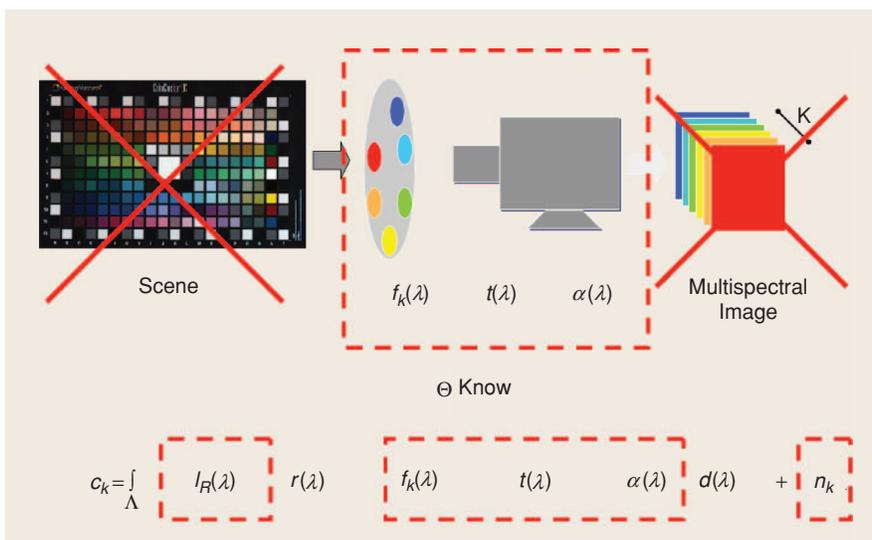
[FIG3] For spectral reflectance reconstruction, direct inversion requires the characterization of CCD sensitivity, filters transmittances, optics transmittance, illuminant radiance, and the characterization of the noise.
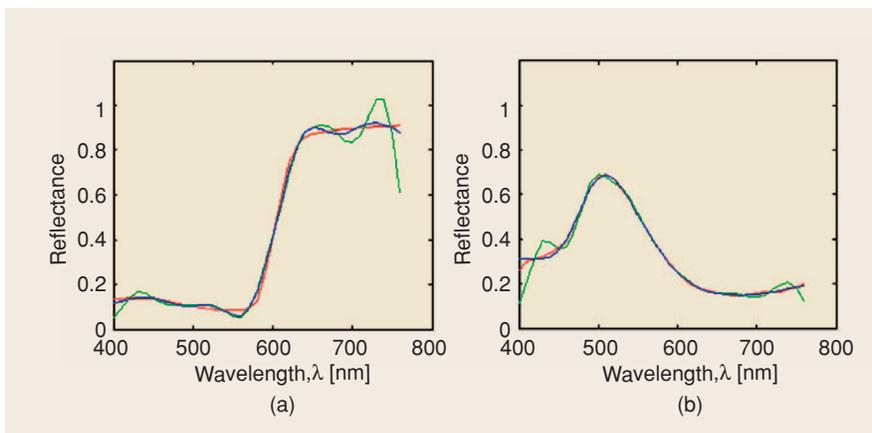
[FIG4] Spectral curves from the Macbeth DC color chart (red) along with their reconstruction using the pseudo-inverse (green) and the smoothing inverse (blue): (a) sample number 4 of the Macbeth DC and (b) sample number 164.
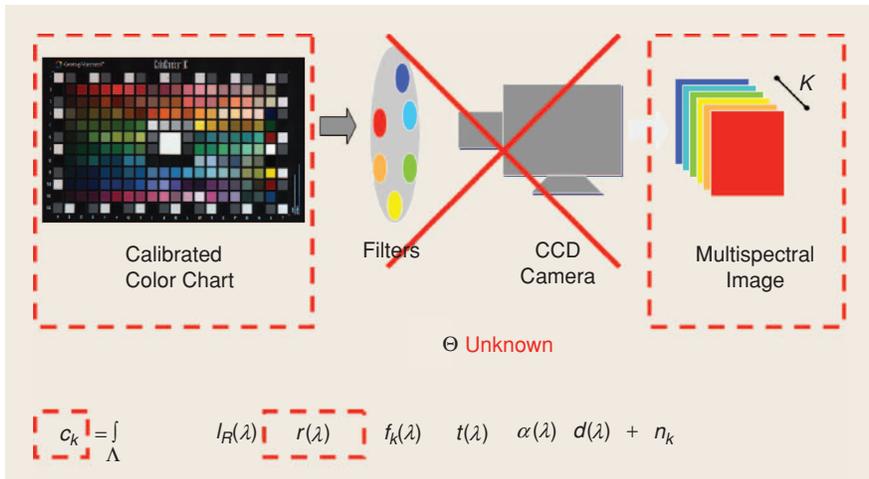
[FIG5] Indirect inversion requires a set of color patches and the acquisition of a multispectral image. From there, corresponding pairs of spectral reflectance curves and their camera responses are extracted. These pairs are later organized as matrices that will be used to calculate the reconstruction operator.

