

Some ‘complexity’ issues for ill-posed problems ¹

Thomas I. Seidman

Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21228, USA

e-mail: <seidman@math.umbc.edu>

ABSTRACT: A ‘correct’ interpretation of the computational complexity of an ill-posed problem is formulated as a cost/effectiveness balance for the use of available data to obtain adequate solutions for an application. This composition with an application, is seen as the real problem, leading to the conclusion that some apparently ill-posed problems are, in context, really well-posed with a reasonable associated ‘compositional complexity’.

KEY WORDS: *ill-posed problems, inverse problems, computational complexity, information-based complexity.*

1. INTRODUCTION

¹Thanks are due to the organizers of the SIAM–AMS Conference on Inverse Problems for Partial Differential Equations (Arcata, CA; July–August, 1989) — and especially to W. Rundell — for the opportunity to participate and to present this material. This paper appears in the proceedings of that conference:

T.I. Seidman, *Some ‘complexity’ issues for ill-posed problems*,
in *Inverse Problems in Partial Differential Equations*,
(D. Colton, R. Ewing, W. Rundell, eds.), pp. 17–35, SIAM, (1990).

In addition, this is related to the author’s participation in the special emphasis year on Inverse and Ill-Posed Problems (1986) at the Centre for Mathematical Analysis, Australian National University. Thanks are due to R. S. Anderssen and to N. Trudinger for that opportunity and for the support and hospitality of the Centre during the visit.

This research has been partially supported by the U.S. Air Force Office of Scientific Research under the grant AFOSR-87-0190 and by the National Science Foundation under the grant ECS-8814788.

Our concern in the present discussion is with the identification² of some issues, formulated here as issues of computational complexity, arising in connection with the analysis of ill-posed inverse problems. Our hope is to propose a suitable framework for analysis — perhaps in the sense of the title of [15] — rather than to present much in the way of technical results within such a framework. We also note a likelihood that some of the comments on computational complexity, made here in connection with applications to ill-posed problems, are actually of more general applicability.

Beginning with somewhat malicious paraphrases of the original arguments presented by Hadamard [10], we note three ‘classical’ arguments against ever working with ill-posed problems:

1. hopeless: *Any possible computation from ‘real’ data (finite; with finite precision) must fail since such data is necessarily consistent with arbitrarily widely separated solutions.*

2. artificial: *If one were not so perverse as to make the ‘wrong’ measurements, then one could avoid any necessity of considering ill-posed problems.*

3. illusory: *If one were not so perverse as to select inappropriately mismatched topologies for the notions of small uncertainties of data and of solutions, then these problems would not be ill-posed.*

These arguments are classical because they are well-founded. Except, perhaps, for **2**, which represents an assertion of faith on Hadamard’s part, comparable to [36], they are the natural reaction of a Mathematician to the very definition of an ‘ill-posed problem’ and, of course, are responsible for the introduction of the pejorative term ‘ill-posed’. To these classical arguments we may adjoin a more recent result by Werschulz [33] which provides a somewhat different aspect of the hopelessness of **1**:

1⁺. even more hopeless: *In a precise sense, the computational complexity of an ill-posed problem is infinite.*

Since we must acknowledge the essential ‘rightness’ of these arguments

What are we doing here?

Obviously there must be *something* to be said in response or we would *not* be here. Indeed, a principal stimulus to this analysis was a desire to respond

²As those working with distributed parameter systems know, ‘identification’ is typically an ill-posed problem!

to Werschulz's result $\mathbf{1}^+$ by introducing a different 'precise sense' for which our problems have finite complexity; compare [35].

One possible answer is that, as Mathematicians, we find these problems of interest — so that the analysis of ill-posed problems is enjoyed as an intellectual recreation — and, as in playing a game, that suffices for justification. However, our *present* 'game' is to view this area as a part of Applied Mathematics and the 'rules of the game' then require a more utilitarian answer. For present purposes we take it that one only considers an ill-posed problem when there is some *use* for the answer. This teleological viewpoint will not only determine our concerns but will also suggest our approach. The issues involved, when addressed quantitatively, do seem to be essentially issues of computational complexity, providing the rationale for our title.

