General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

NASA

Technical Memorandum 80301

(NASA-TM-80301) GEOPHYSICAL APPROACHES TO INVERSE PROBLEMS: A METHODOLOGICAL COMPARISON. PART 1: A POSTERIORI APPROACH (NASA) 23 p HC A02/MF A01 CSCL 12A

N79-33925

Unclas

G3/64 38936

Geophysical Approaches to Inverse Problems: A Methodological Comparison Part 1 — A Posteriori Approach

T. I. Seidman and M-J Munteanu

JUNE 1979

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

NOV 1979
RECEIVED
NASA STI FACILITY
ACCESS DEPT.

TM 80301

GEOPHYSICAL APPROACHES TO INVERSE PROBLEMS:

A METHODOLOGICAL COMPARISON

PART I – A POSTERIORI APPROACH

T. I. Seidman
Dept. of Math., Univ. of Md., Baltimore Co.
Baltimore, Md. 21228

and

M-J Munteanu
Goddard Space Flight Center/NASA Code 921
Greenbelt, Md. 20771

June 1979

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

CONTENTS

	Page
ABSTRACT	iii
1. INTRODUCTION	1
2. AN ABSTRACT VIEW OF INVERSE PROBLEMS: *A POSTERIORI* CONSIDERATIONS	2
3. DISCUSSION	7
4. SUMMARY	8
NOTES	9
REFERENCES	16

GEOPHYSICAL APPROACHES TO INVERSE PROBLEMS: A METHODOLOGICAL COMPARISON PART I – *A POSTERIORI* APPROACH

Thomas I. Seidman
Dept. of Math., Univ. of Md., Baltimore Co.
Baltimore, Md. 21228

Marie-Jeanne Munteanu
Goddard Space Flight Center/NASA Code 921
Greenbelt, Md. 20771

ABSTRACT

There exists a variety of general computational methods (and variances) for ill-posed problems such as geophysical inverse problems. These have significant differences in approach and interpretation based on varying assumptions as to, e.g., the nature of measurement uncertainties. This paper addresses the following points: *How are the various approaches related? What considerations should be kept in mind in selecting an approach? To what extent can one confidently rely on the results of such computation?*

PRECEDING PAGE BLANK NOT FILMED

GEOPHYSICAL APPROACHES TO INVERSE PROBLEMS: A METHODOLOGICAL COMPARISON PART I – *A POSTERIORI* APPROACH

1. INTRODUCTION

Most scientific activity can be categorized under the three heads: *model construction, inference, prediction.*[1] Most models, as constructed, involve parameters which must then be experimentally determined; the elements of the relevant *parameter space* (required to specify the presumed physical reality within the general structural framework) consist of a finite or infinite set of numbers and/or functions. In geophysical problems these parameters typically do include specification of functions[2] so that the parameter space is infinite dimensional[3].

Given a specification of the model parameters, the so-called "direct problem" is that of predicting the results of certain observations or measurements. The "inverse problem," with which this paper is concerned, is that of making inferences (about the parameter values) from available observational data. This consists in delineating[4] the set of those parameter values for which the predicted results are consistent with the given data. Much of our present discussion relates to the possible interpretations of this notion of "consistency with the data" and the methodological implications of these possibilities for study of the inverse problem.

We also make a fundamental distinction between *a priori* and *a posteriori* approaches. An *a priori* approach, here, is one which promises an arbitrarily good approximation to the solution *if* it can be furnished "sufficiently adequate and accurate data" while an *a posteriori* approach is one which furnishes an interpretation of a set of data already obtained. Perhaps the most widely acknowledged general approach to geophysical inverse problems is that of Backus and Gilbert ([3]-[9]; see also, e.g., [18]-[21]) which, in terms of this distinction, is to be classified as primarily an *a posteriori* approach. In view of the nature of the organization and funding of geophysical investigation, it will necessarily be approaches of this sort which can typically be of direct practical relevance. On the other hand, we shall argue that the viewpoint of the *a priori* approaches is of fundamental importance

in understanding the underlying nature of ill-posedness and so in understanding *any* approach to dealing with it. Most of the relevant mathematical literature is of this nature; see, for example, Payne [22] and Tikhonov/Arsenin [29] and the further references cited in their bibliographies (T/A [1977] especially, has an excellent bibliographic coverage of the recent Soviet work in this area).

2. AN ABSTRACT VIEW OF INVERSE PROBLEMS: *A POSTERIORI* CONSIDERATIONS

Consider a well-posed *direct* problem for which the data x can be taken ina space X with the corresponding solution y in a space Y. The assumed well-posedness of this direct problem means that there is a well-defined continuous map[5] $A: x \to y$. The *range* $R = \{ y : y = A x \text{ for some } x \text{ in } X \}$ is then the set of y in Y for which the equation

$$Ax = y \tag{2.1}$$

has a solution.

The *inverse problem*, then, is to "solve" (2.1) for x. What do we mean by this? From an *a posteriori* viewpoint one assumes that what may actually be available is an "observation of y." Such an observation consists of the results of a (finite) set of measurements. In general, these will be neither adequate nor exact.

