Chapter 9

Inverse Problems

V Rao Vemuri
Professor, Department of Applied Science
University of California, Davis
rvemuri@ucdavis.edu

Summary: Three inverse problems arising in practical applications are presented. The first problem pertains to the modeling of San Fernando Valley groundwater basin and the primary issue is parameter identification in a quasi-linear parabolic partial differential equation. The second problem addresses the question of determining the optimal number and location of wells to pump, clean up and recharge a contaminated aquifer near the city of Livermore, CA. Results pertaining to these two problems have already been published. In contrast, the third problem, still in a formulation stage, addresses the question of chemical process identification; that is, the task of determining what a chemical factory is producing using incomplete and uncertain information gleaned from photographs, type and quantity of raw materials purchased, effluents produced and the basic laws of chemistry, materials science and common engineering practices.

1. Introduction

Inverse problems frequently arise in experimental situations when one is interested in the description of the internal structure of a system given indirect noisy data. Estimating the response of a system given a complete specification of the internal structure, on the other hand, is the *direct* or *forward* problem.

There are several flavors to an inverse problem. Perhaps the simplest among them arises when one has a mathematical description of the internal structure (typically in the form of an equation along with any auxiliary conditions and constraints) and the task is only to estimate the values of the unknown parameters. This is the *parameter estimation* problem. A somewhat difficult problem, picturesquely described by the phrase "can you hear the shape of a drum?" arises when the solution of a partial differential equation (PDE) with specified boundary conditions is known and one is asked to find the shape and extent of the boundary. This may be termed the *boundary identification* problem. A special case of this problem is the *free boundary problem* and is characterized by the occurrence of frontiers or interfaces whose locations are *a priori* unknown. In the so-called *input identification* or *control* problem, one is asked to determine an input or control function that will yield a specified target solution to the problem. Another difficult problem, the *modeling problem*, arises when one is given noisy data observed over irregular intervals of space and time and is asked to develop a mathematical model to fit the observed data. With the advent of high-speed computers and artificial intelligence techniques, this modeling problem went through a metamorphosis and emerged as the *machine learning* problem.

This chapter touches upon a few problems in the evolution of these methods in the later half of the twentieth century starting first from a point of view propounded by Prof. Walter J. Karplus and finally touching upon some of the current trends.

Inverse problems are often formulated by assuming that the underlying phenomenon is a dynamical system characterized by ordinary or partial differential equations, although no such assumption is always essential. In the context of remote sensing experiments, a mathematical formulation often leads to Fredholm integral equations of the first kind [Vemuri, ??]. In both these formulations, often the goal is to build a *mathematical model* of the underlying phenomena. In some contexts a model is only a means to an end. Often, the ultimate goal in such cases is to test the validity of a hypothesis. In these cases, the model is used as a classifier (e. g., neural nets, decision trees) and it matters little whether the model is parametric or non-parametric; the classification accuracy becomes more important. From this point of view the entire field of *Machine Learning* can be treated as an exercise in solving inverse problems. *Data Mining*, a discipline aimed at finding hidden patterns, relations and trends also falls within the scope of inverse problems. While inverse problems associated with data mining represent data-rich situations, there is a class of inverse problems that are data-poor, such as the task of locating hidden structures in an enemy territory. Here the challenge is to combine general knowledge represented by models with specific knowledge represented by data.

By their very nature, inverse problems are difficult to solve. Some times they are ill-posed. A *well-posed* mathematical problem must satisfy the following requirements: existence, uniqueness and stability. The existence problem is really a non-issue in many realistic situations because the physical reality must be a solution. However, due to noisy and/or insufficient measurement data, an accurate solution may not exist. More often, a major difficulty is to find a unique solution; this is especially so while solving a parameter identification problem. Different combinations of parameter values (including boundaries and boundary conditions) may lead to similar observations. One useful strategy to handle the non-uniqueness issue is to utilize *a priori* information as additional constraints. These constraints generally involve the imposition of requirements such as smoothness on the unknown solution or its derivatives, or positivity, or maximum entropy or some other very general mathematical property. A more aggressive approach would be to use Bayesian approach and incorporate prior knowledge probabilistically. Indeed the well-known Tichanov regularization is a special form of Bayesian estimation theory. Another fruitful approach is via search. Given an observed data set, genetic algorithms and genetic programming can be used to probabilistically search a hypothesis space.

