*Proceedings of the National Academy of Sciences*
Vol. 68, No. 2, pp. 291–293, February 1971

# Optimal Solutions to a Linear Inverse Problem in Geophysics

THOMAS H. JORDAN AND JOEL N. FRANKLIN

California Institute of Technology, Pasadena, Calif. 91109

*Communicated by Frank Press, November 16, 1970*

**ABSTRACT This paper is concerned with the solution of the linear system obtained in the Backus-Gilbert formulation of the inverse problem for gross earth data. The theory of well-posed stochastic extensions to ill-posed linear problems, proposed by Franklin, is developed for this application. For given estimates of the statistical variance of the noise in the data, an optimal solution is obtained under the constraint that it be the output of a prescribed linear filter. Proper specification of this filter permits the introduction of information not contained in the data about the smoothness of an acceptable solution. As an example of the application of this theory, a preliminary model is presented for the density and shear velocity as a function of radius in the earth's interior.**

A pertinent problem in the study of the earth's interior is the determination of the material properties as a function of radius from data obtained at the earth's surface. Once determined, such earth models can be used to infer the mineralogical composition and thermodynamic conditions at depth. In an important paper [1] Backus and Gilbert have formalized the linear inverse problem for gross earth data. Subsequent papers by these authors have discussed the resolving power of finite sets of data [2] and the effective resolution of data sets corrupted by noise [3]. These studies adequately demonstrate the inherent nonuniqueness of the solution to this inverse problem and the poor resolution of presently available sets of gross earth data.

However, there exists some valid information concerning the distribution of physical parameters in the earth's interior which cannot be incorporated into the inversion process as gross earth data. An example is the constraint which excludes from geophysical consideration earth models for which the density or intrinsic velocities are negative in any region. Constraints of this form can be extended to exclude models for which the distribution of elastic parameters is inconsistent with laboratory data for plausible mineral constituents. Restriction of the class of earth models obtained from an inversion process to certain bounded sets in the space of all possible models has been discussed by Backus and Gilbert [2] and used by Press [4] in Monte Carlo calculations.

In this paper we propose a method for obtaining optimal solutions to the linear inverse problem for sets of inaccurate gross earth data under the constraint that any solution be the output of some prescribed linear filter. This permits systematic exclusion of solutions which, in some sense, are not *smooth* enough to be compatible with our knowledge about the behavior of physical parameters in the earth's interior. As described in this paper, estimates of the smoothness can be used to restrict the manifold of possible earth models. These estimates can be obtained using high-resolution seismic techniques and experimental and theoretical information on equations of state. The theory presented, which is based on the mathematics of stochastic processes, has been outlined in the general context of ill-posed linear problems by Franklin [5]. We have applied it to the determination of density and shear velocity in the earth. This paper will be limited to a discussion of the basic theory; an exhaustive examination of numerical results will appear in a subsequent publication (T. H. Jordan and D. L. Anderson, in preparation).

## THEORY

Following Backus and Gilbert we consider the separable Hilbert space $\mathfrak{M}$ of all $M$-parameter earth models such that each parameter is a piecewise-continuous, square-integrable, real-valued function of the radius defined on the semi-open interval $(0,R]$ where $R$ is the radius of the earth. For convenience in the following discussion and without loss of generality we may take $M = 1$. Each member of $\mathfrak{M}$ then can be specified by a single function on $(0,R]$. Defined on $\mathfrak{M}$ is the homogeneous scalar product

$$(\mathbf{m}_1,\mathbf{m}_2) = \frac{1}{R^3}\int_0^R m_1(r)m_2(r)r^2dr \tag{1}$$

for all $\mathbf{m}_1,\mathbf{m}_2 \in \mathfrak{M}$. The measure on $\mathfrak{M}$ is a volume measure and is singular at the origin.

Associated with each ordered set $\mathfrak{D}^N$ of $N$ Fréchet differentiable gross-earth functionals and each $\mathbf{m} \in \mathfrak{M}$ are the linear perturbation equations

$$(\mathbf{A}_i,\boldsymbol{\delta}\mathbf{m}) = D_i(\mathbf{m} + \boldsymbol{\delta}\mathbf{m}) - D_i(\mathbf{m}), \qquad i = 1,2,3,\ldots, N \tag{2}$$

which are correct to first order. Here $\mathbf{A}_i$ is the partial differential kernel of $D_i$, the $i$th member of $\mathfrak{D}^N$, and is a member of $\mathfrak{M}$. In general $\mathbf{A}_i$ depends on $\mathbf{m}$. These equations can be written in the form

$$\mathfrak{A}\boldsymbol{\delta}\mathbf{m} = \boldsymbol{\delta}\mathbf{D} \tag{3}$$

$$\text{where } \mathfrak{A} = \begin{bmatrix} (\mathbf{A}_1,\sim) \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{A}_n,\sim) \end{bmatrix} \text{ and } \boldsymbol{\delta}\mathbf{D} = \begin{bmatrix} D_1(\mathbf{m} + \boldsymbol{\delta}\mathbf{m}) - D_1(\mathbf{m}) \\ \cdot \\ \cdot \\ \cdot \\ D_N(\mathbf{m} + \boldsymbol{\delta}\mathbf{m}) - D_N(\mathbf{m}) \end{bmatrix}.$$

