Are Inverse Problems Ill-Posed?

Roel Snieder and John A. Scales

Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines, Golden, CO 80401, USA

(Submitted to the special issue of Inverse Problems
to celebrate the 65th anniversary of Pierre Sabatier)

September 26, 2000

“Really new ideas on inverse problems most times come from
the collision between the needs and accomplishments of people
working in different areas”. [Sabatier, 2000]

1 Introduction

Inverse theory concerns the problem of making inferences about physical sys-
tems from measurements. There are two kinds of measurements, direct and
indirect. In a direct measurement you measure the property of the phys-
ical system that you are trying to estimate. For example, if you want to
estimate the mass of an object you can repeatedly put it on a balance and
take the mean of the measurements. This would be a direct measurement
since a balance measures mass. However, if you put the mass on a conven-
tional scale which actually measures the displacement of a spring, then the
measurement becomes indirect and you must take into account the forward
problem (Hooke’s law), systematic errors (nonlinearity in the spring) as well
as random errors in the data, in order to be able to estimate the mass of the
sample. It is entirely a matter of definition, but it seems reasonable to restrict the scope of inverse problems to inferences from indirect measurements. Since most measurements are in fact indirect, inverse theory is central to all empirical sciences. We don’t know what the first inverse problem was, but certainly by the middle of the 18th century people were “inverting” noisy data to make geophysical and astronomical inferences. For example, in the middle of the 18th century, the Croatian Jesuit Roger Boscovic invented the least absolute deviation criterion to reduce geodetic measurements in order to estimate the figure of the earth. This was even before Gauss had developed the least-squares method for his astronomical work.

Traditionally, people working on inverse problems have approached the subject from one of two distinctly different perspectives. The first perspective relates to the problem of inverting operators. This includes the formal problem of identifying the coefficients of a differential equation from hypothetical “data.” Here the operator maps from a space of parameter functions (the system “models”) to a space of data functions. In an ideal world of perfect data (infinite bandwidth and continuously sampled) these operators may be formally invertible. However, in practice such problems are typically ill-posed [e.g. Dorren and Snieder, 1995] so some generalized criterion for solution is constructed. This criterion is usually based on the optimization of some regularized data misfit function. The ultimate goal then is to construct a stable, uniquely solvable optimization problem, the extremum of which is the solution. A posteriori error analysis becomes an appraisal of this solution. The tool used in the study of operator inversion come primarily from functional analysis, differential equations and optimization.

The second major perspective comes from people who begin the study of the inverse problem with real data, which are always finite in number and subject to a variety of random and non-random uncertainty. Here the overarching principle is the extraction of information from data. Or, more precisely, the goal is to make quantitative statements about the features of the physical system under study that are consistent with the measurements and whatever other data-independent information is available: this is the inference problem [e.g., Lehman, 1983]. Since there can only be a finite number of data and since physical systems are usually characterized by continuous functions, the notion of operator inversion becomes secondary: the operators mapping models to data are never invertible and regularizing this mapping only camouflages the fact there is a null-space of models that have no influence on
the data. The presence of a null-space or kernel in the data mapping implies fundamentally that it is impossible to pin down the true parameters of the physical system with finite uncertainty unless you are willing to incorporate data-independent information into the problem. Not surprisingly, the basic tools used in the data analysis perspective come primarily from statistics.

Mathematicians and theoretical physicists usually approach inverse problems from the operator inversion perspective. For example, a recent survey of inverse theory [Colton et al., 2000] dates the field only to the mid-1960s, since this is date of the papers of Tikhonov [1998], thus ignoring a rich heritage in statistical inference going back to Levenberg [1944], Jeffreys [1931] and before.