The quality of a solution to an inverse problem depends on the constraints imposed. The best constraints are those that not only seek good mathematical properties to the solution but also incorporate prior knowledge about the system. In the context of a problem in geophysics, for example, demanding non-negative permeability is an example of the former category and accommodating abrupt changes in the properties of rock formations (viz., discontinuities) would be an example of the later. It may be difficult indeed to accommodate both the smoothness constraints and the discontinuities simultaneously. That is to say, regularization may prevent the recovery of discontinuities. Recent advances in fractals may eventually provide a natural mechanism to incorporate the fractal or multi-scale nature of the structure of rocks and soils.

2. Modeling an Aquifer

Inverse problems arising in geophysics are of particular interest in this paper. Efficient environmental cleanup of subsurface chemical spills, enhanced oil recovery, safe containment of gases and fluids generated by underground nuclear tests, underground storage of nuclear waste, accurate characterization of water-supply aquifers - all require the modeling and simulation of the flow of fluids (air, water, contaminants) through porous media. The mathematical equations describing these processes are typically non-linear, non-homogeneous partial differential equations.

One of the first inverse problems I had the privilege of working is the task of estimating two parameters, namely transmissibility T (x, y) and storage coefficient S(x, y) of the unconfined aquifer in the San Fernando Valley groundwater basin (Figure 1). The starting point was the equation of water flow through porous medium:

$$\frac{\partial}{\partial x}[T(x,y,h)\frac{\partial h}{\partial x}] + \frac{\partial}{\partial y}[T(x,y,h)\frac{\partial h}{\partial y}] + Q(x,y) = S(x,y,h)\frac{\partial h}{\partial t} \qquad \textbf{(1)}$$

where $h(x, y, t)$ is the elevation of the water table above the mean sea level at the spatial point (x, y) at time t. Nonlinearities enter the equation because the transmissibility and storage coefficient, the two most important parameters, are mildly influenced by the elevation of the water table h, the dependent variable in the PDEs [Vemuri and Karplus, '69]. Irregularly sampled values of $h = h(x, y, t)$ were made available from historical well logs kept at the

Department of Water and Power, City of Los Angeles. These wells were located at irregular points in the *xy*-plane and the well logs were recorded at irregular time intervals. The task had been to treat these data points as the sampled solution of the above equation.