The operator $\mathfrak{A}$ maps a small change $\boldsymbol{\delta}\mathbf{m} \in \mathfrak{M}$ of the model into a perturbation $\boldsymbol{\delta}\mathbf{D}$ of the data functionals. The vector $\boldsymbol{\delta}\mathbf{D}$ is a member of the $N$-dimensional Euclidian space $E^N$ associated with the set $\mathfrak{D}^N$. For finite $N$ the problem of determining the solution to (3) is ill-posed in the sense that the solution is not unique [1, 6].

Given a set of $N$ observed gross earth data, we seek an estimate of the difference $\delta\mathbf{m}$ between the representation $\mathbf{m}_E \in \mathfrak{M}$ of the "real" earth and some initial guess $\mathbf{m}_S$ as the results of the application of some bounded linear operator $L:E^N \rightarrow \mathfrak{M}$ to the residual vector $\delta\mathbf{D}$. We take for $\delta\mathbf{D}$ the difference between the observed data and the data functionals calculated for $\mathbf{m}_S$. If the functionals in $\mathfrak{D}^N$ depend linearly on the model and if the data residuals are exact, then $\delta\mathbf{m}$ must satisfy (3). If the set $\mathfrak{D}^N$ contains nonlinear data functionals, as it will in the determination of density and shear velocity using eigenfrequencies of free oscillations, then $\delta\mathbf{m}$ satisfies (3) only to the accuracy that $\mathbf{m}_S$ is $\mathfrak{D}^N$-near $\mathbf{m}_E$, again assuming the residuals are exact.

Of course, in practice the linear system (3) is never exact. Because of inaccurate measurements, inadequacies of the theory (such as neglecting the effects of lateral inhomogeneities), or simply finite arithmetic, each data residual $\delta D_i$ will be associated with some error of "noise" $n_i$. For inaccurate gross earth data, the equation to be solved is

$$\mathfrak{A}\delta\mathbf{m} + \mathbf{n} = \delta\mathbf{D} \tag{4}$$

where $\mathbf{n} \in E^N$ is the vector containing the noise components. The scalar values of these components are unknown; if they were known, we would simply correct the data. However, it is usually possible to describe the noise in terms of its statistics, say, for instance, the variance of each component (the expectation is assumed to be zero). Following Franklin [5], we consider (4) to be a sample of the stochastic equation

$$\mathfrak{A}u_s + u_n = u_d. \tag{5}$$

In this expression $u_s$ is a stochastic process defined over $\mathfrak{M}$ describing the solution, and $u_n$ and $u_d$ are noise and data processes respectively, both defined over $E^N$.

The question we pose is the following: What is the operator $L$ which, when applied to the data process $u_d$, yields the best linear unbiased estimate $\bar{u}_s$ of the solution process $u_s$? The stochastic formulation proves to be advantageous in two respects. First of all, it permits the introduction of information into the operator $L$ about the smoothness of the solution which is not contained in Eq. (4) but specified by the autocorrelation operator of the process $u_s$. Secondly, $L$ is obtained such that the information in inaccurate gross earth data is utilized in an optimal fashion.

As defined by Franklin, we consider $Lu_d$ to be the *best linear estimate* of the solution process if it minimizes the variance of the random variable $(u_e,\mathbf{h})$ for all $\mathbf{h} \in \mathfrak{M}$, where the error process is defined as $u_e = u_s - Lu_d$. Assuming that the random variables $(u_e,\mathbf{h})$ have zero expectation and that the autocorrelation operator $C_{dd}$ of the data process is positive definite, Franklin has shown that if

$$L = C_{sd}C_{dd}^{-1} \tag{6}$$

then $Lu_d$ is the best linear estimate of $u_s$. In this expression $C_{sd}$ is the cross-correlation operator of the solution and data processes. The operator $L$ can be expanded in terms of $\mathfrak{A}$ and the correlation operators for the solution and noise processes:

$$L = (C_{ss}\mathfrak{A}^* + C_{sn})\,(\mathfrak{A}C_{ss}\mathfrak{A}^* + C_{ns}\mathfrak{A}^* + \mathfrak{A}C_{sn} + C_{nn})^{-1} \tag{7}$$

