AN OPTIMAL APPROACH TO THE INVERSE PROBLEM

W. Jeffrey and R. Rosner
Harvard-Smithsonian Center for Astrophysics
60 Garden St.
Cambridge, MA 02138

ABSTRACT. We describe how remote sensing problems can be reformulated
within the framework of optimization theory. This reformulation allows
any prior knowledge about the solution to be naturally incorporated into
the solution process. The inversion problem then reduces to a search for
the global extremum in the possible presence of local extrema. Two algo-
rithms are presented that can be used to solve this global optimization
problem, and their application to the helioseismology inverse problem is
detailed.

1. INVERSION OF REMOTE SENSING DATA

The problem of using helioseismology data to constrain the interior struc-
ture of the Sun can be approached either directly by constructing trial
solar models and then solving the "forward" problem (i.e., predicting
the form of the global oscillation spectrum, and comparing it to the data),
or indirectly by solving the "inverse" problem (i.e., using the data to
invert the convolution of the solar structure and the oscillation re-
sponse kernel to obtain the structure information). Each of these ap-
proaches has its particular strengths, but we shall not discuss these
here; rather we focus on the latter approach, and discuss our solution to
the difficulties one faces in solving the inverse problem.

The fundamental difficulty of the inverse problem is that it is ill-
posed. That is, the solution, if one exists, may be non-unique as well
as unstable to small perturbations in the data. Better data will *not* cir-
cumvent these problems; they are inherent in the mathematics of the prob-
lem itself. To demonstrate this, consider the following typical inver-
sion problem:

$$O(y) = \int_a^b K(x,y)F(x)dx$$

where the observations $O(y)$ are related to the quantity of physical in-
terest, $F(x)$, through an integral transform with $K(x,y)$ being a known
kernel. This may, for example, represent an image deconvolution problem
where $K(x,y)$ is the point spread response function, $F(x)$ is the actual
image, and $O(y)$ is the smeared image. In effect, $K(x,y)$ defines an in-
complete basis set for the problem. The non-uniqueness of the solution

129

is a consequence of the fact that this basis is not necessarily complete, so that any function not spanned by the basis may be added to the solution without affecting the observables. Hence given just the observables $O(y)$, we cannot uniquely specify $F(x)$. Similarly, the sensitivity to noise in the data is a consequence of the fact that the natural basis defined by $K(x,y)$ is not necessarily an orthogonal set; in fact, the more ill-posed the problem, the more linearly-dependent are the basis functions in this set. In such a case, a small perturbation in the data can lead to very large changes in the coefficients of the expansion of $F(x)$ in the basis set.

The first systematic study of this problem dates back to Tikhonov (1963), who attempted a solution through regularization of the problem. Most current solution techniques can be reduced to some type of regularizing, including Backus-Gilbert (1970), Phillips-Twomey (1977), and the Maximum Entropy Methods (see Jaynes 1957 and Frieden 1972). Each of these techniques makes explicit use of additional information (e.g., smoothness) to generate a unique solution. Since additional constraints are imposed on the solution, it is important to note that each of these techniques solve a *different* problem from that originally posed. All one can then hope for is that the solution to the new problem will sufficiently approximate the actual solution. For example, the Maximum Entropy methods will find a solution that agrees with the observations as well as possible subject to maximizing the entropy of the solution. There are no assurances, however, that the Sun knows anything about its "configurational entropy".

2. INCORPORATING ALL AVAILABLE INFORMATION

A more natural approach, yet still in the spirit of regularization, is to incorporate all the known or assumed facts about the solution directly into an objective function which one then seeks to minimize. This approach avoids incorporating constraints into the solution which may have no basis in reality (unlike traditional approaches that all apply smoothness constraints which the user has little or no control over).

The difficulty in solving the inversion problem is then converted from the ill-posed nature of the initial problem to the global minimization of a function in the possible presence of multiple local minima. This conversion is an improvement because in many instances, the constraints can be written in such a way as to make the objective function convex; that is, it has only one minimum: the global minimum. If this is the case, then the problem is easily solved: Any descent algorithm, no matter how naive, can be used to find the unique solution. Consider for example the problem of determining the solar interior differential rotation rate from the observed frequency splittings of the 5-minute oscillations. We may have the following information:

$$\text{Observations } \Delta\nu : [\frac{\Delta\nu - \Delta\nu^*}{\Delta\nu}]^2, \text{ where } \Delta\nu^* = \int K\Omega^* dr d\Theta$$

$$\text{Surface Rotation } \Omega(r = R, \Theta) : [\Omega^*(r = R, \Theta) - \Omega(r = R, \Theta)]^2$$

