Probabilistic residual statics

Daniel Rothman

Abstract

Conventional least squares residual statics solutions are known to fail when noise contamination causes gross errors ("cycle skips") in observed time deviations. The statics problem is presented here as a combination of information in the form of probability density functions (Tarantola and Valette, 1982), allowing more flexibility in the determination of time deviations and less tendency towards gross errors. Solution of a very large optimization problem with many local maxima is necessary. The applicability of a recently proposed technique of Monte Carlo optimization (Kirkpatrick et al, 1983) is examined.

Introduction

Residual statics solutions are conventionally based on a set of observed constant timing differences between input traces and some spatially varying model. The observed time deviations T_{ij} for a trace associated with the ith shot and jth geophone are expressed as a linear combination of surface-consistent and subsurface-consistent parameters S, R, G, and M as

$$T_{ij} = S_i + R_j + G_k + M_k x_{ij}^2 \tag{1}$$

described by Wiggins, Larner, and Wisecup (1976). S_i and R_j refer to the ith shot and jth receiver static, respectively. G_k represents the geologic component of traveltime for the kth CMP, and M_k is a time averaged residual normal moveout coefficient which is multiplied by the offset, x, squared. We wish to find the parameters S_i, R_j, G_k, and M_k, given the observations T_{ij}. The problem is linear, overdetermined, and underconstrained. Using general linear inverse methodology, Wiggins et al solved for the parameters that, in a least squares sense, best satisfied the observed time deviations.

Errors in the theory

The travel-time model in equation (1) employs, like most models, simplifying assumptions that help make the problem tractable. Perhaps the most arbitrary assumption made here concerns vertical raypaths - near surface waves are assumed to travel vertically for all shot-receiver pairs. Other assumptions concerning frequency-independence have been discussed by Chuck Sword (SEP-35). Also restrictive are the notions of surface- and subsurface consistency (for example, velocity anomalies may not be at the surface, and reflectors are not always planar). Thus errors arise in residual statics solutions that are independent of data quality.

A more general, probabilistic model incorporates the errors that exact theory denies. The relationship (1) can be used to derive a conditional pdf $\theta_{ij}(T_{ij} \mid S_i,R_j,G_k,M_k)$. This incorporates the exact solution to the forward problem [find T_{ij}, given S_i, R_j, G_k , and M_k] *and* acknowledged inadequacies in the model. For compactness, we let

$$\mathbf{T} = f(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M}) \tag{3}$$

represent the system of equations described by (1). We choose the errors in the model to be Gaussian, which yields the multivariate pdf

$$\theta(\mathbf{T} \mid \mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M}) = \exp\left\{-\frac{1}{2}\Big[\mathbf{T}-f(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M})\Big]^T \mathbf{C}_m^{-1}\Big[\mathbf{T}-f(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M})\Big]\right\} \tag{4}$$

where $\mathbf{C}_m$ is a covariance matrix.

Solution of the inverse problem

Tarantola and Valette (1982) have approached the general inverse problem as a combination of information given in the form of probability density functions. Their analysis is Bayesian in nature, though more general. Following their conclusions, [specifically, equation (6.7) of their paper] we obtain the pdf of the desired parameters given the pdf's of the data and theory described in the previous two sections. Thus

$$\sigma(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M}) = \int \rho(\mathbf{T})\,\theta(\mathbf{T} \mid \mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M})\,d\mathbf{T} \ . \tag{5}$$

If the Gaussian distributions (2) and (4) are assumed, then the pdf of the parameters becomes [equation (10.4) of Tarantola and Valette]

$$\sigma(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M}) = \exp\left\{-\frac{1}{2}\Big[\mathbf{T}_0-f(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M})\Big]^T (\mathbf{C}_0+\mathbf{C}_m)^{-1}\Big[\mathbf{T}_0-f(\mathbf{S},\mathbf{R},\mathbf{G},\mathbf{M})\Big]\right\} \tag{6}$$

The solution of Wiggins et al is based on the single quantities T_{ij}, weighted in the least squares equations according to their estimated accuracy. Following the work of Tarantola and Valette (1982), the residual statics problem is reexpressed here in more general terms in which the observed data and the theoretical relationship (1) are given as probability density functions, instead of single quantities or exact equations. It is seen that Gaussian assumptions for the error in both the observed data and the model relationships yield the least squares solution described by Wiggins et al. However, as discussed by Donoho (1979), the errors in the observed timing deviations are not Gaussian when the signal-to-noise ratio is low. These non-Gaussian errors manifest themselves as the familiar "cycle-skips." Thus conventional approaches to the statics problem employ erroneous assumptions for the distribution of errors in the observed data when S/N is poor.