Here $\mathfrak{A}^*$ is the transpose of $\mathfrak{A}$. If the solution and noise are uncorrelated (which we will assume), then $C_{ns} = C_{sn}^* = 0$ and

$$\bar{u}_s = C_{ss}\mathfrak{A}^*(\mathfrak{A}C_{ss}\mathfrak{A}^* + C_{nn})^{-1}\,u_d. \tag{8}$$

In this form the data autocorrelation $C_{dd}$ will, in general, be positive definite if the noise autocorrelation is positive definite. An expression analogous to (8) was first obtained by Wiener [7] in the context of optimal filtering of stationary time series. For a particular sample of the data process, the solution given by (8) is

$$\overline{\delta\mathbf{m}} = C_{ss}\mathfrak{A}^*(\mathfrak{A}C_{ss}\mathfrak{A}^* + C_{nn})^{-1}\,\delta\mathbf{D}. \tag{9}$$

The square of the relative error of estimation for $\overline{\delta\mathbf{m}}$ is

$$\epsilon^2 = (\overline{\delta\mathbf{m}},C_{ee}\overline{\delta\mathbf{m}})/(\overline{\delta\mathbf{m}},C_{ss}\overline{\delta\mathbf{m}}). \tag{10}$$

## CONSTRUCTION OF THE AUTOCORRELATION OPERATORS

In the previous section Eq. (9) was presented as the optimal solution to the system (4). The remaining task is to construct admissible solution and noise autocorrelation operators in a manner which permits their physical interpretation and numerical calculation. For the noise autocorrelation $C_{nn}$, the form is quite simple once we make the reasonable assumption that the noise components are statistically independent. In this case their autocorrelation has the diagonal representation

$$C_{nn} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} \tag{11}$$

where the diagonal element $\sigma_i^2$ is the variance of the $i$th noise component. Note that it has the desired property of being nonsingular if each diagonal element is nonzero.

Specification of a justified form for the solution autocorrelation is somewhat more difficult. Inspection of Eq. (9) suffices to show that our solution has the form $C_{ss}\mathbf{h}$ for some $\mathbf{h} \in \mathfrak{M}$. In this sense $C_{ss}$ acts as a filtration operator. Proper specification of $C_{ss}$ should allow us to discard those solutions which are "unreasonable", say, on the basis of physical constraints that we wish to incorporate into the problem, or perhaps as defined by the resolving power of the data itself. The subtleties involved in the question of which models are reasonable will not be discussed in this paper. Instead, we will provide an example which is convenient and yet illustrates a general approach to the problem of parameterizing $C_{ss}$. In this case we obtain $C_{ss}$ as a member of a one-parameter family of smoothing operators.

For this purpose consider a complete ordered orthonormal set $B$ defined on $\mathfrak{M}$. We assume that the stochastic process $u_s$ has the orthogonal decomposition $\sum_{\psi_n \epsilon B} a_n\psi_n$ which converges in quadratic mean uniformly on $(0,R)]$ for some sequence $\{a_n, n = 1,2,\ldots\}$ of orthogonal Gaussian random variables. Then, by the Karhunin-Loève Theorem [8], the eigenvectors of $C_{ss}$ are members of $B$, and its integral kernel has the expansion

$$C_{ss}(r_1,r_2) = \sum_{\psi_n \epsilon B} \alpha_n^2\,\psi_n(r_1)\,\psi_n(r_2). \tag{12}$$

The coefficient $\alpha_n^2$ is the variance of the random variable $a_n$

The sequence of these coefficients represents the (discrete) spectrum of the operator $C_{ss}$.

To particularize the spectral coefficients, we allow $B$ to consist of the eigenvectors of some Sturm-Liouville system $(\mathcal{L} + k_n^2r^2)\psi_n = 0$ with an appropriate set of homogeneous boundary conditions ordered such that $k_1^2 < k_2^2 < \ldots k_n^2 < \ldots$. Now, let $\alpha_n^2 = k^2/(k_n^2 + k^2)$. This specification of the spectral coefficients has the following desirable properties:

(*i*) if the disposable parameter $k$ is finite, $C_{ss}$ behaves as a low-pass filter,
(*ii*) in the limit as $k^2$ approaches infinity, $C_{ss}$ converges in quadratic mean to the identity operator $I$,
(*iii*) the norm of $C_{ss}$ is, at most, equal to one.