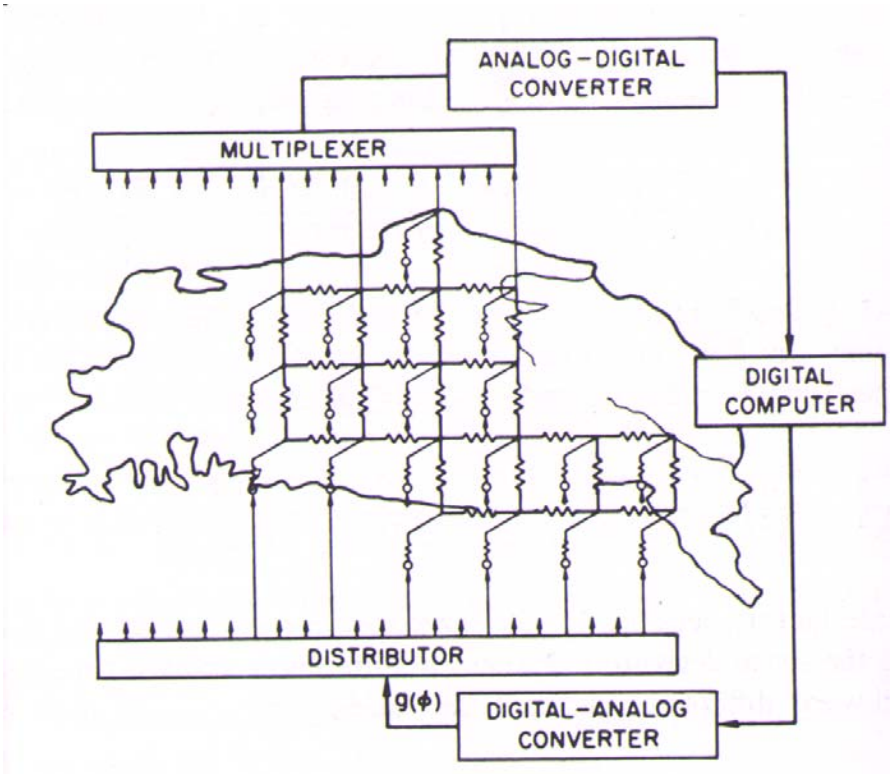

Figure 1: Hybrid Computer Modeling of San Fernando Valley Aquifer

What is left is to define the sources and boundary conditions. The aquifer was recharged periodically using recharge basins. During periods of surplus, water was brought in and allowed to flood large tracts of land, termed recharge basins, and the water was allowed to percolate. This information served to define a "distributed source." Fortunately, much of the San Fernando Valley is surrounded by hills permitting the assumption that the normal derivatives are zero along the impermeable rock formations. Admittedly, the quality and reliability of the data was poor and high-quality solutions were not expected.

All along, the goal had been to use this problem as a vehicle to test the suitability of a newly developed hybrid computer to solve partial differential equations. The hybrid computer, built in Dr. Karplus' laboratory by some of his graduate students, used a hardware subroutine (in this case, a resistance network, as depicted Figure 1) to invert matrices. The idea was to discretize the PDE in question using the usual finite differencing techniques but use the resistance network subroutine to solve the resulting system – rather than using the classical tridiagonal algorithm (in the one-dimesional case) or the alternating direction method (in higher-dimensional cases). As a resistance network relaxes almost instantaneously to its steady state, the solution of the linear tridiagonal or block-tridiagonal system can be obtained instantaneously – at least, theoretically!

With this computational tool in hand, what remained to be done was to solve the parameter identification problem as a multi-point boundary value problem on the hybrid computer. This was accomplished by solving Equation (1) forward in time for the value of *h* and its adjoint backward in time for the value of *v*. A gradient method was used to minimize an error defined in terms of the inner product <*h, v*> [Vemuri and Karplus 1969]. As Equation (1) is non-linear the implementation required linearization and iteration.

Two major difficulties arose in the above process; one is theoretical and one practical. On the theoretical side, it was never clear where the physical boundaries of the aquifer were. The visible topographic boundary of the valley was clearly not the boundary of the aquifer. At this point, it was decided to include the boundary identification problem to be a part of the problem statement. An attempt was made to heuristically adjust the parameters and the boundary, one at a time, in a systematic manner, although in retrospect the results were somewhat lackluster.

One of the causes for disappointment was advancing technology. Although the resistance network inverted the matrix in question instantaneously, the time it took for serial-to-parallel conversion of data, the mechanical crossbar multiplexer to step through the nodes of the resistance network and the subsequent re-conversion of the data from the parallel to serial format – not to mention the A/D and D/A conversion times - more or less sounded the death knell to the hybrid computer. As this work came to an end, solid-state multiplexers came into the market making the crossbar switch obsolete. The SDC 920 computer with its magnetic tape to store and read the programs already looked like a relic in the presence of the newly installed IBM 360 in a room down the hall.

3. Controlling Groundwater Contamination

This example is concerned with the remediation of groundwater in contaminated aquifers. The aquifer under study is a one square mile region once occupied by a petroleum depot in the 1940's. Due to the seepage of hydrocarbons from this facility, the aquifer, located some 90 to 180 feet beneath the ground surface was polluted. The primary goal was the containment of the pollutant.

Remediation is accomplished by pumping the contaminated water out, treating it and recharging the aquifer with the treated water. The problem is to determine the optimum placement of pumping (and recharge or injection) wells and optimum pumping (and recharge) schedules in order to achieve a set of objectives [Cedeno and Vemuri, 1996].

Specifically, the problem solved was the determination of the optimum location of no more than 10 wells, on a 20 x 21 grid, so that three objectives are met. The first objective was to minimize the remediation cost that includes the capital cost for facilities, piping, and running costs associated with water treatment and day-to-day operations. Cost minimization was achieved by picking solutions that stay within a budget. The second objective was to maximize the amount of contaminant removed from the aquifer. This was straightforward and was obtained from the output of SUTRA code in kilograms. The third objective was to prevent unsafe levels of contaminant from leaving the site. The goal was to minimize the concentration of contaminant leaving the site as much as possible. This measure was also obtained from the output of SUTRA in parts per billion (ppb).