Roughly corresponding point-by-point to the rather negative arguments **1–3**, above, we note the following counterarguments:

1'. successful: *We must be doing something right — this stuff gets used, e.g., hospitals buy CAT scanners.*

2'. real: *The selection of problems to be treated (including the nature of the available measured data) is presented to us by the real world and the available technology.*

3'. imposed: *The appropriate choices of topology are imposed by the necessity to reflect the genuine costs involved — of increasing the accuracy of the measurements and of accepting the effect of error in the solution.*

We need a framework which permits analysis with higher resolution to make a start at resolving these (apparently) conflicting sets of arguments; our task is to address the paradox:

How can *both* sets of arguments be right?

It would seem that **2'** and **3'** are adequate responses to **2** and **3** in indicating why our behavior is not, in fact, perverse in the specific contexts of our applications even if one feels that **2** and **3** are normally correct. The principal concern to be addressed is the 'conflict' between **1** (self-evident from the definition) and $\mathbf{1}^+$ (proved) as against **1'** (an observable fact). In general, the thrust of our discussion is that it is precisely the utilitarian aspect of **2'** and **3'** which enables us to resolve this conflict — suggesting that the *interpretation* of 'success' is to be made in terms of some 'real' problem so we anticipate that ill-posed problems can successfully be treated precisely when this real problem is such that the ill-posedness is no longer an issue.

We will indicate, by examples, a number of ways in which this may reasonably be the case.

In view of **1'**, above, one expects that some modification of the framework for analysis in [33], giving **1**⁺, must permit a *new* analysis showing that what is actually being done should not be impossible. Our approach here is to focus on the interpretation of **3'** as an argument for a ‘balance of costs’. A quantitative approach to that can be formulated along the lines of the theory of ‘information-based complexity’ [31], extended to include a notion of ‘compositional complexity’, and we will indicate some of these considerations.

The ‘flavor’ of the present discussion, develops some considerations arising in a number of earlier papers of the author’s (e.g., [5], [25], [28], [26]) and in conversations with A.G. Werschulz, R.S. Anderssen, and P. Sabatier, among others. This has been influenced by the teleological viewpoints of Control Theory and Statistics and we must also acknowledge the influences of [33] and [31], of Hadamard [10], of the discussions of Backus and Gilbert [2] and Parker [22], of the treatments in [12] and [1], etc.

Of particular relevant interest in the Statistical literature we note, e.g., [21], [32], [9] as well as the developments of ‘maximum entropy’ and of ‘sequential decision theory’; we also point out, for the historical parallel to regularization [29], the recent article [23] and its bibliography. From the Control Theoretic literature we note the development of the theory of ‘adaptive (optimal) control’ and, as instances of the introduction of information-based complexity analysis into control theory, the recent papers [4] and [11].

2. WHAT IS THE REAL PROBLEM?

Let us expand on our earlier utilitarian ‘rule’ that one only seeks to solve an ill-posed problem if one has a use for the solution. For definiteness we consider an inverse problem

$$(2.1) \quad \text{Solve } \mathbf{F}(x) = \bar{y} \text{ for } x = \bar{x}$$

which will be an ill-posed problem if, e.g., \mathbf{F} is (injective but) compact. The point is: *Why* do we want to know \bar{x} ?

We emphasize the situation in which the inverse problem under consideration is to be used as a precursor to some particular³ *application* (possibly

³In many applied settings the application may *not* be specifically known at the time of

involving other, known, data which we do not consider here) for which \bar{x} appears in the model as, e.g., a coefficient. From this point of view,

The solution of the ill-posed problem is only an intermediate construct intervening between the available data and the intended application.