containing the set of the most representative PCA coefficients. A relationship between representative PCA coefficients and channel responses C can be established by the operator $A C^t (C C^t)^{-1}$. This example is interesting as it is clearly a variation of a multivariate regression. Let's note that another operation is needed to obtain the sampled spectral reflectance curve from the set of estimated PCA coefficients. This is obtained by a matrix multiplication, a projection from the orthogonal space defined by the PCA to the original reflectance space. If we call this matrix E_p, we can redefine the operator to estimate directly the spectral reflectance curves as

$$\Theta_{\text{pinvPCA}}^- = E_p \, A C^t \, (C C^t)^{-1}. \qquad (33)$$

If we compare (32) with this operator, we see that matrix R is simply replaced by $E_p A$. In fact, the concept of "representative PCA coefficients" is very similar to the use of a Truncated SVD in the context of regression; see [9].

NONAVERAGED PSEUDO-INVERSE
This technique is based on (32). It uses a conceptually easy and interesting way of introducing noise information in the system. It was first introduced in [27]. The key idea is simple, but the experimental detail must be explained in order to understand how this method works. Determining R and C implies classically the following two-step experiment based on a chart containing color patches:

1) Measure the spectral reflectance of all the patches of the color chart using a spectroradiometer. Each sampled spectral curve is stored in a column of the matrix R.
2) Acquire a multispectral image of the color chart. A window is superposed onto each patch and its mean value calculated per patch and per channel. The mean values of a patch form a vector of camera responses that is stored in a column of the matrix C. This column has the same index as the corre-

sponding spectral reflectance column in matrix R.

The technique is based on taking advantage of the information acquired in the second step above. Instead of calculating a mean value for the windows superposed onto the patches, this technique uses all the values contained in the window to build a large C matrix. Of course, matrix R is expanded to have a corresponding spectral curve column for every column of C. This makes the matrices very large. For instance, a typical color chart can contain 200 patches. When using the classical mean value approach, if ten filters are used and the spectral curves have 40 samples, matrix R has dimensions 40×200 while matrix C is 10×200. When using this new approach and a small window of 10×10 pixels, the sizes of the matrices are multiplied by 100. Matrix C becomes $10 \times 20,000$ and matrix R is $40 \times 20,000$. We can easily imagine that when the window increases its dimensions the size of the system formed by R and C can become huge. This implies much more computation time.

The advantage of this method is that it somehow automatically captures the acquisition noise model. We can consider every pixel of a window as a realization of a random noise process. Then, using all the values in the analysis windows as samples to solve the spectral reconstruction problem means that we implicitly take noise into account. In fact, this approach is interesting because noise is not explicitly modeled, no assumption about its distribution being required.

SOLUTIONS BASED ON INTERPOLATION
A multispectral system can be considered as a tool that samples spectral reflectance curves. Instead of using delta Dirac functions for the sampling as in the classical framework, the spectral transmittance functions $f_k(\lambda)$ of the K filters are considered to be the sampling functions. This approach just requires the camera response itself, c. The methods based on this paradigm interpolate the camera responses acquired by a multispectral camera by using a smooth curve. The smoothness properties of the interpolating curve introduce a natural constraint which regularizes the solutions. However, there are two underlying problems to take into account before representing the camera responses in the same space as spectral curves:

■ Positioning the camera response samples in the spectral range can be a challenge. For instance, in the case of Gaussian-shape filter the camera responses can be positioned at the center of the filter. But, real filters are rarely Gaussian-shaped. In general, it is admitted that if a filter is narrow, positioning the camera responses can be done with low uncertainty. Unfortunately, when wide filters are

used this uncertainty increases with the width of the filter. This is the reason why interpolation methods should be used only with multispectral cameras using narrow band-pass filters.

■ The camera must be radiometricaly calibrated. This means that camera responses must be normalized, i.e., to belong to the [0,1] interval for all the camera channels. In high-end applications, this normalization implies the use of a radiometric standard white patch. This reference patch is imaged for normalization as part of a calibration procedure.

Most practical applications of interpolation to spectral reconstruction are used in cases where the CCD is cooled and Gaussian-like filters are available, see [28]. Such methods are reported not to be well-adapted to filters having more complex wide-band responses, suffering from quite severe aliasing errors [24], [26].

> THE INVERSION OF AN INTEGRAL OR RESOLUTION OF AN INTEGRAL EQUATION CAN BE DIFFICULT TO SOLVE; PROPERLY SPEAKING, IT CAN BE AN ILL-POSED PROBLEM.

Cubic splines have been applied in this context by [29]. They are well-adapted to the representation and reconstruction of spectral reflectance curves because they generate smooth curves, C^2 continuity being assured. A technique called modified discrete sine transform (MDST) was also introduced by [29]. It is based upon Fourier interpolation.