Examples of inverse problems seen as the inversion of an operator include the determination of the mass density of a spring from its eigen-frequencies [Borg, 1946], inverse problems in quantum scattering theory [e.g. Marchenko, 1955; Chadan and Sabatier, 1989; Newton, 1989], the determination of the shape of the vocal tract Sondhi and Gopinath, [1971], inferring the structure of a radially symmetric acoustic [Sabatier, 1974a] or elastic [Sabatier, 1974b] sphere from scattering data, as well as other inverse problems in wave propagation that are based on inverse scattering techniques [Burreidge, 1980] or the Generalized Random Transform [e.g. Beylkin, 1985]. Research in this field has sometimes led to unexpected results in other fields; an example is the inverse scattering transform that plays an important role in the solution of nonlinear partial differential equations [e.g. Ablowitz and Clarkson, 1991]. These advances have proven to be of crucial importance in practical applications that include geophysics, medical imaging, non-destructive testing, ground water modeling, helioseismology and network design.

This type of research also helps us to understand the limits of what aspects of the system can be reconstructed from a given data set. For example, when bound states are present, the inverse problem for the Schrödinger does not have a unique solution [e.g. Chadan and Sabatier, 1989]; a spherically symmetric potential that decreases too rapidly with decreasing radius cannot be reconstructed uniquely from scattering data in the WKB approximation [Sabatier, 1973]; and the seismic velocity in the Earth cannot be retrieved uniquely from travel time data when a low-velocity layer is present [Gerber and Markushevitch, 1966]. In this way applied mathematical research has played an important role in the solution of inverse problem.

Statisticians, on the other hand, have made fundamental advances in the
analysis of optimal inversion algorithms in the presence of uncertain information. For example, given a linear forward operator, a linear functional and a quadratic constraint on the set of plausible models (e.g., a bound on the norm of the models), there are ways to determine how well we may expect to do with the best estimator (inversion algorithm), as well as concrete algorithms for the optimal estimate of uncertainty in the parameters (confidence sets, etc) [e.g., Donoho et al, 1990; Donoho, 1990; Stark, 1992]. In particular Donoho [1994] has shown that there is a close connection between minimax statistical estimation and optimal deterministic recovery of a linear functional of a model known to be in a convex subset of the space of models, from linear observations with Gaussian errors. A modern treatment of the general problem of inference of linear functionals in the presence of random and systematic noise can be found in book of Plaskota [1996].

Figure 1: The conventional view on inverse problems: find the model that predicts the measurements.

2 Inversion made (too) simple

Figure 1 shows the view on inverse problems as is it presented in many texts. The model that one aims to determine from data is an element of a mathematical space that contains all allowable parameterizations of the model properties (or at least those properties relevant to a given experiment); this space is referred to as model space. The physics of the problem determines which data $d$ correspond to a given model $m$. The problem of computing the model response (synthetic “data”) given a model is called the forward problem. The corresponding data reside in a mathematical space that is called data space. In many application one records the data, and the goal is to find
the corresponding model. The task is called the *inverse problem* as shown in figure 1.

Unfortunately, figure 1 is not accurate for practical inverse problems. There is a simple reason for this. In many applications the model that one seeks is a continuous function of the space variables with infinitely many degrees of freedom. For example, the three-dimensional velocity structure in the Earth has infinitely many degrees of freedom. On the other hand, the data space is always of finite dimension because any real experiment can only result in a finite number of measurements. A simple count of variables shows that the mapping from the data to a model cannot be unique; or equivalently, there must be elements of the model space that have no influence on the data. This lack of uniqueness is apparent even for problems involving idealized, noise-free measurements. The problem only becomes worse when the uncertainties of real measurements are taken into account. Although the uniqueness question is a hotly debated issue in the mathematical literature on inverse problems, it is largely irrelevant for practical inverse problems because they invariably have a non-unique solution (if by *solution* we mean an specific model that we reconstruct from the data). It is this non-uniqueness that makes figure 1 deceptive, because the arrow pointing from data space to model space suggests that a unique model corresponds to every data set.