The *inadequacy* of the available observation means only that *even if there were no measurement error*, the set of measurements made would not suffice to specify y exactly. This consideration does not materially affect the present analysis of the implications of inexact measurement and need not be treated separately.

The *inexactitude* or *measurement uncertainty* reflects the limitations of the physical measuring and recording instruments used in the observation. Thus, the nature of the assumptions made about it cannot be viewed as entirely subject to mathematical convenience but should properly be the result of a suitable analysis[6] of the measuring process. At this point one can distinguish two quite different assumptions as to the nature of the inexactitude: either[7]

The measurements come with specifiable error bounds. (2.2a)

or

A statistical distribution (i.e., a joint probability distribution for the measurement components) is available for the errors. (2.2b)

These types lead to different modes of analysis and rather different interpretations[8] of the results; we will proceed to sketch both types of analysis.

For analysis, we assume first that the available data set is *finite*[9] — the results of K scalar measurements so an "observation of y" is a vector ω in $\mathfrak{R}^K$. We write x_* for the *true* (desired) *solution* so $y_* := Ax_*$ is the *true* (observable) *state*. If the taking of exact, error-free measurements is denoted by Ω (this is thus an operator $\Omega : Y \to \mathfrak{R}^K$), then $\omega_* := \Omega y_*$ is the (*true observation* or *ideal data set*; the (vector) error is $E := \omega - \omega_*$ in $\mathfrak{R}^K$.

Our problem, then, is to estimate x_* (or make inferences about it) on the basis of knowledge of ω and analysis of the operators[1c] A and Ω. Actually, even though the nominal domain D of A may be all of X (see note[5]), one typically has additional qualitative information about x_*: on "physical grounds" x_* may be known, e.g., to be everywhere non-negative or to be a strictly increasing function or to be "smoother" (as a function) than an arbitrary element of X need be. Such additional information can often be taken into consideration by suitable modification (restriction) of the domain[11] —say, replacing D by $\tilde{D}$ as the effective domain.

We now discuss—briefly—some approaches available for inverse inference under each of the "uncertainty assumptions" (2.2a,b). For simplicity we concentrate our attention on problems in which X is a (*Hilbert space* and A, Ω (hence, also the composed map $A_K := \Omega A : X \to \mathfrak{R}^K$) are *linear*[12].

Case (2.2a): We take the assumption (2.2a) to mean that a (small) set B in $\mathfrak{R}^K$ is specified[13] in which the error E is known to lie. In the absence of other informaton, all one can then conclude is that x_* lies in the set of "potential solutions"

$$S(\omega) := \{x \in X : [\omega - A_K x] \in B\}. \tag{2.3}$$

We emphasize that *all* elements of $S(\omega)$ are *(equally consistent with the observation* ω and that only the introduction of additional considerations can in any way suggest the selection of any one element of $S(\omega)$ as being a "better" solution than any other. One might hope to select x_K in $S(\omega)$ for which the error bound

$$\| x_* - x_K \| \leqslant \sup \{ \| x - x_K \| : x \in S(\omega) \} \tag{2.4}$$

would give useful information—but, for the problems under consideration, $S(\omega)$ will always be unbounded so the right hand side of (2.4) would, uselessly, be infinite.

On the other hand, given auxiliary information determining an effective domain $\hat{D}$ the set of potential solutions becomes

$$\hat{S}(\omega) := \{ x \in \hat{D} : [\omega - A_K x] \in B \} = S(\omega) \cap \hat{D} \tag{2.3'}$$

(for the case: $\hat{S}(\omega)$ empty, see note[4]). One may now wish to select x_K so (2.4) becomes

$$\| x_* - x_K \| \leqslant \sup \{ \| x - x_K \| : x \in \hat{S}(\omega) \} \tag{2.4'}$$

with the hope that the right hand side of (2.4') might be usefully small.[14]

Since ω, as given, is a point of the K-dimensional space[15] $\mathcal{B}^K$, it is plausible to expect to be able to use this observation to determine K scalar parameters specifying a "nominal solution" x_K to be selected from a suitably chosen K-dimensional subspace $\hat{X}$ of X, i.e., to seek x_K as an estimate (approximation) for x_*. We first consider the selection method:

$$x_K \text{ in } \hat{X},\ A_K x_K = \omega \tag{2.5}$$

(any reasonable choice of $\hat{X}$ makes the restriction $\hat{A}$ of $A_K := \Omega A$ to $\hat{X}$ invertible). Equivalently,

$$x_K := \hat{A}^{-1} \omega. \tag{2.5'}$$

Note that this selection of x_K need not be in $\hat{S}(\omega)$. Now let θ be the angle[16] between $\hat{X}$ and $\eta(A_K)$, let P be the orthogonal projection on $\hat{X}$ and

$$\upsilon := \| \hat{A}^{-1} \| = 1/\inf \{ \| \Omega A x \| : x \in \hat{X} \| \}. \tag{2.6}$$