SIMULATION

When an imaging process is modeled by an integral equation, its simulation is straightforward: in a discrete system, it becomes a series of vector and matrix products. This can be used to better understand the nature of the problem and to perform some analysis as a quantitative study on the performance of different elements integrating the studied imaging process. For instance, in multispectral imaging, the shape of the filters defining the spectral bands of the camera can be optimized for better final image quality [30]. Another interesting example is the simulation of the color of a scene under various illuminants. When the spectral reflectance reconstruction problem is solved an image containing a reflectance curve in each of its pixels is obtained. As the reflectance $r(\lambda)$ is now known in (26), the illuminant $l_R(\lambda)$ can be introduced to simulate a color image. In fact, a color camera is defined by three known color matching functions $\phi_k(\lambda)$, $K = 1, .. , 3$, corresponding to a color space such as sRGB [31].

Figure 6 shows results of the illuminant simulation for an image of the head of *La Baigneuse*, a painting by Renoir (Musée de l'Orangerie, Paris, France). This simulation was part of the European project CRISATEL where a multispectral acquisition system for paintings was developed [32]. This kind of simulation has numerous applications in art preservation. A curator would find a tool implementing illuminant simulation very useful when deciding the appropriate light sources for an art exhibition with the original or, for instance, when having to produce a high-quality printed reproduction of a painting under specific lighting conditions. It is also possible to simulate

the aging of a painting by introducing the spectral transmittance of an old varnish in (26).

EXAMPLE: PARALLEL MRI IMAGING

Linear inverse problems also appear in systems of a nonoptical nature, such as an MRI. Parallel MRI is a rapid acquisition technique considered as one of the modern revolutions in the field of MRI. All current medical MRI equipment offer the ability to use this technique.

For readers not familiar with this imaging modality, it should be understood that a medical MRI acquisition system can be conceived as a machine which physically performs an FT of the properties of human tissue. The acquisition process in a voxel of the produced images is indeed an FT of a tissue contrast function, which depends on the hydrogen spins present in this voxel. This process is mathematically represented by

$$\text{k-space}_l(\mathbf{k}) = \int_V s_l(\mathbf{r})\, t(\mathbf{r})\, e^{-i\mathbf{k}.\mathbf{r}} d\mathbf{r} + n_l(\mathbf{k}), \quad (34)$$

where $\mathbf{r}$ and $\mathbf{k}$ are position vectors in the image domain and in the frequency domain, respectively, and $\mathbf{k}.\mathbf{r}$ represents their inner product. Here $t(\mathbf{r})$ is the tissue contrast function at position $\mathbf{r}$, $s_l(\mathbf{r})$ is the sensitivity map of the lth receiver coil used to receive the MRI signal, $l = 1, .. ,$ and k-space$_l(\mathbf{k})$ represents a measurement performed by the MRI system in the so-called k-space. Finally V is called volume of interest, the volume taken into consideration for the imaging process. Classical MRI systems are equipped with a single reception coil ($L = 1$) whose unique sensitivity map is as spatially flat as possible. This enables taking the term $s_l(\mathbf{r})$ out of (34), and then the reconstruction of an image becomes (apart from artifacts correction) an inverse FT of the k-space.

Parallel MRI systems are equipped with multiple coil receivers. They were first introduced for image signal-to-noise ratio (SNR) improvement [33]. Subsequently, they have been used to accelerate image acquisition in [34] and [35]. When used for rapid acquisition, the k-spaces of each coil are incompletely acquired. Each coil observes the same MRI signal, but this signal is modulated by the coil individual sensitivity profile $s_l(\mathbf{r})$.

In the rest of this section, we consider the parallel MRI reconstruction problem from a linear algebra point of view [36]. The problem can be defined as follows: Different coils having different sensitivity profiles, this introduces a spatial encoding effect; how can this effect be used to compensate the aliasing generated by the incomplete acquisitions?

DISCRETIZATION OF THE INTEGRAL EQUATION

Strictly speaking, (34) is a system of L integral equations with a common unknown $t(\mathbf{r})$. In the general case, this equation can be discretized by considering the following matrix:

$$E_{(l,\kappa),\rho} = s_l(\mathbf{r})\ e^{-i\mathbf{k}_\kappa \mathbf{r}_\rho}, \qquad (35)$$

where $(l, \kappa), \rho$ appear as subindexes emphasizing that matrix $\mathbf{E}$ has dimensions $(L \times K) \times N$, L is the number of coils $(l = 1, .., L)$, K is the number of samples in the k-space $(\kappa = 1, .., K)$, and N is the number of voxels in the image