This problem was formulated as a multi-objective optimization problem and solved using a genetic algorithm (GA). The aquifer dynamics were simulated by repeatedly solving the PDEs describing the fluid flow using the U. S. Geological Survey's SUTRA code [Voss, '84]. SUTRA (Saturated, Unsaturated TRAnsport) is a 2-D hybrid finite-element/ finite-difference model aimed at solving the governing partial differential equations for groundwater flow and solute transport. As for boundary conditions, the northeast and southeast were treated as no-flow fault zones. Flux boundaries were assumed along the eastern and western edges of the site. A hydraulic conductivity of about 10 feet/day was assumed.

The simulation took into account the three "pump, treat and recharge" facilities already existing on the western edge of the site and the one facility on the eastern edge. Three more "pump, treat and recharge" facilities were planned. The task was to find an optimum location for these three new facilities. Based on data available, each facility was expected to cost $2.5 million with a capacity to clean 70 gallons of water per minute by reducing the concentration of contaminants from 550 ppb to negligible quantities. Thus the scenario described was as realistic as one can make it.

The computational mesh used had a total of 2436 nodes and 2385 elements, which covers the extent of the aquifer, which is much bigger than the one square mile region (see Figure 2). Out of this, a sub-region of size 20 x 21 nodes, covering an area of 5200 ft x 4950 ft. was chosen. Each node in this sub-region is considered as a potential pump location. The distance between nodes is 260 feet, and there are 420 possible pumping locations. Strictly speaking, the decision on where to locate the pumps should be governed by considerations such as the concentration of pollutants, pollution gradients, feasibility of drilling, and so on.

In this study another consideration, namely the computational burden, entered the picture. SUTRA is computationally intensive; it took about 6.5 minutes of elapsed time for evaluating the suitability of each proposed well configuration (called “fitness evaluation” in GA parlance). Each fitness evaluation ran SUTRA for 10 time steps, where each step represented one year. Any finer resolution meant much more time and adequate computational resources could not be mustered to do the job. It was felt that some gains in computational time in terms of fitness evaluations could have been made by restricting the potential pumping sites to the sub-region of the total grid and by using larger time steps.

This knowledge of the flow field, obtained by solving the PDE's using SUTRA, constitutes an input to the genetic algorithm. The algorithm's goal is to find sets of well locations in the 20 x 21 grid that best meets the objectives, subject to the constraint that no more than 10 wells were allowed. The algorithm outputs the values of the objective functions for a range of values of the decision variables. Each of the objective functions constitutes the "modes" of the multi-modal function over which the genetic algorithm conducts its search. No explicit fitness function, for use with the GA, was defined in this problem. The fitness of each individual solution was implicitly given by its objective values and its rank against other solutions.

Table 1 shows a set of solutions for a run of the GA. Only improvements from previous generations are shown. Only solutions meeting the regulatory limit and having a higher amount of contaminant removed from previous generations are shown. The last column ranks the solutions according to their cost. Lower rank values indicate lower cost solution. As expected, improvements in one objective had a negative impact in other objectives during the initial generations.

