A Geophysical Inverse Theory Primer

Andrew A. Ganse

Applied Physics Laboratory and Department of Earth and Space Sciences

University of Washington, Seattle, WA

`http://staff.washington.edu/aganse`

`aganse@apl.washington.edu`

March 31, 2008

Abstract

This document is ten pages long, has no equations, and aims to introduce the underlying concepts of inverse theory and place them in perspective. In giving inverse theory lectures and labs in the past, I found that students often became so caught up in keeping up with mathematical details that they did not entirely grasp what they were calculating. The goal of this document is to help with that, laying out the broad ideas before delving into any mathematical details.

A prerequisite for delving further into a geophysical inverse theory class or textbook is a familiarity with linear algebra (rules for doing matrix and vector arithmetic, definitions of matrix inverse, transpose, underdetermined and overdetermined problems, rank, condition number, eigenvalues, and eigenvectors) and basic statistics (what are mean, standard deviation, variance, covariance matrix, and correlation matrix).

Introduction

Inverse theory, or "inversion", is a mathematical tool for interpreting indirect scientific measurements. It is used in many different fields of science; let us mention some examples to illustrate the idea. In seismology one often wants to find out what the structure and composition is deep inside the Earth, but we cannot go there to take measurements, so instead we indirectly figure out what we can about the Earth's interior from seismometer recordings taken at the Earth's surface. As a similar example, in medical imaging one wishes to know about the inside of the body without cutting it open. A common imaging technique is the computed tomography (CT) scan, which puts together many Xray measurements through the body to show 3D structures of the organs and bones. Another example: one wishes to know something about the interior of a moon of Jupiter by flying a spacecraft past it. From the subtle, anomalous tugs in the spacecraft's orbital trajectory one can indirectly figure out something about the density distribution inside that moon which caused the tugs in the trajectory. These examples so far all have the frequently-seen theme of indirectly measuring interior properties from the outside, but there are other variations. For example, in borehole seismology one studies near-surface properties of the Earth by drilling a deep borehole, placing seismometers at various locations deep down inside it, and setting off a charge at the surface so that the Earth properties as a function of depth in that area can be worked out. Hydrologists can similarly place contaminant sensors down in deep holes in a region around a chemical factory and from them indirectly work out the source history of the contamination coming from the factory. In any case, the common elements of all these examples are that: there is a quantity we wish to know but cannot measure directly, there is some other set of measurements made, and there is a mathematical relationship that links the two so we can map one to the other. The trouble is, the mapping generally is not perfect and there is noise on the measurements, and because of these we need to quantify the uncertainty and limited resolution of the desired quantity we wish to estimate. Inverse theory quantifies that mapping and the specifying of the uncertainty and limited resolution.

Some Important Terminology

Inversion is a type of statistical estimation, but "there is estimation and then there is estimation." Just to get some terminology straight, in the next several paragraphs let us distinguish between some related but different computations. First let us first clarify a few preliminary terms. It was mentioned previously that in inverse problems there is a mathematical relationship that links the measured quantities – let us call these the **data** – with the quantities one really wants to know – we will call these the **model**. Note that folks in other, non-geophysical communities often wish to call the mathematical relationship the "model", and get confused by this term used for the quantity being estimated. It is purely a vocabulary issue; it comes from geophysicists estimating a "model of the Earth".

Just because it always trips everybody up, let us say it one more time. In geophysical inverse theory classes and textbooks, the **model** is the term for the quantity being estimated, *not* the mathematical relationship. The mathematical relationship is called the **forward problem**. They are just words, but they are the words that the geophysicists use.

In any case, the forward problem is of the form $data = somefunction(model)$, so that if we already have the model and wish to compute the data everything is fine. But often we have a mathematical relationship of that form, and instead wish to know the model that corresponds to data that we have, so we really want a relationship of the form $model = somefunction^{-1}(data)$. Since that is the inverse of the forward problem we call that the **inverse problem**. Unfortunately, with the mathematical relationships often encountered in physics and scientific imaging, finding that inverse function can be a really tall order.

There are two main communities that focus a lot on inverse theory, and ironically they rarely talk to each other because their approaches are so different. First there is the pure-math community, which takes an analytical, just-solve-the-equation (not that it is easy) approach to finding the $model = somefunction^{-1}(data)$ form. Since pure mathematicians are not necessarily focused on problems that have application in the sciences, the forward problems that they invert this way are frequently very idealized and thus often not directly useful for scientific work. There are some notable exceptions however. For example, computed tomography (CT) scans use this approach via a mathematical relationship called the **inverse Radon transform**. Laboratory nuclear scattering studies use analytically-based **inverse scattering theory**. In both cases, these problems were solved for the function $somefunction^{-1}()$ assuming continuous (i.e. perfect geometric coverage) data without noise. Then the real world application works if we plug data points into those problems as long as their noise is very low, there are a lot of data, and their geometric coverage is very good. In the 1980s there was a flurry of journal papers by geophysicists (especially in the field of oil exploration seismology) who desperately tried to work out analytical inversion approaches like this for seismic problems, and after enough years they finally concluded that seismic inverse problems, like many other geophysical problems, were just too **unstable** for such analytical inverse solutions. We will define what "unstable" means in just a moment.