Errors in the data

Trace-to-trace time deviations are usually estimated from crosscorrelations. Crosscorrelation is performed for a range of lag times; the lag time yielding the maximum value of the crosscorrelation spectrum is the "picked" timing difference. These picked times T_{ij} are the observed data in the residual statics inverse problem. If the data are relatively uncontaminated by noise, errors in T_{ij} are Gaussian with zero mean. Noise, however, causes crosscorrelation spectrums to contain several maxima of nearly equal magnitude. Choosing T_{ij} according to the absolute maximum in the spectrum often results in gross errors, or "blunders." Donoho (1979) showed that these gross errors are uniformly distributed over the lag range used. Recognizing this problem, we choose to specify probability density functions $\rho_{ij}(T_{ij})$, to be used in lieu of the single, possibly grossly erroneous quantities normally employed. This pdf contains *all* the available information on T_{ij}. For data with low S/N, $\rho_{ij}(T_{ij})$ would likely be multimodal; the modes (maxima) of the distribution would correspond to both cycle-skips and (to some extent) the true value of T_{ij}.

Let **T** represent the set of all possible time deviations for a given dataset, and let $\mathbf{T}_0$ represent the set of true time deviations. If the errors in the time picks are assumed to be Gaussian, we obtain the multivariate pdf

$$\rho(\mathbf{T}) = \exp\left\{-\frac{1}{2}\left[\mathbf{T}-\mathbf{T}_0\right]^T \mathbf{C}_0^{-1}\left[\mathbf{T}-\mathbf{T}_0\right]\right\} \qquad (2)$$

where $\mathbf{C}_0$ is a covariance matrix. We emphasize, however, that for the important case of low S/N, $\rho(\mathbf{T})$ will generally *not* have a Gaussian nor an analytic expression.

In a large problem such as this, the joint pdf in (5) or (6) need not be fully computed. Calculations may be restricted to the maximum likelihood point of the distribution, that choice of parameters which maximizes the pdf. For the Gaussian assumptions of (6), this leads to the minimization of $| \mathbf{T}_0 - f(\mathbf{S,R,G,M}) |^2$, which is the least squares solution of Wiggins et al. We are most interested, however, in $\rho(\mathbf{T})$ as a non-Gaussian pdf, and find ourselves left with the apparently enormous problem of finding the maximum likelihood point of (5).

The main difficulties here are the size of the problem and the multimodality of the general, non-Gaussian $\rho(\mathbf{T})$. How big is the problem? If a line has 1000 midpoints, then we can expect 500 shot and receiver static values, respectively. And also 1000 values for the structural term. We can conservatively assume only one, average RNMO term. If each of these parameters can take on one of 10 possible discrete values, then an undirected, brute force approach necessitates computing $10^{500} \cdot 10^{500} \cdot 10^{1000} \cdot 10 = 10^{2001}$ (i.e., infinity) different solutions to equation (5) on the way towards finding the optimal combination in the parameter space. We have, then, a *very large* problem.

Simplifications

Reductions in computational effort are possible. Assuming that all the $\rho_{ij}(T_{ij})$ and $\theta_{ij}(T_{ij} \,|\, \mathbf{S,R,G,M})$ are independent, equation (5) becomes

$$\sigma(\mathbf{S,R,G,M}) = \int \prod_{i,j} \rho_{ij}(T_{ij})\, \theta_{ij}(T_{ij} \,|\, \mathbf{S,R,G,M})\, dT_{ij} \quad . \tag{7}$$

The range of the product may be constrained to extend over, say, only the length of the cable [effectively the limit of a reliable solution (Wiggins et al, 1976)] instead of the entire seismic line. Now the number of possible solutions is proportional to something like $10^{NSHOTS + NRECS + NCMPS}$, where the terms in the exponent refer to only those terms within the desired maximum wavelength. For a typical seismic survey, this would be something like 10^{100}, still, unfortunately, prohibitively large.