$(\rho = 1, .., N)$. The construction of matrix $\mathbf{E}$, the encoding matrix, leads to the following discretization of (34):

$$\mathbf{k\text{-}space} = \mathbf{E}\, \mathbf{t}, \qquad (36)$$

where **k-space** is a vector containing $(L \times K)$ k-space samples from all coils and $\mathbf{t}$ is a N vector containing all the voxels of the
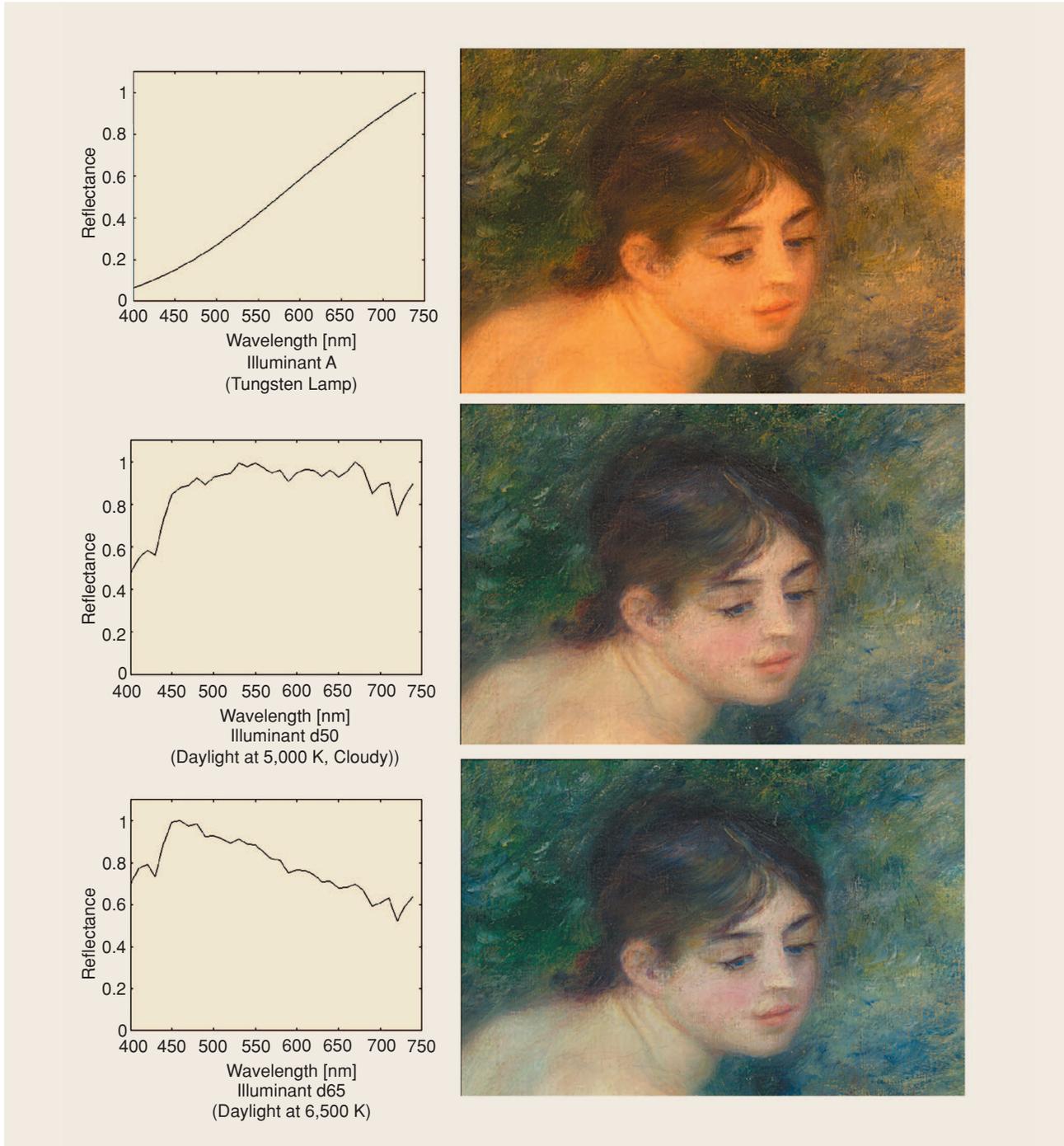
[FIG6] Illuminant simulation to produce an sRGB color image of the head of "La Baigneuse," painted by Renoir (Musée de l'Orangerie, Paris, France). Using spectral reflectance curves reconstructed at each pixel, we simulate the appearance of the painting under different illuminations by simply changing the spectral radiance $I_R(\lambda)$. Note that blue being a prominent element in this painting, the visual appearance is seriously altered by the change of illuminant.

image to be reconstructed. This equation is equivalent to the general expression in (19), but the size of matrix E is huge. The methods presented in this survey are not directly applicable to such a matrix; iterative algorithms should be used. In [37], an iterative solution of (36) is presented; however, it is computationally intense. This kind of solution is found in current problems, for instance, when the k-space is sampled following a spiral trajectory [38].