![Diagram](image)

**Figure 2:** An improved view on inverse problems.
3 Inversion with real data is complex

A more realistic scheme of inverse problems is shown in figure 2. Given a model \( m \), the physics of the problem determines the predicted data \( d \); this is called the forward problem. For a given data set, one determines an estimated model \( \hat{m} \). We refer to this as the model estimation problem. (Later we will consider the generalization to the problem of estimating properties of models, rather than the models themselves.) Note that there may be many reasonable model estimates for a given data set and that the estimation procedure may be nonlinear even when the forward problem is linear. Thus the mean of a set of numbers is a linear function of the numbers, while the median is a nonlinear function. Yet both the median and the mean may be reasonable estimators of the “center” of the set of numbers. Part of the art of solving inverse problems comes from the need to define what it means for an estimate to be reasonable.

Since the mapping from data space to model space is non-unique, the estimated model may also depend on the details of the algorithm that one has used for the estimation problem as well as on the regularization and model parameterization that has been used. In general, the estimated model \( \hat{m} \) differs from the true model \( m \). For example, in seismic inversion the estimated model may be a blurred version of the true model. In addition, the data are always contaminated with errors; these errors represent an additional source of discrepancy between the estimated model and the true model.

One is not finished when the estimated model is constructed; it is essential to somehow quantify the error between the estimated model and the true model. This is called the appraisal problem. In this problem one determines the uncertainty in the estimated model. This uncertainty has a statistical component related to the propagation of errors in the data, and a nonstatistical component that accounts for the finite resolution that is attained in the model estimation [e.g. Snieder and Trampert, 1999].

For linear inverse problems, resolution kernels and confidence set analysis are powerful tools for formulating the appraisal problems, and the theory of linear error propagation is sufficiently well-developed to account for the errors in the estimated model due to the errors in the data. For nonlinear inverse problem the only tool available for the appraisal problem (at least for Bayesian inverse problems) may be Monte Carlo methods, where one estimates the statistical properties of the model by assimilating data
and other information via repeated sampling of the model space [e.g. Sambridge and Drijkoningen, 1992; Lomax and Snieder, 1994; Mosegaard and Tarantola, 1995; Gouveia and Scales, 1998]. These techniques are only applicable to practical problems where the number of parameters is relatively small. Furthermore, the interpretation of large ensemble of models with associated uncertainties is nontrivial [e.g. Douma et al., 1996]. For large-scale inverse problems such as the determination of 3D Earth structure, Monte Carlo methods are not computationally feasible. This means that there is presently no operational theory to account for the appraisal problem of nonlinear inverse problems with large number of parameters [Snieder, 1998]. Developing such a theory is a theoretical and practical challenge that is much more important than establishing uniqueness proofs of idealized mathematical problems.

4 Defining the goal of inverse problems

In practice, one solves inverse problems with a certain goal. For example, one may use an estimated model obtained from an inversion of seismic data as a basis for deciding where to drill or how to optimally exploit a reservoir. In practice, one is never interested in the seismic velocity at a certain spot in the subsurface (velocity being a typical model parameter used in the forward problem for seismic waves), but for a seismic interpreter it is crucial to know whether at a certain place a syncline or an anticline is present. This means that for practical inverse problems one is interested in patterns that can be used in a meaningful way for making decisions. Similarly, in medical imaging a doctor will base her decisions for treatment on the pattern that she sees (e.g. is there a tumor?) rather than on the properties of the image at one particular location in the patient’s body. In practice, these decisions are usually not based exclusively on the estimated model \( \hat{m} \), but involve the integration of other data as well as human expertise; the accumulated experience of a skilled seismic interpreter or a radiologist is crucial in using the obtained model effectively. In addition, the uncertainty in the estimated model is an important factor in making decisions. This means that figure 3 gives a more realistic view on inverse problems, because it shows explicitly that decisions rather than model estimation is the endpoint of practical inverse problems.

