Introduction

When a physical quantity is not directly accessible for measurement, it is common to proceed by observing other quantities that are connected with it by physical laws. The notion of an *inverse problem* corresponds to the idea of inverting these physical laws to gain indirect access to the quantity we are interested in.

For example, in electromagnetism, calculating the electric field induced by a known distribution of electric charges is a *direct problem*, i.e., a problem posed in "the natural direction" of physics as we are used to practising and controlling it. Deducing the distribution of the electric charges from measurements of the field is, on the other hand, an *inverse problem*.

Similarly, in signal processing, modeling a transmission channel that introduces distortion, interference and parasitic signals corresponds to solving a direct problem. Reconstructing the shape of a signal input to the channel from measurements made at the output is an inverse problem.

The situation where the quantity of interest is directly accessible for measurement is obviously more favorable. Nevertheless, *direct measurement* does not signify *perfect measurement*: the *instrumental response* of a piece of measuring apparatus and the various error sources connected with the observation process (systematic error, fluctuations connected with the physical sensors or electronic components, quantization, etc.) are degradations that can also be encompassed in the question of inversion.

When all is said and done, the concept of inverse problems underlies the processing of experimental data in its broadest sense. In the experience of the authors of this book, making it explicit that a data processing chain actually carries out an inversion is also often a very worthwhile exercise. It brings to light *ad hoc* hypotheses and arbitrary

---

Introduction written by Jérôme IDIER.

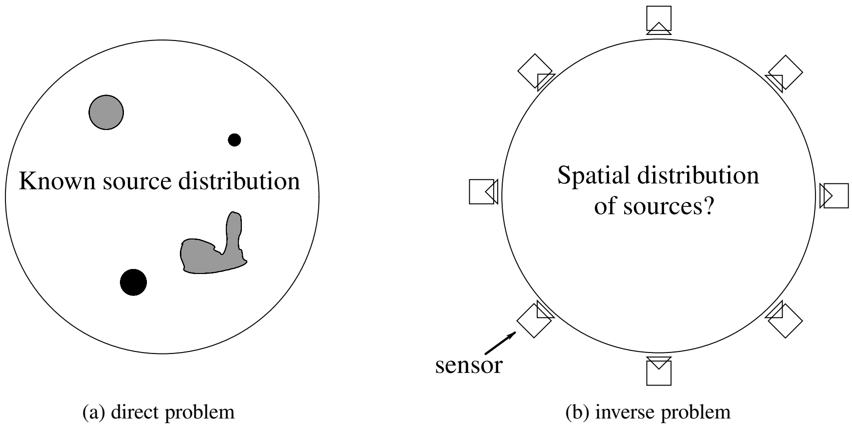


Figure 1. *A simple example of the direct problem/inverse problem pair in electromagnetism: the direct problem consists of finding the field at the circular boundary of the domain from the distribution of the sources; the inverse problem consists of deducing the distribution of the sources from measurements of the field at the boundary*

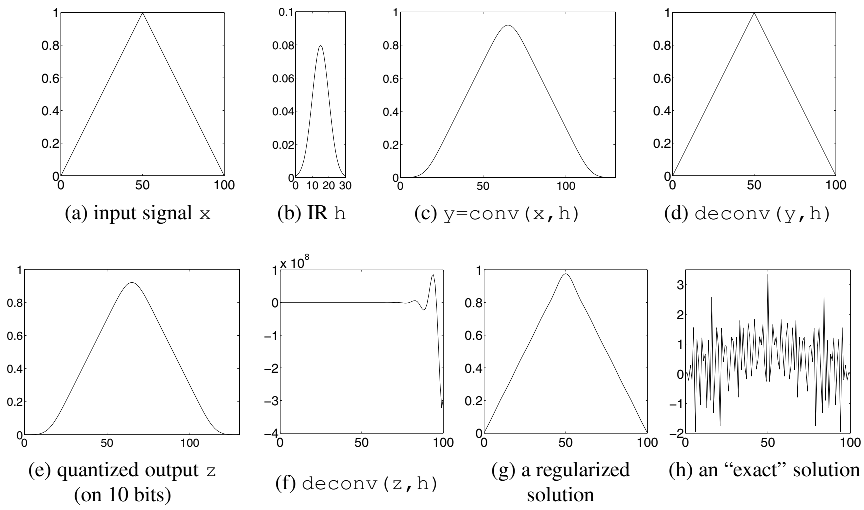
constraints, and provides a rational, modular framework for designing data processing methods and analyzing their efficiency and their faults.