More insights into this problem can be made by taking a close look at the structure of the encoding matrix. We just emphasize the fact that the terms $e^{-i\mathbf{k}_\kappa \mathbf{r}_\rho}$ in (35) can be considered in a separate matrix that is the FT operator $\tilde{\mathbf{F}}$. Then (36) becomes

$$\mathbf{k\text{-space}} = \tilde{\mathbf{F}}\,\mathbf{S}\,\mathbf{t}\,, \qquad (37)$$

where $\mathbf{S}$ is the sensitivity matrix representing the sensitivity profiles of the coils. From this new equation, it is obvious that an inverse FT, $\tilde{\mathbf{F}}^{-1}$, can be applied to the vector $\mathbf{k\text{-space}}$ to obtain $\mathbf{i} = \mathbf{S}\,\mathbf{t}$, where $\mathbf{i}$ is a vector containing all the voxels of the image. This transforms the reconstruction problem to the so-called image domain. However, the reconstruction can be performed directly in the k-space domain: a method that historically exemplifies this approach is SMASH [39].

SOLUTIONS BASED ON DIRECT INVERSION

In fact, (37) is rarely solved in the general case but in a simplified framework. The most common simplification is to consider the k-space as a regular or Cartesian grid. This casts the full reconstruction problem into a series of small reconstruction problems, and then they are solved separately. We show in Figure 7 an example where Cartesian k-spaces have been subsampled by a factor of two, and where four coils have been used for acquisition. The corresponding pattern in the image domain is shown on the left. The four acquired k-spaces have been inverse Fourier transformed. The resulting images are shown on the right side of the figure. They present strong aliasing. From this data the left-hand side image should be reconstructed.

DIFFICULTIES CHARACTERIZING THE DIRECT PROBLEM

In parallel MRI, characterizing the direct problem means that the sensitivity profiles or maps of the coils have been measured. Unfortunately, direct

measurement of the sensitivity is not possible. Thus, the maps must be estimated indirectly from patient images. We show in Figure 8 four images of a water phantom: as this phantom is homogenous, $t(\mathbf{r})$ in (34) becomes constant and the four images can be considered as a physical approximation of the coil sensitivity maps $s_l(\mathbf{r})$. Estimates of the actual sensitivity maps for each coil are generally obtained from patient data by using nonsubsampled images: $\hat{s}_l(\mathbf{r}) = i_l(\mathbf{r})/i_{\text{uniform}}(\mathbf{r})$, where i_l is the full image from coil l and i_{uniform} is the full image with uniform sensitivity profile, obtained either using a traditional homogeneous sensitivity coil or the sum of squares of the individual coil images. Maps computed this way only cover areas where intensity $t(\mathbf{r})$ and sensitivity $s_l(\mathbf{r})$ are high. Interpolation-extrapolation methods must be applied to expand the sensitivity maps. For example, local polynomial fitting is used in [40] to reduce noise in the maps and provide local extrapolation. However, generating a sensitivity map remains a nontrivial task; see [41] for a description of an automatic method for their calculation. Moreover, it can be a problem in images presenting large areas with very low signal. That is the case,
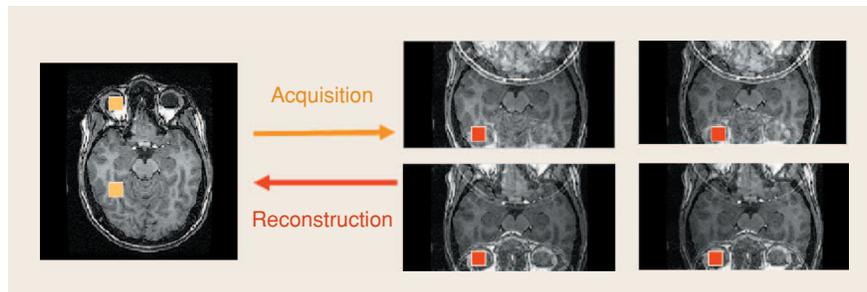
[FIG7] Parallel MRI inverse problem. The k-spaces have been subsampled by a factor of two in order to accelerate the acquisition. The MRI system was equipped with four coils. Both orange left-hand side voxels superpose on the four red voxels belonging to the acquired images.
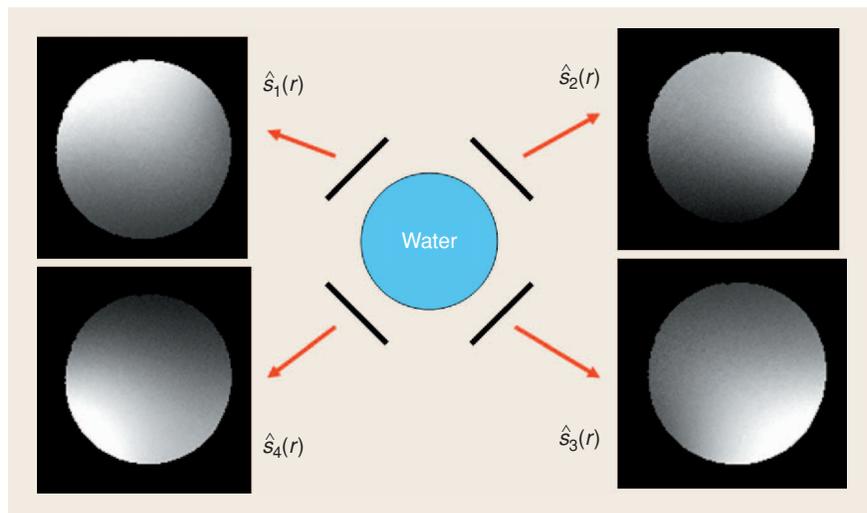
[FIG8] A water phantom is imaged with four coils (center) in order to obtain a rapid approximation of the sensitivity maps $\hat{s}_l(\mathbf{r})$, $l = 1, \ldots, 4$ (left and right sides). These images come from a real experiment on a General Electric 1.5 MRI using a four-coil TORSO system.