An easy geometric argument then shows that

$$\| x_* - x_K \|^2 \leq \| x_* - P x_* \|^2 + [\| P x_* - \hat{A}^{-1} \omega_* \| + \| \hat{A}^{-1} \omega_* - x_K \|]^2 \tag{2.7}$$

$$\leq \| x_* - \hat{X} \|^2 + [\| x_* - \hat{X} \| \cot \theta + \upsilon \| E \|]^2 .$$

Similarly, we may follow Backus-Gilbert [1968] in considering not the approximate determination of x_* by x_K but the approximate determination of $\xi \cdot x_*$ by $\xi \cdot x_K$ (where ξ is a linear functional,[17] normalized so $\| \xi \| = 1$). Letting φ be the angle between ξ and the null space of A_K, an argument much like the derivation of (2.7) gives the error bound

$$| \xi \cdot x_* - \xi \cdot x_K | \leq \| x_* - \hat{X} \| \csc \theta \cos \varphi + \upsilon \| E \| . \tag{2.8}$$

We note that control of the factor $\cos \varphi$ in (2.8) is independent of the choice of $\hat{X}$ and is precisely the effect of the "criterion of δ-ness" applied to A_K in Backus-Gilbert [1968].

Minimization of the factors $\cot \theta$ in (2.7) and $\csc \theta$ in (2.8) is achieved by choosing $\hat{X}$ to be $\mathcal{R}(\hat{A}_K)$, the range of $\hat{A}_K$, as is done in Seidman [1975]. There, this choice of $\hat{X}$ appeared as the result of a *variational formulation*, choosing x_K so as to minimize $\| x \|$ subject to the constraint $A_K x = \omega$. More generally (compare, e.g., Seidman [1978a]), one might replace (2.5) by the constrained variational method:

$$x_K \textit{ giving } \min \{ \| x - \tilde{x} \| \; x \in \hat{S}(\omega) \} . \tag{2.9}$$

where $\tilde{x}$ is some plausibly guessed approximation to x_*. If one temporarily ignores any restriction to $\hat{D}$ and if B is the ball[18] in $\mathcal{R}^K$ of radius β, then (for β small enough) it is easy to see that (2.9) is

equivalent to the *unconstrained* variational method:

$$x_K \textit{ giving } \min \left\{ a \| x - \tilde{x} \|^2 + \| \omega - A_K x \|^2 : x \in X \right\} \tag{2.9'}$$

for some suitable choice[19] of the parameter a. The method (2.9') is, of course, the well-known *Tikhonov regularization* approach to ill-posed problems[20,21]. Note that (2.9') gives $(x_K - \tilde{x})$ in $\hat{X} = \mathcal{R}(A_K^*)$ and so can be handled K-dimensionally. In the case of a Euclidean norm for $\mathcal{R}^K$, (2.9') is equivalent to a *linear* system in $\hat{X}$ (the regularized *normal equations*; cf., e.g., Nashed [1971]):

$$(a + Q)(x_K - \tilde{x}) = A_K^* (\omega - A_K \tilde{x}) \tag{2.10}$$

where $Q := \hat{A}^* \hat{A}$. This leads to the error bound[22]:

$$\begin{aligned} \| x_* - x_K \|^2 &= \| (a + Q)^{-1} [a P (x_* - \tilde{x}) + A_* E] \|^2 + \| (I-P)(x_* - \tilde{x}) \|^2 \\ &\leq \| (x_* - \tilde{x}) - \mathcal{R}(A_K^*) \|^2 + \left(\frac{\nu^2}{a\nu^2 + 1} \right)^2 [a \| P x_* - P \tilde{x} \| + \| A_K^* E \|]^2. \end{aligned} \tag{2.11}$$

Case (2.2b): The probabilistic analysis assumes a probability distribution P_E over $\mathcal{R}^K$ for the measurement error[23] $E := \omega - \omega_*$; for simplicity we adopt the usual normality assumption[24]—that P_E is a multivariate normal distribution with mean O and covariance matrix $\sigma^2 I$ (corresponding to the assumption that the means in the individual measurements are independent with common variance σ^2, perhaps after scaling). One can then proceed in either of two ways, corresponding to the standard statistical procedures of hypothesis testing (establishing "confidence intervals") and of parameter estimation. We consider only the latter.

The construction of x_K given above by (2.9') may be viewed as defining an estimator with the choices of $\tilde{x}$ and a now corresponding to an essentially Bayesian attitude toward prior information[25]. Note that choice of a large means a small variance for the estimator but its mean, which depends on the variance for the estimator but its mean, which depends on the earlier estimate $\tilde{x}$, reflects less of the effect of the new observation ω. Comparably to (2.11), consider the expectation[26] of $\| x_* - x_K \|^2$ based on this P_E:

$$\mathrm{Exp}\,\| x_* - x_K \|^2 = \| (x_* - \tilde{x}) - \hat{X} \|^2 + \| (a+Q)^{-1}\, a\, P(x_* - \tilde{x}) \|^2 + \sigma^2\, \mathrm{tr}\,[Q(a+Q)^{-2}]$$

$$\leq \| (x_a - \tilde{x}) - \hat{X} \|^2 + \left(\frac{a\nu^2}{a\nu^2 + 1}\right)^2 \| P x_* - P\tilde{x} \|^2 + \frac{1}{4a}\ \mathrm{Exp}\,\| E \|^2 . \tag{2.12}$$