Our efforts are thus teleologically directed: we have some *application* which may be taken in the form of a map $\mathbf{A} : \mathcal{P} \rightarrow \mathcal{R} : x \mapsto z$ where ‘what we really want’ is the result z . In isolation, this problem would be a ‘standard’ well-posed problem but of course, we have an obvious difficulty — the ‘value’ of $x \in \mathcal{P}$ is no longer ‘given’ *a priori* but is to be obtained from the data y via the inverse problem: (ill-posed) inversion of \mathbf{F} . Symbolically we have: $y \xleftarrow{\mathbf{F}} x \xrightarrow{\mathbf{A}} z$ and in this form it becomes clear that our *real* problem is the approximate implementation of a composed map

$$(2.2) \quad \mathbf{G} : \begin{array}{ccccc} y & \xleftarrow{\mathbf{F}} & x & \xrightarrow{\mathbf{A}} & z \\ \mathcal{D} & \xrightarrow{\mathbf{F}^{-1}} & \mathcal{P} & \xrightarrow{\mathbf{A}} & \mathcal{R} \end{array}$$

Observe that it is the continuity of this ‘genuine’ map $\mathbf{G} = \mathbf{A} \circ \mathbf{F}^{-1} : \mathcal{D} \rightarrow \mathcal{R}$ which really concerns us and that it is entirely possible for \mathbf{G} to be continuous even though *one* of its factors, corresponding to the ill-posed problem (2.1) of inverting \mathbf{F} is not, when considered in isolation. We emphasize that

Our involvement with an ill-posed problem stems from this factorization of the ‘whole’ problem⁴ \mathbf{G} as $\mathbf{A} \circ \mathbf{F}^{-1}$ — the composition of an ill-posed inverse problem with an application.

To the extent that, at each stage, our computational procedures maintain this factorization, we do actually work with the ill-posed problem but the representation and topology for the *intermediate* variable $x \in \mathcal{P}$ are now seen to be arbitrary — subject to our convenient choice with respect to (2.2). We can thus acknowledge to a certain extent the validity of the original argument **3**, that the ill-posedness may be somewhat illusory.

Let us look at some examples:

measurement and analysis — one is gathering data, etc., to be used later for a possible *variety* of applications. While we recognize this possibility — indeed, the recommendation below for a *sequential* mode of analysis is related to this — our treatment here considers only the case in which this application *is* already known.

⁴This observation was already made in [5], the author’s first foray into the application of ill-posed problems.

EXAMPLE 1: We begin by noting a class of nominally ill-posed problems which really are well-posed. A standard theorem in Topology asserts: Let $\mathbf{F}_* : \mathcal{P}_* \rightarrow \mathcal{D}$ be a continuous injection with $\mathcal{P}_*, \mathcal{D}$ Hausdorff; let $\mathcal{D}_* := \mathbf{F}_*(\mathcal{P}_*) \subset \mathcal{D}$. If \mathcal{P}_* is compact, then \mathcal{D}_* is also compact and the inverse map $\mathbf{F}_*^{-1} : \mathcal{D}_* \rightarrow \mathcal{P}_*$ is also continuous.

We note also that this compactness and the resulting uniform modulus of continuity for \mathbf{F}_*^{-1} lead to convergence rates for computational approximations, etc.

For a typical situation we have $\mathbf{F} : \mathcal{P} \rightarrow \mathcal{D}$ (without any compactness of \mathcal{P}) but the model provides an assumption that the desired solution \bar{x} of $\mathbf{F}(\bar{x}) = \bar{y}$ is actually in a space \mathcal{X} with a compact embedding $\mathbf{E} : \mathcal{X} \rightarrow \mathcal{P}$. (This does *not* yet make the theorem above applicable!) Now suppose one also has a bound M on $\|\bar{x}\|_{\mathcal{X}}$ — again given by the model as an *a priori* assumption. If we now let \mathcal{P}_* be the closure in \mathcal{P} of the image $\mathbf{E}(\mathcal{B}_M)$ (where $\mathcal{B}_M := \{x \in \mathcal{X} : \|x\| \leq M\}$), then this \mathcal{P}_* is compact and we may take \mathbf{F}_* to be the restriction of \mathbf{F} to \mathcal{P}_* and apply the theorem. With the additional information from the modelling assumptions, the problem has ceased to be ill-posed. See [25] for some specific estimates in the linear case in a Hilbert space setting. We do note that to use this effectively it is important that the bound M be known explicitly — this M (together with the estimates of measurement accuracy which we take as part of the data) is an essential ingredient in approximating \mathbf{F}_*^{-1} computationally. ■