for example, in pulmonary MRI where the air inside the lungs produces no signal [42].

SENSITIVITY INVERSION

One of the most well-known reconstruction techniques in parallel MRI is SENSE (sensitivity encoding) [40]. This reconstruction technique works on the image domain and corresponds to the simplified framework presented in Figure 7. In this figure, we can see that the four red pixels on the acquired images match somehow the two orange pixels on the reconstructed image. In fact, this match corresponds to a linear noisy mix of signals coming from known equidistant positions in an ideal nonaliased image.

We note R the number of voxels superimposed in the original image (or reduction factor) and L the number of coils used in the acquisition system. In the core of a reconstruction system, a vector $\mathbf{r}$ of complex numbers containing R reconstructed image intensities is generated from another complex vector $\mathbf{f}$ containing L measurements. These measurements correspond to the same voxel position in the aliased images acquired by the coils. This is performed by means of an $R \times L$ complex unfolding matrix, Θ_{U}^{-}. Mathematically, the relationship is expressed as

$$\mathbf{r} = \Theta_{\mathrm{U}}^{-}\, \mathbf{f}. \qquad (38)$$

Equation (38) is indeed the definition of any linear reconstruction method working in the image domain. The manner of building matrix Θ_{U}^{-} is what differs among various methods. For SENSE [40], this matrix is

$$\Theta_{\mathrm{SENSE}}^{-} = (\mathbf{S}^{H}\Psi^{-1}\mathbf{S})^{-1}\mathbf{S}^{H}\Psi^{-1}, \qquad (39)$$

where $\mathbf{S}$ is the sensitivity matrix and Ψ is the receiver noise covariance matrix. The superscript ^{H} indicates a Hermitian transposition as the MRI signal is a complex signal. Equation (39) corresponds to a pseudo-inverse modified by a covariance matrix, a special case of a Wiener filter shown in (9). Note that the problem is overdetermined as the number R of voxels to be reconstructed is smaller than the number L of coils. This is easily seen in Figure 7 where the two orange voxels ($R = 2$) have to be reconstructed from the four red voxels ($L = 4$); in this case, the reconstruction matrix Θ_{U}^{-} has dimensions of 2×4.

Efforts have been made to increase the SNR of the images reconstructed by SENSE. Regularization has been proposed for the first time by [43], and we can see state-of-the-art results in [44] where Tikhonov regularization is used and the parameter λ is optimized by the use of the L-curve [9].

SOLUTIONS BASED ON INDIRECT INVERSION

Indirect reconstruction applied to parallel MRI has one strong advantage: sensitivity maps of the reception coils do not need to be estimated. In parallel MRI, existing indirect reconstruction techniques work in the k-space and not in the image domain. This means that a regression operator is constructed using elements of the Fourier domain. This kind of technique is also called autocalibrated because it is based on the acquisition of some extra information, the auto calibration signal (ACS) lines; they correspond to lines from each coil located at the low-frequency coordinates of the k-space. The most popular among these methods is currently GRAPPA (generalized autocalibrating partially parallel acquisitions) [45]. In fact, GRAPPA seeks to form estimates of the full k-space of each coil, which are combined to obtain the reconstructed image. Each coil's full k-space is estimated by the use of a regression between neighboring k-space points, matrix $\mathbf{M}$, and ACS lines, matrix $\mathbf{O}_{\mathrm{ACS}}$:

$$\mathbf{O}_{\mathrm{ACS}} = \Theta_{\mathrm{GRAPPA}}^{-}\, \mathbf{M}. \qquad (40)$$

Equation (40) is equivalent to (20). The exact format of matrices $\mathbf{O}_{\mathrm{ACS}}$ and $\mathbf{M}$ can be deduced from [45]. Note that in [45] the regression is not shown in matrix form but explicitly.