3. DISCUSSION

Summarizing the above we see that the data set ω can provide direct information only of the projection of x_* along $\eta(A_K)$ onto $\mathcal{R}(A_K^*)$. Any inference beyond this must involve the restriction to a "small" set $\hat{D}$ by virtue of *a priori* auxiliary information. The real value of such an *a posteriori* analysis is often just the possibility alluded to in vote[4] of rejecting a hypothetical $\hat{x}$ (or a suggested value for a functional) as "inconsistent with the observation" ω if $\hat{x}$ is not in $\hat{S}(\omega)$. In the probabilistic case, the normality assumption implies that, regardless of what x_* may be, *no* ω is entirely impossible. Given ω, one cannot, then, with absolute certainty rule out any $\hat{x}$ but can only reject the hypothesis that this $\hat{x}$ is, indeed, x_* on the grounds that this would require belief in the occurrence of an event so unlikely (i.e., $\| E \|$ so large) as to constitute a "miracle."

In either deterministic or probabilistic settings, the method (2.9′) is a useful procedure for estimating x_* on the basis of an earlier estimate $\tilde{x}$ together with a new set of observational data ω. From the bounds (2.11), (2.12) we conclude that increase in K will presumably decrease the first term (as $\hat{X} = \mathcal{R}(A_K^*)$ increases) but $\| Px_* - P\tilde{x} \|$ and ν will increase[27] so a must decrease if the second term is to decrease whence, finally, one must require greater accuracy of measurement[28] to make the last term decrease as well. Note that while the bounds (2.11), (2.12) are correct, they are strongly dependent for their precise form on the specific hypotheses made and have been presented here only as suggestive. Along the lines of e.g., Backus/Gilbert [7–9] or Parker [18–20], one could at this point elaborate on the expectation that intrinsically nonlinear problems would *qualitatively* exhibit essentially the same behavior with the linear operator A of the analyses above now obtained via linearization—e.g., the derivative[29] at $\tilde{x}$ of the actual nonlinear map.

The paragraph above, of course, initiates consideration of (asymptotic) *a priori* analysis: viewing the procedure described in the preceding section as merely one a *sequence* of increasingly

more accurate (one hopes!) approximations It should be clear from the above that, unless considerable care is taken in matching one's expectations and procedures to the increasingly poor conditioning of the computation, the approximation could actually get *worse*[30] rather than better.

It is worth noting that the usual procedure—adopted above—of using the data to construct an approximant (estimate) x_K is one which makes sense *only* in this asymptotic context. While we were able to obtain bounds (2.11), (2.12), the first term $\| (x_* - \tilde{x}) - \mathcal{R}(A_K^*) \|$ will (unless one can somehow make effective use of a restriction[31] to a "small" $\hat{D}$) decrease arbitrarily slowly and, in any case, depends on the unknown x_* so one cannot really *know* how good an approximation any particular computed x_K may be. Asymptotically, however, one can ensure convergence to x_* of the sequence of approximants by employing data which "in the limit" is both adequate and accurate[32] $\eta(A_K) \to \{o\}$ and $\| E \| \to o$ sufficiently rapidly as $K \to \infty$).

4. SUMMARY

The principal use of experiment lies in the possibility of rejecting a model. To the extent that the predicted observations are comparatively insensitive to (certain aspects of) the model, one finds it difficult to make useful distinctions without imposing unrealistic requirements on the extent and accuracy of the observations. The role of "inverse theory" must lie primarily in making efficient use of what data may be available and in evaluating the *range* of models consistent with these rather than in selecting a single model. Nevertheless it *is* possible to construct estimates of model parameters (functions) from observations. The flexibility in choosing norms for measuring errors leads to considerable arbitrariness in constructing algorithms - the primary criteria are that norms for measurement errors reasonably reflect the properties of the physical instruments and that norms for parameter functions reflect the combination of theoretical assumptions and the uses to which the results will be put. In general only strong *a priori* assumptions (permitting advance restriction to a "small" set) will make possible useful explicit error estimates but, instead, procedures can be compared on the basis of the asymptotic properties of *approximation schemes* in which they can be embedded[33]. Other than that, procedures can appear quite different (e.g., the explicit pseudo-

solution (2.5) and the variational formulation (2.9')) yet (cf., note[21]) be quite closely related. In general, the relative computational convenience of procedures will strongly depend on the availability of bases for $\dot{X}$ which make the matrix representations of Q (as of $\tilde{A}$ and $\tilde{A}^*$) relatively sparse.

NOTES

[1] To these one may add (e.g., in engineering contexts) a fourth category *design*—to the extent that some parameters or structural elements are at our disposal, select these so that the predicted results are as desired (perhaps, also, optimizing the selection with respect to some specified criterion).