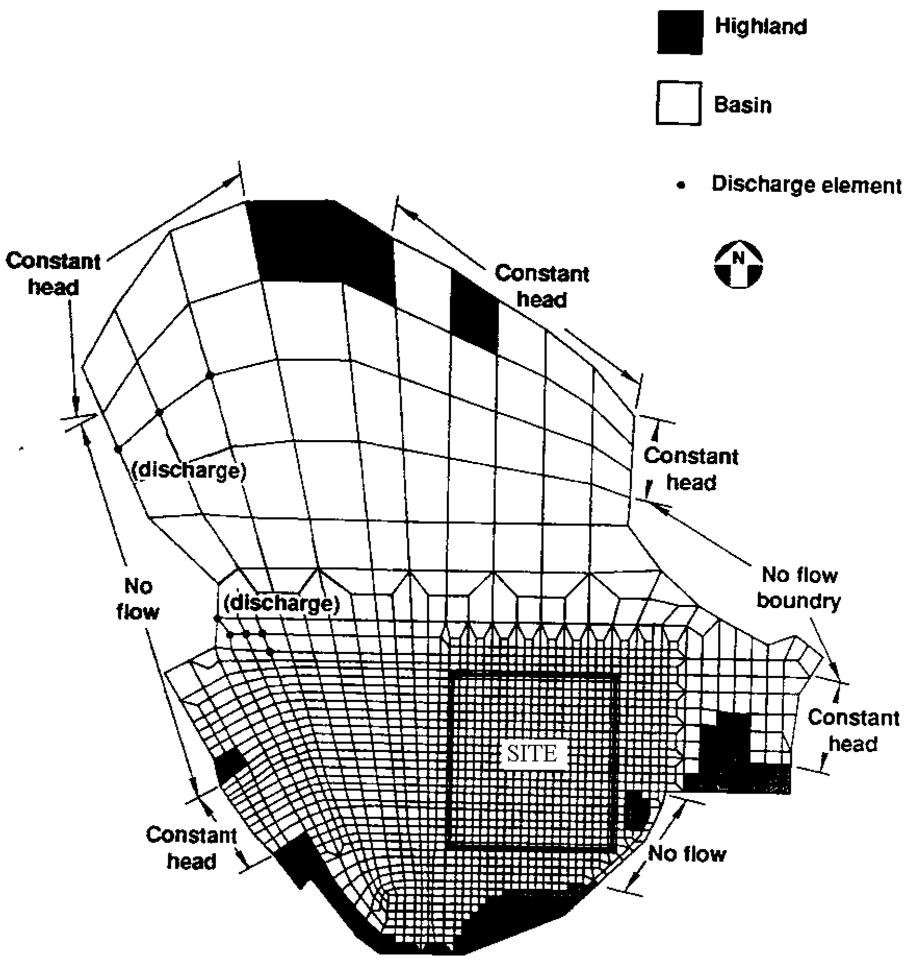

Figure 2: SUTRA nodes and elements in mesh

The last column of Table 1 ranks the solutions according to their cost. Lower rank values indicate lower cost solution. As expected, improvements in one objective had a negative impact in other objectives during the initial generations. The competition by solutions having above average ranking in particular objectives was visible again and as a result offspring with good ranking in must objectives appeared in later generations.

4. Iterative Inversion with Genetic Algorithms

An interesting problem in geophysics is to obtain information about the distribution of material underground. A related problem is that of locating hidden structures and facilities involved in clandestine operations. A relevant problem of contemporary interest is the task of identifying what, if any, chemical warfare agents are being produced at some location. As many chemical processes are "dual-use" the same raw ingredients may go into the production of a fertilizer or an explosive; the difference is the production pathways used by the processes in question.

Consider, for instance, the problem of locating invisible underground structures. It is relatively straightforward to gather field data from airborne surveys and use this data for inversion purposes. In the so-called *model-free* methods data from gravity or magnetic surveys can be used to find the distribution of material that satisfies the observations in a least-square sense. In the so-called *assumed source* methods, the unknown bodies are described with simple, regular shapes such as spheres, polyhedra, thin layers, etc., and the geometrical parameters of these objects and their locations are the target of inversion.

No matter what method is used, solution to a geophysical inversion problem is always non-unique. Additional subjective information is often necessary to resolve the problem further. Such additional information may be provided in the form of, say, (a) a specific starting model, (b) specific parameterization, restricting the search to predetermined geometric shapes, or (c) extra mathematical requirements by demanding solutions that exhibit some unique features (say, the set of all smooth solutions). However this leaves the actual extent of the *ambiguity domain* (i.e., the range of variability within the class of acceptable functions) unknown. Indeed, additional "knowledge" introduced into the problem in a quest to narrow down the scope of the inverse problem may introduce un-intended parameterization that may affect the final result. The question to be answered is: Is it possible that two radically different hypotheses (representing two completely different geological interpretations) satisfy the observed data exactly? In other words, how reliable is the inversion?

A possible solution to this dilemma is to conduct a systematic search within the ambiguity domain and guide this search with expert judgment. Such technique can be used to build a prototype and use the prototype model as a reference model during inversion. During ambiguity search, for instance, a misfit measure can be defined between the prototype and the inversion candidate. For example, the misfit could be the squared difference between the auto-correlation between the prototype model and the candidate inversion. Utility of a metric such as the above misfit measure may or may not be valid for the particular problem at hand. An alternative is to use global search using a method such as a genetic algorithm (GA).