When approached with no special precautions, the inversion problems we meet with in practice, unlike direct problems, have a nasty tendency to be "naturally unstable": if there are errors, however tiny, on the data, the behavior of "naive" inversion methods is not robust.

Let us take the example of *inverse filtering* (or *deconvolution*), a classic in signal processing. Figure 2 proposes two digital experiments:

– The first consists of inverting a discrete convolution relationship, $\boldsymbol{y} = \boldsymbol{h} \star \boldsymbol{x}$. The input signal $\boldsymbol{x}$ is triangular, of length $M = 101$ (Figure 2a), and the impulse response (IR) $\boldsymbol{h}$ is a discretized, truncated Gaussian of length $L = 31$ (Figure 2b). The output $\boldsymbol{y}$, of length $N = M + L - 1 = 131$, is calculated using the Matlab language in the form `y=conv(h,x)` (Figure 2c). Matlab also offers a deconvolution method (by polynomial division) that, here, faithfully gives $\boldsymbol{x}$ again from $\boldsymbol{y}$, in the form `deconv(y,h)` (Figure 2d).

– Now let us suppose that output $\boldsymbol{y}$ was measured imperfectly, e.g. with no measurement error but uniformly quantized to 10 bits, i.e., $1{,}024$ levels: $\boldsymbol{z}$ is the quantized output (Figure 2e), calculated in Matlab in the form `z=round(y*2^10)/2^10`. The difference between $\boldsymbol{y}$ and $\boldsymbol{z}$ is imperceptible; and yet, `deconv(z,h)` (Figure 2f) is very different from $\boldsymbol{x}$. Note in passing that the on-line help provided by Matlab (version 7.5) for `deconv` contains no warning of the unstable nature of this operation.



(a) input signal x (b) IR h (c) y=conv(x,h) (d) deconv(y,h)

(e) quantized output z (on 10 bits) (f) deconv(z,h) (g) a regularized solution (h) an "exact" solution

Figure 2. *An example of a "naturally unstable" inverse problem: deconvolution. A tiny difference between* z *and* conv(x,h) *is enough to make* deconv(z,h) *very different from* x*. Like "*conv*", "*deconv*" is an instruction in Matlab, version 7.5; it uses a non-regularized deconvolution method which is thus not robust. In comparison,* (g) *corresponds to a robust solution, obtained by Tikhonov regularization. Finally,* (h) *is an "exact" solution in the same way as* x*, in the sense that it reproduces* z *exactly by quantized convolution*

In the early 20th century, Hadamard characterized these unstable problems mathematically, qualifying them as *ill posed*, in the sense that they did not lend themselves to being satisfactorily solved mathematically (and physically) [HAD 01]. Chapter 1 of this book develops the notion of ill posed problems and makes the non-robust behavior of "naive" inversion methods mathematically explicit.

In the 1960s, the Russian mathematician Tikhonov laid down the theoretical basis of modern inversion methods by introducing the concept of *regularized solutions* [TIK 63]. These solutions result from two ingredients being brought together. They are faithful to the data but this ingredient is not discriminating enough if the problem is ill posed. Among other solutions that are faithful to the data, they are the most regular, in a pragmatic sense that depends on the context. Tikhonov formalizes this trade-off between fidelity to the data and regularity by defining regularized solutions as those that minimize a composite criterion. He shows that the problem thus reformulated is *well-posed*. The principle of regularization in Tikhonov's sense is one of the main subjects of Chapter 2.

Figure 2g illustrates the use of a method, regularized in Tikhonov's sense, that reproduces input $\boldsymbol{x}$ very acceptably from imperfect data $\boldsymbol{z}$. The calculation of this type of solution is dealt with in Chapters 3 and 4. The only imperfection visible on Figure 2g concerns the point of the triangle, which is slightly blunt. This is a logical effect of the regularity imposed uniformly on a signal which is, in fact, locally irregular at its midpoint (its derivative is discontinuous). The case of signals or images that are globally regular but have localized irregularities is a very important one in practice. It is treated in the context of deconvolution in Chapter 5 for irregularities that are "bright spots" and in Chapter 6 for irregularities that are borders between homogeneous areas. The latter situation is precisely the one we have in Figure 2.