[2] As e.g., the density distribution of the earth or the form of the nonlinearity in a problem of flow through a porous medium.

[3] With a suitable choice of metric, this may often be conveniently represented as (a subset of) some Hilbert space.

[4] Note that this includes the notion of *verification* or *validation* of a theory or model: if the "consistent parameters set" were void, one would have disproved the theory!

[5] The map A need not be defined for *every* x in X and we let D be the set of x for which Ax *is* defined (as A is continuous we always take D to be closed in X). Thus, it is really D which is the set of "admissible data" x and the true domain of A but we abuse notation slightly by continuing to write $A:X \to Y$.

[6] The complexities of such an analysis lead us to sacrifice some potential accuracy to permit estimating the measurement uncertainties in a "more standardized," simplified form. Considerable mathematical work remains to be done to justify this procedure in terms of some notion of "robustness."

[7] This typology is simplifying but not complete: it would certainly be possible, for example, to consider mixed types.

[8] Neither treatment of errors can be considered *superior*—e.g., in the simplest (scalar) setting it must clearly be a matter of "personal taste" to have a preference between a Gaussian error distribution with the standard deviation $\sigma = 10^{-3}$ as against an absolutely certain error bound of 10^{-2}.

[9] Some measurement modes *do* give a continuous record (e.g., a seismographic track) but information theoretic considerations (recording resolution, pen width, etc.) can be used to show equivalence of these with discrete (sampled data) observations.

[10] What is actually under consideration in this section (i.e., for *a posteriori* data interpretation) is just the composed map $\Omega A : X \to \mathfrak{A}^K$. In this context, from this viewpoint, any concern for the operator A as such (rather than as part of the analysis of ΩA) is irrelevant. Similarly, any direct involvement of the intermediate space Y—say, in first using ω to construct an estimate y of the state y preparatory to attempting to invert A to approximate x by A^{-1}—would be essentially misplaced.

[11] Making effective use of such information is often the most significant problem in treating such situations. One special form of this is of particular importance. It can happen that the set $\hat{D}$ (taken as the effective domain of A through the use of auxiliary information) is a *compact* subset of X. In this case, assuming A would be one-to-one (uniqueness in $\hat{D}$ of solutions of (1.1)), a standard theorem in Topology assures us that *after restricting its domain to* $\hat{D}$ the operator A is actually *continuously* invertible so in this case the inverse problem is no longer ill-posed—provided computationally effective use can be made of the compactness of $\hat{D}$. The composed map ΩA is unlikely to be one-to-one, of course, so the above does not apply directly; nevertheless, this compactness is exactly the consideration needed to obtain useful error estimates. This compactness typically follows from smoothness assumptions on x_* via the Relich-Kondrachov Theorems (cf., e.g., Adams).

[12] For example: the direct problem relating the gravitational field in the observable regions to the internal density distribution of the earth is linear. However, the typically measurement process

here obtains only the field *strength* so Ω would not be linear. Indeed, the linearity assumption seems to exclude almost all problems of direct geophysical interest. Nevertheless, the discussion here is of more than merely tutorial significance but may also be relevant to consideration of intrinsically nonlinear problems through linearization; compare the penultimate section of Parker [1977].

[13] A more general formulation would permit B to depend on x_*, e.g., if it would be the *relative* magnitude of errors which could be controlled or estimated. In particular, this would be the case if a significant source of uncertainty in the result would be uncertainty in specification of the model: if (2.1) is used when the "true" operator is A_*, then—even in the absence of any further error—one would have observed $A_* x_*$ while computing on the assumption that $\omega_* - Ax_*$ so $E = [A_* - A]\ x_*$. The nature of the modifications needed for the analysis in the more general case are conceptually clear but inordinately more complex computationally. Compare, e.g., Theorem 4.5 of Seidman [1978a].

[14] The substitution of (2.4′) for 2.4) will not help if $\hat{S}(\omega)$ is still unbounded or is bounded but too large. A fuller understanding of the reason depends on more detailed consideration of the asymptotic analysis (cf., Seidman [1978a]) but we see that if $\hat{D}$ is *compact* (see note[11]) and B small enough then useful estimates may be available.

[15] In practice one may well use measurements embodying some redundancy (linear dependence, in the present context) so that the range of A_K is actually a proper subspace R in $\mathcal{R}^K$. In this case our first step might be to replace the observation ω (which, due to measurement error, might not actually be in the subspace R) by, say, $\tilde{\omega}$—the nearest point in R to ω. In some sense, what we wish to do is to replace the set $[\omega + B]$ by the set $[\omega + B] \cap R$, which typically has the form $[\tilde{\omega} + \tilde{B}]$, and replace $\mathcal{R}^K$ by R. Note that the use of $\tilde{B}$ instead of B *may* correspond to a substantial decrease in the actual uncertainty: this would be the "payoff" for the redundancy. Having replaced $\mathcal{R}^k$ by R, we return to the original notation with the assumption that A_K is "onto."