EXAMPLE 2: The simplest example in which ill-posedness ‘disappears’ in our present framework of considering the composed problem is that in which $\mathcal{R} = \mathbb{R}$ in (4.5), particularly when \mathbf{A} is linear; compare [1].

Let us suppose that the (direct) map $\mathbf{F} : \mathcal{P} \rightarrow \mathcal{D}$ is also linear and write $\mathbf{A} = \mathbf{a} \in \mathcal{P}^*$. Since we have $\mathbf{G} := \mathbf{F}^{-1}\mathbf{A} : \mathcal{D} \rightarrow \mathcal{R} = \mathbb{R}$ by definition, we now can solve the equation

$$(2.3) \quad \mathbf{F}^* \mathbf{g} = \mathbf{a}$$

to obtain $\mathbf{g} \in \mathcal{D}^*$ — provided, of course, one has $\mathbf{a} \in \text{ran}(\mathbf{F}^*)$. Since \mathbf{F}^{-1} is unbounded (by assumption), so is $[\mathbf{F}^*]^{-1}$; the problem (2.3) is ‘just as ill-posed’ as the original ill-posed problem: solving $\mathbf{F}x = y$ for $x \in \mathcal{P}$, given data $y \in \mathcal{D}$. BUT: while y is only ‘given’ approximately (by measurement), the functional \mathbf{a} is given analytically *a priori* in defining the application so we may use *analytic* methods to solve the new ‘ill-posed problem’ (2.3). Once

one has precomputed \mathbf{g} in this way the problem is no longer ill-posed: the genuinely significant map: $y \mapsto z$ is now easily realized as: $y \mapsto \langle \mathbf{g}, y \rangle$ and *this* computation is well-posed.

Somewhat similar considerations apply to nonlinear problems although the analysis may then be rather more complicated. The issue, however is the same: to precompute (analytically) an effectively usable realization of the functional $\mathbf{g} = \mathbf{a} \circ \mathbf{F}^{-1}$ under hypotheses ensuring that this is continuous.

One can also treat, in almost the same way, the situation in which \mathcal{R} is finite dimensional by working with \mathbf{A} ‘component by component’. For the linear case (compare [24]) we would now precompute $[\mathbf{F}^*]^{-1}\mathbf{A}$ and then proceed as above. ■

EXAMPLE 3: Now suppose $\mathbf{A} : x \mapsto z : \mathcal{P} \rightarrow \mathcal{R}$ represents finding the solution z of a PDE with coefficients involving a parameter $x \in \mathcal{P}$ (system identification). Then the result may well be quite insensitive to certain (structured) perturbations of $x \in \mathcal{P}$. In this context it is very possible that \mathbf{F} is roughly the same as \mathbf{A} , say, with a different choice of initial or boundary data for the PDE. This can be expected to make inversion of \mathbf{F} ill-posed as we well know — **but** *the structure of the uncertainty is precisely such that this uncertainty probably does not really matter much!* The general principle, here, is that if some ‘aspects’ of the unknown parameter x do not (significantly) affect the output y for purposes of identification, then those same aspects will not (significantly) affect the result z desired for the application — the compactness of \mathbf{A} ‘cancels’ the compactness of \mathbf{F} so one may reasonably hope that $\mathbf{G} = \mathbf{A} \circ \mathbf{F}^{-1}$ is continuous.