The problem of Figure 2 could also be solved by positioning straight lines through the minimization of a least squares criterion. A *parametric* approach of this kind, which regularizes the problem by *dimension control* (Chapter 2, section 2.1.1), can be considered as the oldest of the inversion methods as it was invented around the same time as the least squares method in the late 18th century. One of the first times the principle was put into practice was when Gauss estimated the coefficient of ellipticity of the Earth from arc length measurements, having modeled the Earth's profile in the form of an ellipse [STI 81]. This was indeed a case of an inverse problem being solved by a parametric approach, even though the concept actually appeared much later.

Finally, we could think that the inversion would naturally become stable if the direct problem could be modeled with no errors. If this were true, it is a more detailed description of the direct problem that would lead to stabilization of the inversion. Let us take the example of Figure 2, for which the exact mathematical relation linking $\boldsymbol{x}$ and $\boldsymbol{z}$ includes quantization: the inversion of this exact relation remains unstable. In fact, $\boldsymbol{x}$ is only one of an infinite number of solutions, some of which stay remarkably far from $\boldsymbol{x}$; Figure 2h is an example.

Figure 2 is instructive but simplistic. In a more realistic situation, attaining a description of a direct problem – including the measuring system – in a mathematically perfect form is, in any case, more than we can hope for. In the inversion field, it is widely accepted that a credible inverse method must possess a minimum of robustness with respect to imperfect modeling of the direct problem. Adding pseudo-random noise to simulated data is a way of testing this robustness. Testing inversion only with "exact" simulated data is sometimes called the *inverse crime*.

Regularization in the Tikhonov sense, dimensionality control, adoption of such or such a parametric model, etc., there is no universal way of stabilizing an ill-posed inversion problem. The regularity of the solution must be defined case by case, in a form that may therefore appear subjective. Due to this, the concept of regularization is sometimes criticized or misunderstood. In fact, regularization is part of an application-oriented approach: it is not a question of inverting abstract problems that can be characterized by an *input-output* equation, but of solving *real problems*, where there is always advantage to be drawn from a few general characteristics of the quantity we are interested in, which may have been neglected in the initial formulation.

The practical success of Tikhonov's regularized approach and its subsequent evolutions has demonstrated that the approach is well founded. It is now accepted that an ill-posed inverse problem cannot be satisfactorily solved without some prior information, and this prior information is often qualitative or partial. For example, in an image restoration problem, it is desirable to take into consideration the fact that an image is generally composed of *homogeneous* regions, but this characteristic is qualitative and does not directly correspond to a mathematical model.

The encoding of uncertain or partial information can be envisaged within a *Bayesian probabilistic* framework. Work published a considerable time ago [FRA 70] showed that Tikhonov's contribution could be interpreted in this framework. For direct problems formulated deterministically, the handling of probabilistic rules for the inversion gives rise to comprehension difficulties. It has to be understood that these probabilistic rules are inference rules: they enable states of knowledge to be quantified and their evolution, through measurements, is itself uncertain because of errors. So we are not judging the fundamentally deterministic or random nature of the observed phenomena or even of the measurement errors. In this respect, Jaynes' work, brought together in [JAY 03], provides a reference for understanding the Bayesian approach in the data processing field. Chapter 3 of this book takes its inspiration from this work.

Modern methods for solving inverse problems have been arousing increasing interest since their beginnings in the 1960s. The question of inversion concerns a variety of sectors of activity, such as Earth and space sciences, meteorology, the aerospace industry, medical imaging, the nuclear electricity industry and civil engineering. Added up over all these sectors, its scientific and economic impact is enormous.

As far as the structure of inverse problems is concerned, very different fields may have very similar needs. However, the compartmentalization of scientific disciplines makes it difficult for ideas and methods to circulate. In this respect, it is the role of the signal processing community to respond to needs common to other disciplines in terms of data processing methods and algorithms. It was with this in mind that this book was written. Its 14 chapters are grouped together in four parts.