[16] The angle a between two subspaces U, V is defined by $\cos a := \sup \{u \cdot v : u \in U, v \in V, \| u \| = 1 = \| v \| \}$. Thus, this definition of θ is equivalent to $(\pi/2 - \theta)$ being the angle between the two K-dimensional subspaces, $\dot{X}$ and the range of A_K^*. Note that the minimum value of δ_n, taken over all possible n-dimensional subspaces X_n, is called the *n-width* of the set $\dot{D}$; this notion has been extensively studied in certain contexts (cf., e.g., Jerome [1967]).

[17] Actually taking ξ to be point evaluation, as Backus and Gilbert do, is, of course, possible only if X consists of smooth enough functions for such functionals to be continuous. In that case X will be a *reproducing kernel Hilbert space* (cf., e.g., Aronszajn [1950]) and it may be convenient to express some of our formulas a bit differently in terms of the reproducing kernel. Typical cases lead to the use of spline spaces; see, e.g., Schoenberg [1946], Schumaker [1980], Greville [1969].

[18] This is actually the most reasonable possibility for B *if* one uses an appropriate norm in $\mathcal{R}^K$, related to the nature of the measurement uncertainties. This appropriate norm will *not*, in general, be the usual euclidean norm for $\mathcal{R}^K$ but is more likely to take the form

$$\| \omega \|_R := \max \{ |\omega^{(k)}| / a_k : k = 1, \ldots, k \}$$

where a_k measures the (relative) uncertainty of measurement of the k-th scalar component.

[19] Clearly the choice of a depends on the radius ϵ of B (with $a \to o$ as $a \to o$) but, except asymptotically (compare the considerations of our next section), the appropriate choice is hard to determine; cf., e.g., Morozov [1967], Craven & Wahba [1977] we choose this point at which to note that the norm for X is also somewhat arbitrary (for example, it can always be chosen to make a fairly arbitrary $\dot{X}$ orthogonal to the nullspace of X_K without changing the topology of X) and some study has been given to optimal choice of the X-norm for (2.9'); cf., Cullum [1979].

[20] See Tikhonov [1963] and, for a more complete treatment and an extensive bibliography (especially of the more recent Soviet literature—including over 20 specifically geophysical applications, especially by Glasko and by Prilepko) see Tikhonov-Arsenin [1977].

[21] In the limiting case $\epsilon = 0+$ (corresponding to assumed error-free measurement : $\omega = \omega_*$) one would theoretically be solving (2.5), corresponding to (2.9′) with $a = o$. However, it is characteristic of such "inverse problems that $\hat{A}$, although invertible, is quite badly conditioned even for moderate values of K. We note that regularization is a recommended approach (cf., e.g.,) to solution of such badly conditioned systems in the finite-dimensional case—so *computationally* (2.5) and 2.9′) are not really very distinct.

[22] The bound (2.11) reduces to (2.7) with $\cot \theta = o$ on taking $a = o$. In general, (2.11) offers no improvement over (2.7)—the advantage of (2.9′) is the greater computational stability provided by regularization for the ill-conditioned system (2.5).

[23] As in case (2.2a), our assumption is essentially that the uncertainty is wholly due to measurement uncertainty which is *additive* and *uniform*. Modifications of uniformity would also be possible here along the lines of note[13]. Note that if P_E were to have (small) bounded support B. then all the considerations would apply as well.

[24] See any elementary statistical text on hypothesis testing, especially for the multivariate case (e.g., Morrison [1976]), for the implications and justification (in terms of the Central Limit Theorem) of this assumption.

[25] Here $\tilde{x}$ represents an estimate based on earlier observations and a will be large or small depending on one's degree of confidence in the accuracy of $\tilde{x}$ (actually, in the relative accuracies of $P\tilde{x}$ and of $\hat{A}^{-1}\ \omega$). In some contexts one could simply include the earlier observations with the new in a single step but (2.9′) is a far simpler "updating procedure." Further, the format of the earlier measurements might be quite different—e.g., $\tilde{x}$ might correspond to an earlier estimate of an interior density distribution for the Earth based on seismic data whereas the new observations may be gravimetric—and with a different type of error statistics; indeed, the old measurements may no longer be available at all.

[26] One can carry this analysis a step further by viewing $\tilde{x}$ as itself a random variable (e.g., its randomness might derive from measurement errors in the earlier observations from which it had been obtained); assume $\tilde{x}$, E independent with $P(x_* - \tilde{x})$ having mean o and covariance matrix $\tilde{\sigma}^2 I$ (in $\dot{X}$). Then the expectation in (2.12) must be modified by considering the distribution of $\tilde{x}$. It can be shown that this is minimized by taking $a := \sigma^2 / \tilde{\sigma}^2$ in (2.9′) which gives, altogether,

$$\text{Exp} \| x_* - x_K \|^2 = \text{Exp} \| (x_* - \tilde{x}) - \dot{X} \|^2 + K \frac{\tilde{\sigma}^2 \nu^2}{\sigma^2 \nu^2 + \tilde{\sigma}^2} . \qquad (2.12')$$

Similar analyses have been used (e.g., Franklin 1970) to estimate the optimal a with $\tilde{x} := o$ but viewing the unknown x_*, itself, as a random variable, rather than as strictly determinate.