We call $k$ the mean wave number of $C_{ss}$. To obtain $C_{ss}$ in closed form, we observe that $C_{ss} = -k^2G$ where $G$ satisfies $(\mathcal{L} - k^2r^2)\, G = I$ and the boundary conditions. If we choose for $B$ the normalized spherical Bessel functions of zero order regular at the origin and complete on $(0, R]$, then

$$C_{ss}(r_1,r_2) = \frac{k}{2r_1r_2} \times \left\{e^{-k|r_1-r_2|} - \frac{\cosh k(R - r_1 - r_2)}{\sinh kR} + e^{-kR}\frac{\cosh k(r_1 - r_2)}{\sinh kR}\right\} \quad (13)$$

Note that in this expression the first term is dominant.

## NUMERICAL RESULTS

In the numerical studies completed to date, we have sought to determine the shear velocity and density distribution in the earth assuming that the compressional velocity is known. With present machine capability (an IBM 360-75 computer), it has been possible to invert simultaneously up to eighty gross earth data. A representative model (designated 435002 by our computer) is given in Fig. 1. The mean wavelength $\lambda = 2\,\pi R/k$ used in the solution autocorrelation for this calculation was 3000 km. The perturbations of the density and shear velocity were uncorrelated across the core-mantle boundary and uncorrelated with each other. The compressional velocity is that of Johnson's CIT 204 [9] with minor modifications. The data used in the inversion include the mass and moment of inertia of the earth, the eigenperiods for the fundamental mode and two radial overtones (${}_1S_0$ and ${}_2S_0$) of free oscillation, the travel times of direct $S$ waves between 30 and 95°, and the travel times of ScS phases between 0 and 30°. The root mean square (RMS) relative error assigned to the data set, based in part upon the scatter in the observed values, was 0.35%. The model was obtained in two iterations and fits the data with an RMS relative error of 0.23%.

Space limitations prevent a complete discussion of these results; however, certain features should be noted. A strong solution autocorrelation requires that the perturbations be quite smooth, and therefore, for this model, the structural detail in the upper mantle was determined by the starting model. This detail cannot be resolved by the data used in the inversion. However, it is sufficient to satisfy the data; no density reversal in the upper mantle is required.

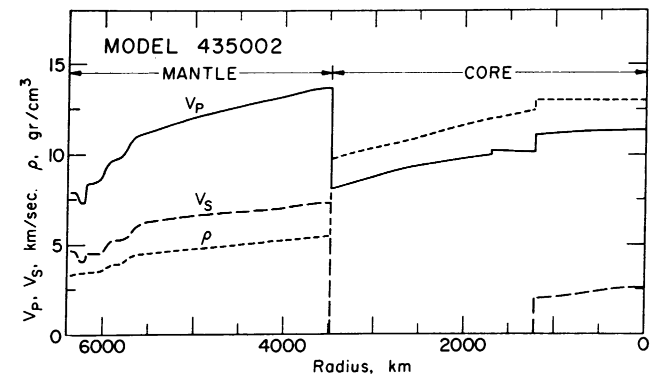


FIG. 1. Compressional velocity, $V_p$, shear velocity, $V_s$, and density, $\rho$, as a function of radius in the Earth determined by the inversion of gross Earth data.

T.H.J. is indebted to Don L. Anderson for his guidance and assistance during the course of this research. The stimulating discussions and helpful comments provided by Bernard Minster and Charles Archambeau are also gratefully acknowledged. The normal-mode programs used as subroutines in the inversion computations were written by Martin Smith and utilized techniques developed by Freeman Gilbert. The travel time routines were developed by Bruce Julian. This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract F44620-69-C-0067. Thomas H. Jordan is affiliated with the Seismological Laboratory, Division of Geological Sciences and Joel N. Franklin is a member of the Division of Engineering and Applied Sciences, California Institute of Technology.

This work is contribution No. 1933 from the Division of Geological Sciences, California Institute of Technology.

1. Backus, G., and F. Gilbert, *Geophys. J. Roy. Astron. Soc.*, **13**, 247–276 (1967).
2. Backus, G., and F. Gilbert, *Geophys. J. Roy. Astron. Soc.*, **16**, 169–205 (1968).
3. Backus, G., and F. Gilbert, *Proc. Royal Soc.*, **266**, 123–192 (1970).
4. Press, F., *J. Geophys. Res.*, **73**, 5223–5234 (1968).
5. Franklin, J. N., *J. Math. Anal. Appl.*, **31**, 682–716 (1970).
6. Smith, M., and J. N. Franklin, *J. Geophys. Res.*, **74**, 2783–2785 (1969).
7. Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (MIT Press, Cambridge, Massachusetts, 1949).
8. Loève, M., *Probability Theory* (D. Van Nostrand, Princeton, N.J., 1955), p. 478.
9. Johnson, L., *J. Geophys. Res.*, **72**, 6309 (1967).