In a GA, one starts with the solutions of a family of forward problems. Each solution is termed a chromosome. Problem parameters are genes and a concatenation of these parameters is the chromosome. Associated with each chromosome is a fitness function that describes how well that solution satisfies the requirements of the problem. Fitness functions can be defined in terms of subjective judgments and objective measures. Then those chromosomes that exhibit high fitness values are selected and a new family of solutions is generated through mathematical operations that are picturesquely termed mating, crossover and mutation.

5. Inverse Problems in Chemical Process Identification

The purpose of the proposed study is to study a new class of inverse problems involving chemical reactions. Our goal is to detect clandestine activities by observing and correlating multiple indicators or "signatures." As the best way to conceal an illegal activity is to make it look like a legal activity, it is not enough to look at obvious signatures such as components and their layout, traces left by testing activity, presence of precursors or degradation byproducts in waste streams, and so on. It is also necessary to impose regularization constraints derived from prior subjective knowledge. Furthermore, each signature has its own precise pathway including the choice of technology. This is a difficult problem to solve and requires a careful formulation from first principles.

Central to our ability to succeed in this effort is a deeper understanding of the reaction pathways. A reaction pathway is nothing but a sequence of elementary reactions through which the precursors of a reaction (starting reactants) are routed through until the final target state is reached. The determination of these pathways entails two phases. The first phase entails the identification of all feasible candidate mechanisms and the second phase requires the selection of the ultimate pathway. Once an understanding is reached, mechanisms for more complex pathways can be determined through a synthesis of plausible elementary reactions. If a rigorous algorithmic method is available to perform this synthesis (which is known to be difficult due to its combinatorial complexity) then that gives us an initial capability to explore.

Insight into this problem can be obtained by studying an analogous inverse problem associated with compartmental models that are very popular in biomedical engineering. In these models, some material enters the system from the external world (say food, medicine, etc.) and is transported through various compartments. There is no tangible output or observable other than the state of the system (e. g. the health) that can be deduced by observing by-products and effluents. Various nutrients, water, drugs and oxygen are typical inputs to the system and various excrements and carbon dioxide, urine, feces are typical effluents and energy is a typical output. Given some measurements on these inputs and outputs can we deduce information about the functioning of the body? This is what a physician does routinely? This task has become routine because of the accumulated knowledge about the anatomy and physiology of human body and the fact that the anatomy of a human body is fixed. This analogy stops here because there is some room for variability in the anatomical configuration of a chemical plant. Nevertheless, this would be a useful metaphor to gain an understanding of the problem.

References

1. Garcke, J. M. Griebel and M. Thess, “Data Mining with Sparse Grids,” *Computing*, Vol. 67, pp 225-253, Springer-Verlag 2001.
2. Tenorio, L. Statistical Regularization of Inverse Problems, *SIAM Review*, Vol. 43, No. 2, pp 347-366, 2001.
3. Karplus, W. J. “Hardware Subroutines,” *Simulation*, Vol. ??, No.??.
4. Vemuri, V. and W. J. Karplus. Identification of Nonlinear Parameters of Ground Water Basins via Hybrid Computation. *Water Resources Research* 5:172-185, 1969.
5. Cedeno, W. and V Rao Vemuri, Genetic Algortihms in Aquifer Management, *Journal of Network and Computer Applications*, Vol. 19, pp 171-187, 1996.
6. Vemuri, V. "Multiple Objective Optimization in Water Resource Systems," *Water Resources Research*, 10:44-48, 1974.
7. Voss, C. I. *SUTRA, Saturated-Unsaturated Transport: A finite-element simulation model for saturated-unsaturated, fluid-density-dependent groundwater flow with energy transport or chemically reactive single-species solute transport*, U. S. Geological Survey, Water Resources Investigations Report 84-4369, 1984.

About the Author.

The author joined UCLA in 1964 and received his doctorate in1968 under the guidance of Dr. Karplus. Now he holds a joint appointment as a professor in the departments of Applied Science and Computer Science. Prof. Vemuri also holds a position as a Computer Scientist at the Lawrence Livermore National Laboratory.