CONCLUSIONS

Classical techniques for solving linear inverse problems have been presented. Our aim was to show how these classical techniques are applied in current state-of-the-art imaging systems. Moreover, we have provided a classification of the techniques into four families: FT-based, direct reconstruction, indirect reconstruction, and interpolation. We hope that this classification will guide the curious reader into a discipline with a rich bibliography and sometimes sophisticated mathematics. In this survey, we skipped complicated methods to solve inverse problems. Through our examples, we have tried to emphasize the large variety of applications of linear inverse problems in imaging. Two main examples have been examined more deeply in this survey. We hope they have helped the reader to understand the application of the general techniques in two interesting contexts: multispectral imaging and magnetic resonance imaging.

AUTHORS

Alejandro Ribés (alejandro.ribes@gmail.com) received a computer science and engineering degree (bachelor's and master's) from

the University Jaume I, Castelló, Spain. He also holds a master's in image processing and computer vision from the University of Nice-Sophia Antipolis, France, and a Ph.D. in multispectral imaging applied to fine art paintings from the Ecole Nationale Supérieure des Télécommunications, Paris, France. He has been a postdoctoral fellow at the French Atomic Energy Commission, Orsay, France working on parallel MRI reconstruction. He has been a research assistant at the University of Oxford, U.K., and a lecturer at the Computer Science Department of Ecole Polytechnique, Palaiseau, France. He is currently a visiting scholar at the National Yang-Ming University, Taipei, Taiwan.

Francis Schmitt (francis.schmitt@telecom-paristech.fr) received an engineering degree from Ecole Centrale de Lyon, France, in 1973 and a Ph.D. degree in applied physics from the University Pierre et Marie Curie, Paris, in 1979. He has been a member of Télécom ParisTech (Ecole Nationale Supérieure des Télécommunications) since 1973. He is currently a full professor at the Image and Signal Processing Department and head of the image processing group. His main interests are in color image analysis and synthesis, multispectral imagery, colorimetry, computer vision, 3-D modeling, and image and 3-D object indexing. He has authored or coauthored about 200 publications in these fields.

REFERENCES

[1] I. Kant, *Critique of Pure Reason*, Cambridge, U.K.: Cambridge Univ. Press, 1787.

[2] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Bull. Univ. Princeton*, 1902, pp. 49–52.

[3] C. Lahanier, F. Schmitt, P. Le Bœuf, and G. Aitken, "Multi-spectral digitisation and 3-D modelling of paintings and objects for image content recognition, image classification and multimedia diffusion. An ontology access to the C2RMF database and library using the CIDOC-CRM," in *Proc. Int. Conf. Museum Digitization, Antiquities, Painting and Calligraphy*, National Palace Museum, Taiwan, 2003, pp. 157–201.

[4] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems. Mathematics and Its Applications*. Norwell, MA: Kluwer, 1996.

[5] G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, vol. 14, no. 4, pp. 651–667, 1977.

[6] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*. New York: Academic, 1972.

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

[8] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA: MIT Press, 1942.

[9] P.C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia: SIAM, 1998.

[10] R. Penrose, "A generalized inverse for matrices," in *Proc. Cambridge Philosophical Society*, 1955, vol. 51, no. 3, pp. 406–413.

[11] G.H. Golub and C.F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1983.

[12] G.H. Golub, M.T. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[13] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[14] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameters Estimation*. Philadelphia, PA: SIAM, 2005.

[15] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley Series in Probability and Statistics), 3rd ed. New York: Wiley, 2003.

[16] D.C. Montgomery, E.A. Peck, and G.G. Vining, *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley, 2001.

[17] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, 2nd Ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[18] W.K. Pratt, *Digital Image Processing*, 2nd ed. New York: Wiley, 1991.

[19] H. Brettel, J.Y. Hardeberg, and F. Schmitt, "WebCam for interactive multispectral measurements," in *Proc. Colour Image Science CIS 2000 Conf.*, Derby, U.K., 2000, pp. 160–168.

[20] J.E. Farrell and B.A. Wandell, "Scanner linearity," *J. Elect. Imag. Color.*, vol. 2, no. 3, pp. 225–230, 1993.

[21] H. Maître, F. Schmitt, J. Crettez, Y. Wu, and J.Y. Hardeberg, "Spectrophotometric image analysis of fine art paintings," in *Proc. IS&T and SID Fourth Color Imaging Conf.*, Scottsdale, AZ, 1996, pp. 50–53.

[22] J.Y. Hardeberg, F. Schmitt, H. Brettel, J. Crettez, and H. Maître, "Multispectral image acquisition and simulation of illuminant changes," in *Color Imaging Vision and Technology*, L.W. MacDonald and M.R. Luo, Eds. New York: Wiley, 1999, pp. 145–164.