Part I is devoted to introducing the problems and the basic inversion tools and is the most abstract. It comprises three chapters:

– Chapter 1 introduces the problem of inversion as a whole, in a structured mathematical framework. It gives the characteristics of inverse problems posed in a continuous or discrete framework, and of ill-posed problems. It introduces the ideas of the pseudo-solutions and generalized inverse;

– Chapter 2 introduces the essential characteristics of regularization theory, and describes Tikhonov's approach and its subsequent evolutions together with other approaches to regularization. Reminders are then given of methods for minimizing criteria, followed by techniques for estimating the regularization parameter, so that an automatic choice can be made for the trade-off between fidelity to data and regularity;

– Chapter 3 deals with solving inverse problems in the framework of statistical inference. It becomes apparent that the conventional estimation technique, known as *maximum likelihood*, corresponds to a non-regularized solution, whereas a Tikhonov regularized approach finds a natural interpretation in the framework of Bayesian estimation. A number of questions are then re-examined in this context: the automatic choice of parameters, and the building of models and criteria. The end of the chapter is devoted to the Gaussian linear framework, which constitutes a fundamental special case.

Part II is made up of Chapters 4, 5 and 6, and is entirely given over to deconvolution, as a case that is very widespread in practice and also as a very instructive case where many of the tools introduced in these chapters can be adapted to the inversion of problems that are structured differently:

– Chapter 4 deals with deconvolution methods yielding solutions that are linear functions of the data. It first studies the general properties of the solutions, then the various algorithm structures that enable them to be calculated. Traditional signal processing tools, such as Wiener and Kalman filters, figure among these structures;

– Chapter 5 looks at the more specific problem of deconvolution when the signal of interest is a series of pulses. This situation is very common in numerous domains such as non-destructive evaluation and medical imaging. Taking the pulse character of the input signal into account leads us to two classes of nonlinear solutions according to the data. One follows a *detection-estimation* approach and the other uses the minimization of convex criteria and *robust estimation*;

– Chapter 6 is devoted to deconvolution when the unknown signal is "regular almost everywhere" and, in particular, takes this characteristic into consideration for images rather than monovariate signals. As in the previous chapter, we find two classes of solutions, according to whether the problem is approached in terms of detection of edges or as a problem of image restoration in robust form.

Part III groups together two chapters introducing "advanced tools" specific to the Bayesian framework presented in Chapter 3:

– Chapter 7 is concerned with imaging from a probabilistic point of view. The composite criteria of Chapter 6 are reinterpreted in this framework, which leads us to the Gibbs-Markov models. Various sub-classes are introduced as models for images. We next look at the statistical aspects connected with calculating the estimators and evaluating their performance. Finally, the principle of iterative methods for random sampling of the Gibbs-Markov models is presented;

– Chapter 8 is entirely devoted to the problem of *non-supervised* inversion, i.e., inversion without a regularization parameter fixed by the user. This question, already mentioned in Chapters 2 and 3, is of considerable theoretical and practical interest but brings together several types of methodological and algorithmic difficulties. Chapter 8 studies the case of a linear Markov penalization function with respect to the parameters to be estimated and proposes, in particular, deterministic or stochastic techniques for implementing *maximum likelihood*, exact or approximate estimators.

Part IV, in six chapters, presents some inverse problems in their applications. This is by no means a complete review of all the domains involving inversion; several important fields such as heat, mechanics and geophysics are not covered[1]. For the applications that are mentioned, we do not give an overall synthesis of the inversion problems encountered but rather some typical examples chosen by the authors as concrete illustrations of how regularized solutions are implemented in a particular domain. This last part also contains some important methodology extensions – myopic deconvolution (Chapters 9 and 10), Fourier synthesis (Chapter 10), spectral estimation and handling of *hidden Markov chains* (Chapter 11), and tomography problem solving (Chapters 12, 13 and 14):

– Chapter 9 concerns *industrial non-destructive evaluation using ultrasound*. It compares the implementation of the spike train deconvolution methods presented in Chapter 5. The problem of an impulse response that is poorly known or that introduces deformation is specifically studied. The result of this is some extended versions of the algorithms looked at in Chapter 5;