We note an instructive example of this phenomenon in a rather different context, adaptive control; cf., [3]. Consider the linear discrete-time control problem $x_{k+1} = ax_k + bu_k + e_k$ where u_k is the control and e_k is noise. One wishes to stabilize the system (i.e., minimize the expected long term average variance of the random variable $\{x_k\}$) but the parameters $a, b \in \mathbb{R}$ are unknown. If one controls ‘optimally’ (on the basis of the current maximum likelihood estimate of the pair $[a, b]$), then one’s expectation is (asymptotically) as good as if the parameter pair had been correctly known initially — despite the fact that with probability 1 the sequence of estimates $[a_k, b_k]$ converges to some *incorrect* pair $[\bar{a}, \bar{b}]$. The explanation of the unexpected success, here, is that the optimal control strategy depends only on the ratio a/b and this ratio *is* correctly obtained in the limit. ■

EXAMPLE 4: We emphasize that the argument of Example 3 about the ‘inevitable cancellation of structured insensitivity’ depends essentially on the presumed structural similarity of the two problems represented by \mathbf{F} and \mathbf{A} . On the other hand, if we look, e.g., at the situation where \mathbf{F} is the Radon transform involved in a CAT scan and the application corresponds to surgical intervention, requiring information as to the location of a tumor, then *that* argument is entirely inapplicable. We may ask, then, whether the ‘success’ here is again explicable in terms of well-posedness of a composed ‘genuine’ problem. The relevant observation, here, is that inversion of the Radon transform is, indeed, an ill-posed problem but that the application requires only limited resolution; we do not need the *exact* density distribution corresponding to the tumor. The issue is very much one of ‘ δ -ness’ in the language of Backus and Gilbert [2]. Mathematically the question is like having infinitely many cases of Example 2: $[\mathbf{A} \in \mathcal{A}]$ with each \mathbf{A} corresponding to a ‘local’ functional. The ‘limited resolution’ noted earlier corresponds, in this formulation, to precompactness of the family \mathcal{A} of functionals. It is this compactness which permits effective computation, uniformly for $\mathbf{A} \in \mathcal{A}$ especially since one also has some compactness for the set of densities ‘admissible’ in this model.

We remark that one might also think of this situation as part of a sequential investigation in which, to limit the expected radiation dose, one first asks (at a level of rather low resolution) only for *detection* — with a ‘Yes/No’ answer or, perhaps, ‘Possible: need more information’, leading to a more refined repetition of the detection phase. If/when the answer is ‘Yes’, one then might proceed to localization at a higher resolution. Similar considerations arise in oil exploration, etc. ■

3. THE ROLE OF THE MODEL

At any stage when we might actually be doing computation/estimation we will always have:

the model: This specifies the spaces and mappings involved and also includes all *a priori* assumptions, constraints, and ‘prior information’.

the data: The data set is here always a finite set of numbers (resulting from a finite number of measurements of finite precision); an estimate of the measurement accuracy is part of the data.

a representation: Both the *input* (data set) and *output* of computation are finite sets of numbers. These correspond to parameters in some fixed representation giving the interpretation within the model.

teleology: Specification of what constitutes an ‘adequate’ solution in relation not only to the intrinsic ‘economics’ of the problem but also to the application in which the output will be used.

We wish here to emphasize the roles of the teleology and of the model, which in part, is to be given in the form:

$$(3.1) \quad \begin{array}{ccc} \mathcal{P} & \xrightarrow{\mathbf{F}} & \mathcal{D} : \\ & & \downarrow \mathbf{M} \\ & & \mathbb{R}^n \end{array} \quad \begin{array}{ccc} x & \xrightarrow{\quad} & y \\ & & \downarrow \\ & & \eta \end{array}$$

where \mathbf{M} is the ‘measurement operator’ and the data we are given is $\hat{\eta}$ together with some estimate ε for (some \mathbb{R}^n -norm of) the ‘error’ $\nu := [\hat{\eta} - \eta]$.