The second community that focuses a lot on inverse theory, then, is geophysicists – particularly seismologists, but gravimetry and magnetotellurics and glaciology researchers too. The geophysicists worked out a much more general, optimization-based approach to inverse problems, which may not always give as much information as an analytical solution would, but since stable analytical solutions to geophysical inverse problems generally do not exist (or at least have not been found yet) it is a vastly better tool than nothing. It is this general approach of the geophysicists which we learn in **geophysical inverse theory** classes.

Here now we are always talking about problems that will be done on a computer, so that the data is always discrete, not continuous – the data is always a finite number of data points in a computer file somewhere. From those data points the model is estimated, and now let us further clarify how the estimations of different classes of models have different names.

Parameter estimation is the estimation of a finite number of discrete model parameters from a set of data; it might be only a couple parameters, or thousands of parameters. Often parameter estimation problems are designed to have fewer parameters than data points so that the problem is overdetermined, and thus really straightforward to solve. **Inversion** is the estimation of a continuous model function (be it 1D, 2D, 3D, or even multiple continuous functions) from a set of data points. If we think of the continuous function as an infinitely close together, infinite string of model parameters, then by comparison to the finite number of data points we see that inverse problems are *always* underdetermined by their very nature. This feature of inverse problems makes for limits on how fine a resolution the continuous function can be estimated at. Inverse problems have a lot more complexity than parameter estimation problems because we have to work out the balance between the limited resolution of the continuous function and the goodness of fit to the noisy data in a process called **regularization**. Ironically, part of solving a geophysical inverse problem is parameterizing the continuous function with a finite set of model parameters, turning part of the problem into a parameter estimation problem. But there is more to it than that, via that misfit/resolution business, and at the end the model parameters are converted back to a continuous model function again. Lastly, **filtering** is a special case of inversion, in which the continuous function is in a category called a *Gauss-Markov process*, and we would like to keep estimating the function further as more data points come in. Filtering (and a closely related estimation process called **smoothing**) is frequently used in vehicle dynamics problems like satellite tracking, where the Gauss-Markov process is in the form of gravitational and inertial dynamics, the continuous function is the satellite trajectory over time, and the data points are radar range and velocity measurements coming in over time. One final distinction will not be defined until the next section, but parameter estimation and inverse problems can be done in either the **frequentist** or **Bayesian** statistical frameworks, while filtering and smoothing problems are only defined in the Bayesian framework.

It may be useful to note a common misnomer: people often call parameter estimation problems inverse problems, often enough to mention it here so that we know to translate it inside our heads. Again just vocabulary really, as long as one understands that while geophysical inverse problems include parameter estimation in their solution process, they also require the extra regularization machinery for the misfit/resolution issues, while plain-old parameter estimation problems do not and are thus simpler than inverse problems.

Frequentist vs. Bayesian Frameworks for Inversion

It turns out the statistical world has two camps, so the geophysicists who use statistics (such as in inverse theory) tend to fall into two camps as well. For a few hundred years now this theoretical rift in statistics has raged and produced endless arguments as to whether the **frequentist** or the **Bayesian** camp is more correct in its statistical philosophy. The conclusion is likely not for geophysicists to work out, and perhaps it never will be worked out at all; in the meantime it is useful to be familiar with both.

Both statistical frameworks are rigorous and legitimate and have a long history, and both approaches to geophysical inversion usually involve transforming the estimation of a continuous function into an estimation of model parameters which represent that function. However, these two approaches define probability differently, which affects their respective formulations and results of inverse problems.

Frequentists define probability in the sense of frequency of repeatable events, so they do not allow for the notion of prior knowledge about the model parameters before an inversion takes place (i.e. before an event which provides measured data). This constraint causes one to only be able to estimate weighted averages of model parameters rather than the individual model parameters themselves in frequentist inversion. This is a result of the fact that without prior information about the model, the regularization (the process that handles the inherently underdetermined nature of inverse problems) messes up the estimator such that it can only give those weighted averages. But frequentists prefer this limitation over accepting prior information. **Bayesians**, on the other hand, define probability in the sense of degree of belief, and so do allow for prior knowledge. In Bayesian inverse problems, prior probabilities of the model parameters are updated using information from the measured data to give posterior probabilities of the model parameters. The Bayesian inverse problem estimates probabilities of the parameters themselves, rather than of the weighted averages, so one must be careful when comparing frequentist and Bayesian inversion results for the same problem.