$$\text{Mean Rotation} : [\bar{\Omega}^* - \bar{\Omega}]^2$$

$$\text{Smoothness (e.g., entropy)} : -\sum \Omega^* \ln \Omega^*$$

where Ω is the actual rotation curve, Ω^* is the trial solution, and the overbars on $\bar{\Omega}$ and $\bar{\Omega}^*$ represent a mean (with depth) rotation curve (which may only be available as an inequality from solar oblateness measurements). This list is not exhaustive of convex constraints that may be of value to the helioseismologist. For this example, however, we would generate the following objective function:

$$E = [\frac{\Delta\nu - \Delta\nu^*}{\Delta\nu}]^2 + \phi_1[\Omega^*(r = R, \Theta) - \Omega(r = R, \Theta)]^2 + \phi_2[\bar{\Omega}^* - \bar{\Omega}]^2 - \phi_3 \sum \Omega^* \ln \Omega^*$$

where the ϕ's are Lagrange multipliers that weight the relative importance of each term. We thus seek that Ω^* such that E is minimized. Since this function is convex, there is only one minimum, and the solution can be readily obtained.

3. FINDING THE GLOBAL MINIMUM

If one is, however, unlucky enough to require the use of a non-convex constraint, then more sophisticated numerical techniques are needed to discover the optimal solution. Two relatively new techniques (simulated annealing and neural network processing) show promise of efficiently solving this problem. These algorithms are described in more detail elsewhere (see Jeffrey and Rosner 1986), but a brief description follows: *Simulated Annealing* mimics the way a thermal system will achieve its minimum free energy state if its temperature is lowered slowly. Avoiding the local minima is accomplished because at any finite temperature, there is a finite probability (given by the Boltzmann probability) that a fluctuation consistent with local thermal equilibrium will knock the solution out of a given local minimum, and allow it to proceed to find the global minimum. This procedure can easily be simulated on a computer for any objective function. *Neural network processing* has its origin in the way neurobiologists model the interaction of neurons in the storage and recall of human memory. For the discrete optimization case, Hopfield and Tank (1985) have worked out the details of implementation. For analog problems, Jeffrey and Rosner (1986) implement a version of neural networks that is equivalent to a type of stochastic gradient descent. Both methods give satisfactory results on a variety of simple optimization problems.

As an example of the performance of these techniques, the global minimum of the following function was sought:

$$E(x, y) = [4 - 2.1x^2 + \frac{x^4}{3}]x^2 + xy + [4y^2 - 4]y^2, \quad |x + 1, y| \leq 2$$

which has six minima, two of them global. The results of using neural networks, a modified annealing approach, and a conjugate gradient method are shown in Table 1. Both relative computation time and success in determining the global minimum are listed. The modified annealing uses an

idea proposed by Szu and Hartley (1986) which suggests that a Lorentzian probability distribution rather than Boltzmann should be used to determine changes to the trial solution. We took this alteration one step further, and ran it at zero temperature: that is, only downhill moves would be accepted. Since the Lorentzian has infinite variance, the trial solution can quickly span the entire domain of interest, allowing the solution to "step out" of local minima and find points at lower energy. For this problem, the modified annealing is then sped up by several orders of magnitude over the traditional simulated annealing.

TABLE 1

METHOD	COMPUTATION TIME	GLOBAL MINIMUM FOUND (%)
CONJUGATE GRADIENT	1	56
MODIFIED ANNEALING	2.4	98
NEURAL NETWORK	1.8 4.2	76 100

4. CONCLUSIONS

The following conclusions can be drawn:
- The solar inverse problem can be posed as a variational problem.
- This reformulation allows *any* constraints to be naturally incorporated into the problem.
- If the objective function is convex, then any descent algorithm can be used to determine the global minimum.
- If the objective function is not convex, algorithms do exist to find the global minimum.
- The solution obtained in this way is often more accurate and more stable than solutions found by using traditional techniques.

5. REFERENCES

Backus, G. and Gilbert, F. 1970, *Philos. Trans. R. Soc. Lond. (A)*, **266**, 123.
Frieden, B.R. 1972, *J. Opt. Soc. Am.*, **62**, 511.
Hopfield, J.J. and Tank, D. 1985, *Biol. Cybern.*, **52**, 141.
Jaynes, E.T. 1957, *Phys. Rev.*, **106**, 620.
Jeffrey, W. and Rosner, R. 1986, *Ap. J.*, (to appear Nov. 1).
Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. 1983, *Science*, **220**, 671.
Parker, R.L. 1972, *Geophys. J. Roy. Astron. Soc.*, **29**, 123.
Szu, H. and Hartley, R. 1986, (private communication).
Tikhonov, A.N. 1963, *Sov. Math. Dokl.*, **4**, 1035.
Twomey, S. 1977, *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements* (New York: Elsevier).