– Chapter 10 is about the inversion problems encountered in optical imaging in astronomy and, more specifically, for ground-based telescopes. In this case, atmospheric turbulence considerably reduces the resolution of the images. Various configurations intended to limit this degradation are considered. On the one hand, it is possible to approach the image restoration problem through myopic deconvolution; on the other, the effects of turbulence can be partially compensated by a technique known as *adaptive optics*, where the deconvolution of the images thus acquired remains a helpful step. The end of Chapter 10 is devoted to optical interferometry, which leads to a Fourier synthesis problem complicated by aberration due to atmospheric turbulence;

– Chapter 11 is devoted to *Doppler ultrasound velocimetry*, an imaging technique that is widespread in medicine. Two data inversion problems are seen to arise: *time-frequency analysis* and *frequency tracking*. These are *spectral characterization* problems that are particularly difficult for two reasons: firstly, the number of observed data points is very small and, secondly, the Shannon sampling conditions are not always respected. Chapter 11 covers these two problems in the regularization framework in order to compensate, at least partially, for the lack of information in the data;

– Chapter 12 considers the problem of reconstruction in X-ray tomography using a small number of projections. The Radon transform is introduced, and after a brief reminder of the various conventional approaches for its inversion, the algebraic and probabilistic methods are developed more specifically. In fact these are the only methods that can be used effectively in cases where the projections are limited in number and contain noise;

---

1. See [BEC 85, BUI 94, DES 90] as respective entry points to these fields.

– Chapter 13 looks at how the Bayesian approach can be used to solve the problem of diffraction tomography. For this type of problem, the measurements collected are the waves scattered by an object and depend nonlinearly on the physical parameters we are trying to image. This chapter deals with diffraction tomography without the usual linear approximations, whose domains of validity do not cover all the situations encountered in practice. The wave propagation equations provide a integral direct model in the form of two coupled equations. These are discretized by the *method of moments*. The Bayesian approach then allows the inversion to be approached by minimizing a penalized criterion. However, the nonlinearity of the direct model leads to non-convexity of this criterion. In particularly difficult situations where local minima exist, the use of global optimization techniques is recommended;

– Chapter 14 studies, in the framework of medical imaging techniques such as positron emission tomography, the situations in which the corpuscular character of the counting measurements needs to be taken into account. Poisson's law serves as the reference statistical distribution here to define the likelihood of the observations. There are also composite cases, in which Gaussian noise is added to the Poisson variables. It is sometimes possible to approximate the Poisson likelihood by a Gaussian law. When this is not the case, the properties of the log-likelihood criterion are studied and algorithms are put forward for various situations: emission or transmission tomography and composite cases.

Bibliography

[BEC 85] BECK J., BLACKWELL B., ST. CLAIR C., *Inverse Heat Conduction, Ill-posed Problems*, Wiley Interscience, New York, NY, 1985.

[BUI 94] BUI H. D., *Inverse Problems in the Mechanics of Materials: An Introduction*, CRC Press, Boca Raton, FL, 1994.

[DES 90] DESAUBIES Y., TARANTOLA A., ZINN-JUSTIN J., Eds., *Oceanographic and Geophysical Tomography*, Amsterdam, The Netherlands, Elsevier Science Publishers (North-Holland), 1990.

[FRA 70] FRANKLIN J. N., "Well-posed stochastic extensions of ill-posed linear problems", *J. Math. Anal. Appl.*, vol. 31, p. 682-716, 1970.

[HAD 01] HADAMARD J., "Sur les problèmes aux dérivées partielles et leur signification physique", *Princeton University Bull.*, vol. 13, 1901.

[JAY 03] JAYNES E. T., *Probability Theory – The Logic of Science*, Cambridge University Press, Apr. 2003.

[STI 81] STIGLER S. M., "Gauss and the invention of least squares", *Annals Statist.*, vol. 9, num. 3, p. 465-474, 1981.

[TIK 63] TIKHONOV A., "Regularization of incorrectly posed problems", *Soviet. Math. Dokl.*, vol. 4, p. 1624-1627, 1963.