Aside from the centuries-long argument over whether the frequentist or Bayesian definition of probability is better, it might be argued that the frequentist formulation is more appropriate for problems in which one knows little or nothing about the model parameters before the inversion, and the Bayesian formulation is more appropriate for problems in which one can specify prior probabilistic information about the model parameters (say from a supplemental set of direct measurements).

Uncertainty and Resolution

In inverse problems, the **noise** on the data gets mapped through the problem to become **uncertainty** on the model being estimated. Often both the noise and the model uncertainty are deemed Gaussian (or close enough), so that the vector of measured data points has a measurement **covariance matrix** associated with it, and the vector of resultant model parameters also has a covariance matrix. In the Bayesian framework, that is "all" there is too it, and analyzing the limited resolution of the model estimate consists of studying the improvement from the prior to the posterior covariance matrices. In the frequentist framework one still computes a covariance matrix for the model parameters estimated, but unlike the Bayesian framework, here the estimated parameters are unfortunately weighted averages of the true ones. So in addition to correlations between those weighted averages as seen in the covariance matrix, there is another matrix calculated to give the weights, called the **resolution matrix** – it is via this matrix that we analyze the limited resolution of the frequentist model estimate. The name of this matrix comes from the fact that the estimated parameters are generally weighted averages of neighboring parameters, limiting the ability to resolve them individually.

It was mentioned earlier that many geophysical inverse problems are too **unstable** for analytical inverse methods, without discussing what being stable meant. Many of the forward problems we will come across in geophysics are of a **convolution** form, such that the model is in some sense getting smoothed out by the forward problem, and large variations in the model become small variations in the data. When we want the inverse, this means trouble because small variations on the data (like noise for example) map to large variations in the model that we are trying to find, so that the solution we find for the model is very sensitive to the data noise. The solution keeps jumping around wildly for tiny fluctuations in the noise and we cannot pin it down.

To mitigate this instability so that we can find a solution, we must either add prior information (Bayesian approach) or mess up the estimator such that we can only find weighted averages of the model parameters (frequentist approach). In both cases that process is called regularization, and in both cases the regularization limits the resolution at which we can estimate the continuous model function. The level of regularization changes the weighting among the parameters in the frequentist framework, and changes the relationship between the prior and posterior probabilities in the Bayesian framework.

Quick Descriptions of Common Techniques

You will likely only concentrate on one or two of these in a given inverse theory class, while glossing over the others. This brief summary list may help keep them straight and distinguish them for you, and act as a guide for looking up other techniques in the future.

Frequentist Techniques

- **Backus and Gilbert method:** In the late 1960s, some seminal papers by these guys were a major leap for frequentist seismic inversion. They laid out a process that addressed the tradeoff between the variance of the estimated model and its limited resolution. Their specific method in those papers is not used as widely anymore however as it produces an estimate for only one x point in the continuous model $m(x)$ at a time, making it less computationally efficient than the other frequentist techniques below. But the B&G concept of an averaging kernel to explore the resolution effects are carried over to the below frequentist methods via the model resolution matrix, and the below methods really followed from these guys.
- **Gram matrix method:** This is the technique focused on in Robert Parker's book. In this approach the continuous model $m(x)$ is parameterized by the same number of parameters as there are data points, which can be a problem when there is a huge number of data points. Then a Gram matrix is developed by evaluating integrals that linearly relate the parameters and data points, the parameters are solved for, and used to create the continuous model function estimate. Another common complaint about this method is having to come up with those integrals – while they can be numerically evaluated if needed, not all problems are easily expressible in the required inner product integral form. But like B&G, this method does address the tradeoff between variance and resolution of the model.
- **More general orthonormal basis functions:** This is the approach focused on the most in the Aster/Borchers/Thurber and Menke books, also related to the B&G method above but slightly different. While the Gram matrix technique requires the same number of parameters as data points, this technique does not, and allows for many other choices for the basis functions (which relate the discrete model parameters with the continuous model function). A common one is to just finely discretize the continuous function, which makes for many little square-pulse basis functions. Whatever the choice of orthonormal basis functions, the handy thing about using them is that the norm of the model parameters equals the norm of the continuous function, which comes into

play when translating the estimation of the continuous model function into a parameter estimation problem. Also, with this more general method you can treat the forward problem as a black box and even approximate its derivatives with finite differences, so you can just call the forward problem as a piece of computer code. (The Gram matrix method above requires you to formulate those integrals so you cannot take the black box approach for that.)