[27] This is an inescapable consequence of the ill-posedness of the problem: from (2.6), $\nu^2 \geqslant \lambda_K^{-1}$ where λ_K is the K-th eigenvalue ($\lambda_1 \geqslant \lambda_2 \geqslant \ldots > o$) of $A_* A$. For *a priori* analysis an asymptotic knowledge of $\lambda_1, \ldots$ can be useful in *asymptotically* making an appropriate choice of a in (2.9′). This increase in ν, often rapid with K, makes the computation ill-conditioned even though $\hat{A}$ is invertible (see note[21]) which dominates the computational aspects of the problem.

[28] For the probabilistic case it is instructive to "fudge" the interpretation slightly and suppose the measurement operator were really $\tilde{\Omega} : y \to \mathcal{R}^{K'}$ with K′ a multiple of K corresponding to repetition of a "basic" measurement operator. If $\tilde{\sigma}^2 I$ were the covariance matrix in $\mathcal{R}^{K'}$ for measurement errors, then defining Ω by averaging the values of repeated measurements in $\tilde{\Omega}$ gives $(K/K')\tilde{\sigma}^2 I$ as the covariance matrix in $\mathcal{R}^K$ for the errors in Ω. More complicated analyses cover situations in which the actual observational data is "highly redundant" although not just repetitive but the general strategy remains that one makes a "K-dimensional estimate x_K" based on data reduction from an originally K′-dimensional observation ($K' \geqslant K$), using the variance-reducing property of averaging independent errors to produce a statistically more accurate "pseudo-observation."

[29] Note that A, the weak Gateaux derivative at $\tilde{x}$ of $F : X \to Y$ is defined by $\xi \cdot A\, h = d[\xi \cdot F(\tilde{x} + th)]/dt\,|_{t=0}$ provided this determines $A : x \to y$ as a linear map. This is effectively what one obtains (implicitly) if, e.g., (2.9′) is employed (approximately) for nonlinear problem—provided $\tilde{x}$ is already a good estimate for x_*.

[30] A case in point is the analysis in Seidman [(1978b)] (noncovergence) of the method of least squares. Consider selection of x_K by x_K *giving* min $\| y - A\, x \|$ *subject to* $x_K \in \hat{X}$ (dim $\hat{X} = K$) (here it is assumed that $y \approx A\, x_*$ is observable in Y but one might also be considering this with $\| y - A\, x \|$ replaced by $\| \omega - \hat{A} x \|$ if the actual observation $\lambda\omega$ were in $\mathcal{R}^{K'}$ with $K' \gg K$; compare note[28]). Even with entirely error-free observation and calculation, it can be shown that if $\hat{X}$ is viewed as one of a sequence expanding to "fill" X, then the correspondingly computed sequence x_K of "approximants" need not coverage to x_* (one need not even have $\{ \| x_K \| \}$ bounded as $K \to \infty$) if "bad" choices of $\hat{X}$ may be made corresponding to permitting ν to increase faster than necessary.

[31] Even with $\hat{D}$ compact, the estimates available are typically asymptotic—as for the n-widths mentioned in note[16]. Of course, even for well-posed problems in numerical analysis the error estimates typically involve constants (e.g., bounds on higher derivatives) which are not known so one's degree of confidence in the results in "asymptotic"—although "extrapolation" may permit some direct estimation: use partial information and compare the results to estimate accuracy (this also is the crux of Wahba's "cross-validation" analysis of regularization when, in the probablistic case, the variance σ^2 for measurement errors is assumed not to be known in advance).

[32] The vagueness of this assertion is respect to explicit estimates of the errors is, in part, closely related to the considerable arbitrariness involved in the choice of norms. Even restricting oneself as above to quadratic (Hilbert space) norms as measures of approximation, one is free to adopt any of a wide variety of equivalent or inequivalent norms (e.g., $\| x \|^2$ could be $\int x^2 (s)\, ds$

or $\int x^2(s)\,\omega(s)\,ds$ with weighting function $\omega > o$ or, inequivalently, $\int [x^2(s) + |\text{grad } x(s)|^2]\,ds$ or . . .). The choice made will certainly affect both the interpretation and the computational convenience of the method. Particularly convenient are settings for which there are well-studied computational algorithms with the matrices (associated, e.g., with (2.10) taken with respect to an appropriately chosen basis) sparse and not too ill-conditioned; typically this might be the case of spline representations are used in connection with Sobolev norms. Compare note[19].

[33] Presumably one would avoid a method for which the scheme would not be convergent (to x_*) at all. Compare note [30].