It is interesting to consider the relation between the appraisal problem
and the process of decision making. An important aspect of the appraisal problem is the statistical treatment of error propagation. This means that the appraisal problem has probability as an important component. Decisions are usually based on risk rather than probability. This may be economic risk (where to drill?), environmental risk (what happens when pollutants spread?) or even academic risk (am I sure enough to publish my results?). Risk is always concerned with probability plus another component such as profit or environmental impact. This means that in this stage those who produce the models and their uncertainty must interface with others for making decisions effectively.

Figure 3 offers an overview of the landscape of geophysical inversion. Those working on wave propagation problems focus on the forward problem. Those focused on the development of migration algorithms work on the estimation problem. Statisticians and a limited number of scientists in the inverse problem community are concerned with the appraisal problem. Seismic interpretation is usually based on an estimated model, this activity is an important aspect of the decision making problem and it is an interesting exercise to relate the range of activities in other fields of inverse problems as well to figure 3. It is illuminating to see how different researchers work on
different parts of this problem. One may wonder whether our research efforts could be more effective when their activities are seen in the context of the anatomy of the inverse problem as shown in figure 3.

![Diagram of the inverse problem](image)

Figure 4: The inverse problem seen as an inference problem.

5 Do we need a model?

In figure 3 the estimated model $\hat{m}$ forms an essential part of the inversion process. But is it necessary or desirable to produce an estimated model in the process of inversion of data? We are of course conditioned to produce models from our data. However, given the fact that this estimated model differs from the true model one can be led astray by features in the estimated model that are artifacts of the inversion process. Another view on inverse problems is given in figure 4 where the goal is to determine the range of models that are consistent with the data (as well as other information). This range of models can simply be a box within which all the models are believed to fit the data (there being no comparative relation among the models in the box), or it could be a probability distribution on the space of models $P(m)$ (in which case one can speak of the best or most probable model). Both the box and the probability are determined by the data, the uncertainties and whatever other data-independent information is available. People taking
the former approach are called frequentists (this includes most statisticians) while people taking the latter approach are called Bayesians. (Some pitfalls in the use of Bayes’ theorem are discussed by Scales and Snieder [1997] and Scales and Tenorio [2000].) In either case, the estimation and appraisal problem is replaced by the inference problem. Inference means characterizing somehow the set of models that explain the data (and satisfy whatever other information is available). One can then use this box, or the probability distribution \( P(m) \), as a basis for making decisions.

The reader may be put off by the idea of solving inverse problems without constructing models. Our minds are conditioned to making models from data. However, we have seen that these estimated models can only be the endpoint of research when one ignores the fact that models are being produced with the goal of making decisions. When seen in this larger context (figures 3 or 4) it is worth considering whether one might better served by knowing the probability distribution of the set of models than by knowing a single estimated model and a measure of its uncertainty.

It is interesting to consider how we would use a probability density function in a high-dimensional model space. The simplest way to use such a function is to compute the mean and variance of each model parameter, and one can visualize this information relatively easy. However, as we have seen the patterns in the model are much more interesting than than the estimates of individual model parameters. Assessing the robustness of certain patterns in the model is much more difficult, especially since this entails the use of the correlation between different model parameters. In a high-dimensional model space it is extremely difficult to characterize and interpret the correlations of the model parameters. In order to interrogate the resulting model-space probability density function in a meaningful way research is needed in the cluster and feature analysis of (possibly multi-modal) probability density functions of high-dimensional problems and in the development of an interface between these probability density functions and decision-making procedures.

It will be clear from this that in order to treat inverse problems in ways that are different from current practice requires significant theoretical and numerical advances. However, the optimal use of inverse problem methods in the process of decision making should be an important factor in the research agenda of practitioners of inverse methods. This requires an ongoing collaboration between researchers in both theoretical and applied aspects of inverse
problems. Although the collaboration between researchers from different communities is often difficult [Snieder, 2000], it can be rewarding because as stated in the quote of Sabatier [2000] it is often at the interface between different scientific communities where new developments are spawned.

References