[23] W.K. Pratt and C.E. Mancill, Spectral estimation techniques for the spectral calibration of a color image scanner, *Appl. Opt.*, vol. 15, no. 1, pp. 73–75, 1976.

[24] F. König and W. Praefcke, "A multispectral scanner," in *Color Imaging. Vision and Technology*, L.W. MacDonald and M.R. Luo, Eds. New York: Wiley, 1999, pp. 129–144.

[25] H. Haneishi, T. Hasegawa, N. Tsumura, and Y. Miyake, "Design of color filters for recording art works," in *Proc. IS&T 50th Ann. Conf.*, Cambridge, MA, 1997, pp. 369–372.

[26] P.D. Burns, *Analysis of Image Noise in Multitraitement Color Acquisition*, Ph.D. dissertation, Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, 1997.

[27] F.H. Imai, L.A. Taplin, and E.A. Day, "Comparison of the accuracy of various transformations from multi-band image to spectral reflectance," Tech. Rep., Rochester Institute of Technology, Rochester, NY, 2002.

[28] G.P. Herzog and B. Hill, "Multispectral imaging and its applications in the textile industry and related fields," in *Proc. PICS03: The Digital Photography Conf.*, Rochester, NY, 2003, pp. 258–263.

[29] T. Keusen, "Multispectral color system with an encoding format compatible with the conventional tristimulus model," *J. Imag. Sci. Tech.*, vol. 40, no. 6, pp. 510–515, 1996.

[30] M. Mahy, P. Wambacq, L. Van Eyken, and A. Oosterlinck, "Optimal filters for the reconstruction and discrimination of reflectance curves," in *Proc. IS&T and SID Second Color Imaging Conf.*, Scottsdale, AZ, 1994, pp. 140–143.

[31] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "*A standard default color space for the Internet—sRGB*," Int. Color Consortium, Reston, VA, 1996.

[32] A. Ribés, F. Schmitt, R. Pillay, and C. Lahanier, "Calibration, spectral reconstruction and illuminant simulation for CRISATEL: An art paint multispectral acquisition system," *J. Imag. Sci. Tech.*, vol. 49, no. 6, pp. 463–473, 2005.

[33] P.B. Roemer, W.A. Edelstein, C.E. Hayes, S.P. Souza, and O.M. Mueller, "The NMR phased array," *Magn. Reson. Med.*, vol. 16, no. 2, pp. 192–225, 1990.

[34] J.W. Carlson and T. Minemura, "Imaging time reduction through multiple receiver data acquisition and image reconstruction," *Magn. Reson. Med.*, vol. 29, no. 5, pp. 681–688, 1993.

[35] J.B. Ra and C.Y. Rim, "Fast imaging using subencoding data sets from multiple detectors," *Magn. Reson. Med.*, vol. 30, no. 1, pp.142–145, 1993.

[36] W.S. Hoge, D.H. Brooks, B. Mandore, and W.E. Kyriakos, "A tour of accelerated parallel MR imaging from a linear system perspective," *Concepts Magn. Reson. Part A*, vol. 27A, no. 1, pp. 17–37, Sept. 2005.

[37] K.P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding for arbitrary k-space trajectories," *Magn. Reson. Med.*, vol. 46, no. 4, pp. 638–651, 2001.

[38] B.C. Ahn, J.H. Kim, and Z.H. Cho, "High-speed spiral-scan echo planar NMR imaging," *IEEE Trans. Med. Imaging,* vol. MI-5, no. 4, pp. 2–7, 1986.

[39] D.K. Sodickson and W.J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, pp. 591–603, 1997.

[40] K.P. Pruessmann, M. Weiger, M.B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.

[41] C.A. McKenzie, E.N. Yeh, M.A. Ohliger, M.D. Price, and D.K. Sodickson, "Self-calibrating parallel imaging with automatic coil sensitivity extraction," *Magn. Reson. Med.*, vol. 47, no. 3, pp. 529–538, 2002.

[42] M.A. Gonzalez Ballester, Y. Machida, Y. Kassai, Y. Hamamura, and H. Sugimoto, "Robust estimation of coil sensitivities for RF subencoding acquisition techniques," in *Proc. 9th Ann. Meeting ISMRM*, Glasgow, Scotland, 2001, p. 799.

[43] K.F. King and L. Angelos, "SENSE image quality improvement using matrix regularization," in *Proc. 9th Ann. Meet. ISMRM*, Glasgow, Scotland, 2001, p. 1771.

[44] F.H. Lin, K.K. Kwong, J.W. Belliveau, and L.L. Wald, "Parallel imaging reconstruction using automatic regularization," *Magn. Reson. Med.*, vol. 51, no. 3, pp. 559–567, 2004.

[45] M.A. Griswold, P.M. Jakob, R.M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, 2002.

SP