REFERENCES

1. Adams, R. A. 1975. Sobolev Spaces Academic Press, New York

2. Aronszajn, N. 1950. Theory of reproducing kernels, Trans, Amer, Math. Soc. 68, pp 337-404

3. Backus, G. E. 1968. Application of a nonlinear boundary-value problem for Laplace's equation to gravity and geomagnetic intensity surveys. Q. J. Mech. Appl. Math. 21:195-221

4. Backus, G. E. 1970. Non-uniqueness of the external geomagnetic field determined by surface intensity measurements. J. Geophys. Res. 75:6339-341

5. Backus, G. E. 1970a. Interference from inadequate and inaccurate data, I. Proc. Nat. Acad. Sci. USA 65:1-7

6. Backus, G. E. 1970b. Interference from inadequate and inaccurate data, II. Proc. Nat. Acad. Sci. USA 65-281-87

7. Backus, G. E., Gilbert, F. 1967. Numerical application of a formalism for geophysical inverse problems. Geophys. J. R. Astron. Soc. 13-247-76

8. Backus, G. E., Gilbert, F. 1968. The resolving power of gross Earth data. Geophys, J. R. Astrom. Soc. 16:169-205

9. Backus, G. E., Gilbert, F. 1970. Uniqueness in the inversion of inaccurate gross Earth data. Phil. Trans. R. Soc. London Ser. A 266:123-92

10. Boor, C. De 1978. Practical guide to splines. Springer Verlag

11. Collum, J. 1979. The assertive choice of the smoothing norm in regularization, Math of Computation Vol 33, p 149-170

12. Craven, P. and Wahba, G. 1977. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of cross-validation (preprint). Tech. Report 445, Univ. Wisc. (Dept. Stat.), Madison

13. Franklin, J. N. 1970. Well-posed stochastic extensions of ill-posed linear problems. J. Math. Anal. Appl. 31, pp 682-717

14. Greville, T. N. 1969. Theory and Applications of Spline Functions Academic Press

15. Jerome, J. W. 1968. Asymptotic estimates of the L2 n-width, J. Math. Anal. Appl. 22, pp. 449-464

16. Morrison, D. F. 1976. Multivariate Statistical Methods McGraw Hill, New York

17. Morozov, V. A. 1967. Choice of parameter for the solution of functional equations by the regularization method, (Transl.). Sov. Math. Dokl. 8, pp 1000-1003

18. Parker, R. L., 1977. Understanding Inverse Theory Ann-Rev. Earth. Planet, Sci. 5, 35-64

19. Parker, R. L. 1972. Inverse theory with grossly inadequate data. Geophys. J. R. Astrom. Soc. 29:123-38

20. Parker, R. L. 1975. The theory of ideal bodies for gravity interpretation. Geophys. J. R. Astrom. Soc. 42:315-34

21. Parker, R. L., Huestis, S. P. 1974. The inversion of magnetic anomalies in the presence of topography. J. Geophys. Res. 79:1587-93

22. Payne, L. E. 1975. Improperly posed problems in partial differential equations, CBMS-NSF Regional Conference

23. Nashed, M. Z. and Votruba, F. G. 1974. A unified approach to generalized inverses of linear operators: I. Algebraic, topological, and projectional properties, Bull. Amer. Math. Soc. 80

24. Schoenberg, I. J. 1946. Contributions to the problem of approximation of equidistant data by analytic functions, Quart, App. Math. 4, Part A, pp 45-99-Part 13, pp 112-141

25. Schumaker, L. L. 1980. The theory of spline function and applications. To appear in Springer Verlag

26. Seidman, T. I. 1975. Solution of singular problems, I: linear problems in Hilbert space, Pac. J. Math. 61, pp 513-520

27. Seidman, T. I. Nonconvergence results for the application of least squares estimation to ill-posed problems. JOTA (to appear)

28. Seidman, T. I. 1978. Convergent approximation schemes for ill-posed problems. Math. Res. Report 78. Univ. of MD, Malto. Co., Baltimore

29. Tikhonov, A. N. and Arsenin, V. Y. 1977. Solutions of ill posed problems, Willey, New York

30. Tikhonov, A. N. 1963. Solution of incorrectly formulated problems and the regularization method, Dokl. Akad. Nau, SSSR 151, pp 501-504

BIBLIOGRAPHIC DATA SHEET

1. Report No. 80301	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Geophysical Approaches to Inverse Problems: A Methodological Comparison Part I – A Posteriori Approach		5. Report Date June 1979 6. Performing Organization Code
7. Author(s) T. I. Seidman & M-J Munteanu		8. Performing Organization Report No.
9. Performing Organization Name and Address NASA/GSFC Greenbelt, Maryland		10. Work Unit No. 11. Contract or Grant No. 13. Type of Report and Period Covered TM
12. Sponsoring Agency Name and Address NASA/GSFC Greenbelt, Maryland		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

There exists a variety of general computational methods (and variances) for treating ill-posed problems such as geophysical inverse problems. These have significant differences in approach and interpretation based on varying assumptions as to, e.g., the nature of measurement uncertainties. This paper addresses the following points: *How are the various approaches related? What considerations should be kept in mind in selecting an approach? To what extent can one confidently rely on the results of such computation?*

17. Key Words (Selected by Author(s)) Inverses, probability distribution, potential solutions Tikhonov regularization stability, convergence	18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*

*For sale by the National Technical Information Service, Springfield, Virginia 22151.

GSFC 25-44 (10/77)