We may further assume that the model (1) is an exact relation. Then

$$\theta_{ij}(T_{ij} \,|\, \mathbf{S,R,G,M}) = \delta_{ij}[T_{ij} - f(\mathbf{S,R,G,M})] \tag{8}$$

and (7) may now be integrated to

$$\sigma(\mathbf{S,R,G,M}) = \prod_{i,j} \rho_{ij}[f(\mathbf{S,R,G,M})] \quad . \tag{9}$$

The optimization problem is no smaller, but the calculations are even more trivial than before.

A combinatoric view

This approach does not result in a significantly smaller optimization problem, but it does yield some more insight into the issue. We can think of $\rho_{ij}(T_{ij})$ as being a series of a few spikes, each weighted according to their relative probability. Treating each component separately, we then find the *optimal* least squares solution - that choice of spikes from each $\rho_{ij}(T_{ij})$ which yields the lowest least squares error. We now have a large combinatoric problem. If each $\rho_{ij}(T_{ij})$ consists of two spikes, and there are, say n CMP's of multiplicity m, then there are 2^{nm} possible solutions. Evaluating each solution is still impossible.

Monte Carlo optimization

The large nature of this problem requires an optimization procedure that is sophisticated enough to avoid functional evaluations at every possible point while still unlikely to get trapped in a non-optimal but local maximum. An optimization technique recently developed by Kirkpatrick, Gelatt, and Vecchi (1983) appears to be adaptable to our problem.

Kirkpatrick and his colleagues have based their technique on properties of statistical mechanics. They propose an analogy between the experimental procedure of annealing (slowly cooling a melted substance until it reaches its pure crystalline ground state) and large-scale optimization. They base their procedure on the Metropolis algorithm (Metropolis et al, 1953), a simple computational technique developed to simulate the behavior of a system of atoms in equilibrium at a given temperature. Statistical mechanical theory shows that a system in thermal equilibrium is characterized by the average or near-average behavior of the system when the average is computed from the ensemble of all possible atomic configurations $\{r_i\}$. Each configuration $\{r_i\}$ is weighted by a Boltzmann probability factor $\exp[-E(\{r_i\})/k_B\tau]$. E is the energy associated with a particular configuration $\{r_i\}$, τ is temperature, and k_B is Boltzmann's constant. The Metropolis algorithm simulates the change of configurations $\{r_i\}$ as time evolves.

Where is the connection with optimization theory? The "energy" in an optimization problem is the objective function being minimized (or maximized), and "temperature" is expressed in the same units as the objective function. By using a Metropolis-like algorithm, a system may be first "melted," brought to equilibrium through random perturbations consistent with the Boltzmann distribution, and then slowly cooled by lowering τ, thereby limiting the acceptable perturbations of $\{r_i\}$ allowable through the Boltzmann probability factor. The key idea is that the energy or objective function need only be evaluated locally. Computational effort appears to scale nearly linearly with the number of parameters, instead of exponentially.

With respect to our statics problem, the local energy or objective function may be expressed as

$$E_{ij} = \sum_{\lambda,\mu,\nu} (T_{i-\lambda\,,\,j-\mu} - S_{i-\lambda} - R_{j-\mu} - G_{k-\nu} - M_{k-\nu} x_{ij})^2 \qquad (10)$$

for μ, ν, and λ implying a summation ranging over, say, 10 shotpoints. Random perturbations of the local energy would mean drawing random values of T_{ij} from $\rho_{ij}(T_{ij})$ and making random guesses for S, R, G, and M. If annealing (cooling) is carried out properly, the system should eventually settle to its optimal state, or some reasonable estimate of it.

The Kirkpatrick algorithm has yet to be tested on the statics problem presented in this report. Fruitful applications potentially exist here and elsewhere in geophysics and will be the subject of future investigations.

ACKNOWLEDGMENTS

Discussions with Fabio Rocca were helpful in formulating the problem.

REFERENCES

Donoho, D.L., 1979, Estimation of time delay at poor S/N, Paper presented at the 1979 EAEG, Hamburg.

Kirkpatrick, S., Gelatt, C.D., Jr., and Vecchi, M.P., 1983, Optimization by simulated annealing, Science, 220, p. 671.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Journal of Chemical Physics 21, p. 1087.

Sword, C., 1983, The generalized frequency-dependent surface-consistent statics problem, SEP-35.

Tarantola, A. and Valette, B., 1982, Inverse problems = Quest for information, Journal of Geophysics, 50, p. 159.

Wiggins, R., Larner, K., and Wisecup, D., Residual statics as a general linear inverse problem, Geophysics, 41, p. 922.