- **Occam's inversion:** This one is just a special case of the the above methods – an implementation with one particular choice of regularization. In this special case, out of all the infinity possible solutions that fit the data to within the noise, we choose the smoothest (smallest 2nd derivative) solution, which is the one with the fewest features required by the data fit. The idea is to follow the philosophy of Occam's Razor: all else being equal, the simplest explanation is the best one. There are a few similar variations in how to implement this; officially the name refers to the variation used in the original 1987 paper by Constable, Parker, and Constable, but folks are often a little bit loose with the name. By the way, note when Bayesians talk of Occam's Razor they are not talking about this approach – there is an entirely different discussion/debate relating to Occam's Razor in Bayesian statistics which you can go stumble into if you choose.

Bayesian Techniques

- **Bayesian linear problems:** The Bayesian formulation for the best-fit solution of a linear problem with Gaussian prior and posterior distributions is actually exactly the same as the frequentist one. The inverse of the Bayesian prior covariance matrix equals the frequentist regularization term for the right choice of tradeoff parameter, but in the Bayesian approach one does not make an L-curve to determine that right choice of tradeoff parameter. (This will make more sense by the time you have seen the mathematical details in class.) Rather, the tradeoff parameter is already part of the prior knowledge in Bayesian philosophy. But note that while the best-fit solutions are the same in the frequentist and Bayesian linear problems of this type, the covariance matrix of the solution is not the same, and has a different formulation between the two due to the different philosophical approaches.

- **Conjugate priors leading to filtering:** In Bayesian inversion, a prior model probability is updated with measured data to compute the posterior model probability. In general this does not require any constraints of linearity or of Gaussian noise and model uncertainty. But if you do constrain the problem such that for a particular data misfit metric ("data likelihood function"), the prior distribution has the same parameters as the posterior distribution (e.g. mean and covariance are the parameters of the Gaussian distribution), then you can string together a series of these Bayesian problems and keep updating the parameters of the distributions. The k^{th} posterior distribution computed becomes the $(k+1)^{\text{th}}$ prior. The term for this is that the given data likelihood function has a *conjugate prior* associated with it. This process is exactly what happens in a filtering problem; working out equations for updating those distribution parameters is where all the work lies in developing a new filter.

- **Markov-Chain Monte Carlo (MCMC):** This is a numerical technique, but the most general if you can accept the Bayesian philosophy of specifying prior information. Here there are no constraints at all on the form of the prior and posterior model probabilities. The problem is formulated such that the posterior probability is sampled many times in a random walk, and the random walk is designed to prefer regions of higher probability in just the right proportion to accurately approximate the whole posterior probability. There are a number of variations of MCMC, including the Metropolis-Hastings version and the Gibbs sampling version as the more commonly seen ones. The drawback to this approach is the computation time can be considerable, and for a slow forward problem the computation time might not even be feasible at all. Also, with a numerical approach you cannot develop analytical relationships between quantities, which might be useful sometimes.

Books and Web Resources

- **Books:**
  - Richard Aster, Brian Borchers, and Clifford Thurber, *Parameter Estimation and Inverse Problems* (Academic Press, 2004).
  - Robert L. Parker, *Geophysical Inverse Theory* (Princeton University Press, 1994).
  - Albert Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM: Society for Industrial and Applied Mathematics, 2004).
  - William Menke, *Geophysical Data Analysis: Discrete Inverse Theory*, Revised Edition (International Geophysics Series) (Academic Press, 1989).
  - Per Christian Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial Mathematics, 1987).

- Andrew H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic Press, 1970).

- **Web Resources:**
  - Andy Ganse's geophysical inversion resources webpage (includes lecture notes from past inverse theory computer lab lectures that I gave):
    `http://staff.washington.edu/aganse/invresources`
  - Prof. Brian Borchers' (New Mexico Tech) inverse theory reading list:
    `http://www.ees.nmt.edu/Geop/Classes/GEOP529/Docs/biblio.pdf`
  - *Inverse Problems* journal online (requires UW IP address):
    `http://www.iop.org/EJ/journal/IP`
  - UW Math Department Inverse Problems seminar series:
    `http://www.math.washington.edu/Seminars/IP`

Andrew Ganse is a PhD candidate in the UW Department of Earth and Space Sciences, focusing on inverse theory and theoretical seismology as they apply to underwater acoustics at the ocean bottom. He also has been an engineer at UW's Applied Physics Laboratory for eight years (plus a few years in the early 1990s as well), working on ocean bottom geophysical problems for most of that time, but some ocean surface radar scatterometry too. When he finishes his PhD degree he plans to remain at APL and continue his ocean bottom geophysical research.