We distinguish three possible ‘modes’ of analysis (compare [28]) depending on whether the analyst is able to determine (or influence) the nature of the measurements made. One may alternatively view this as a distinction in timing — are we doing the analysis before or after the data collection. The principal differences, of course, lie in the nature of the relevant questions to be addressed by the analysis, a decision which must precede addressing the question: ‘What is a solution?’

a posteriori The measurement operator is already fixed (e.g., the data may have already been collected) and we can only ask what this is good for:

*What questions can be answered using this fixed set of measurements?
What usable information can be extracted?*

a priori We have a goal, giving a specification of ‘adequate solution’, (i.e., an acceptable level of error in the computed ‘answer’) and we ask:

What combinations of feasible measurements and computations might provide this — and which is ‘optimal’?

sequential This is a mixed mode in which one’s actions at an intermediate time, including possibly the choice of subsequent measurements, may depend on analysis of the (partial) data set obtained so far. We have

a goal (which now may involve a sequence of ‘questions/actions’ to be answered/performed in ‘real time’ with some notions of adequacy and/or optimality) and we ask: *What should we do next?*

We may think of these modes as forming a hierarchy in which such analyses of the *a posteriori* mode as in [2] and [22] are prerequisite to the appropriate *a priori* analysis while this, in turn, is a prerequisite to the more general *sequential* analysis.

It is the *sequential* mode which, perhaps, corresponds most closely to the way we *should* be working.⁵ The appropriate phrase (for which we are indebted to Sabatier) is ‘*navigating in information space*’. We will concentrate our attention in this paper on the other modes but first comment very briefly as to what would then be needed to continue to a *sequential* analysis. At this point we emphasize the value of noting the comparable experience and paradigms in Statistics and in Control Theory for which similar modes of analysis have been developed — in particular, what we have just called the *sequential* mode corresponds roughly to ‘sequential decision theory’ and to ‘adaptive optimal control’ in these fields. The key to this analysis is the observation of ‘invariant embedding’: for such a problem one often may look at the situation at any intermediate (future) time as a problem of exactly the same sort with respect to the ‘future’ relative to that time. If one can make provision for ‘updating’ estimates (efficiently?) as data collection continues, then the estimation made on the basis of the partial information obtained so far simply becomes part of the ‘model’. For Statistics this is essentially a Bayesian approach; in system theory it may be associated with the distinction between the Kalman filter and the earlier Wiener filter. Another interesting paradigm for comparison is that of ordinary conversation in which some uncertainty/ambiguity is initially tolerated and then resolved by asking questions for further clarification.

Now consider the *a posteriori* setting, with the measurements already performed. Since there is inevitable uncertainty in this measured data (indeed, we have taken an estimate ε of this accuracy to be part of the data), we actually have a ball \mathcal{B}_ε in \mathbb{R}^n consisting of ‘all (potential) η consistent with the given data $[\hat{\eta}, \varepsilon]$ ’ and so a *region* $\mathcal{U} \subset \mathcal{P}$, corresponding to this, consisting

⁵Compare the remark in Example 4. Note also that one often wants a *general* procedure/analysis for a *class* of such problems. E.g., one cannot redesign the CAT scanner for each patient! In such a context we should emphasize the *initial* design/analysis — which would put us back in the *a priori* mode.

of ‘all values of x consistent with the given data’, the *uncertainty region*. In one sense, the ‘correct’ solution of the inverse problem is the determination of this set \mathcal{U} . Usually, however, we select some *particular* element $\hat{x} \in \mathcal{U}$ and refer to this as ‘the (approximate) solution’ so one has an ‘approximating solution algorithm’ $\tilde{\varphi} : \hat{\eta} \mapsto \hat{x} : \mathbb{R}^n \rightarrow \mathcal{P}$. Much is often made of the ‘stability’ of this algorithm, i.e., the continuity of this $\tilde{\varphi}$, but it is not clear what useful purpose is really served by this or even why one should care whether $\tilde{\varphi}$ is obtained from a problem with a unique solution or involves some arbitrary ‘selection’ from such solutions. Concern for this stability seems to be primarily a concession to our prejudices. Indeed, observation of ‘wild’ variation of \hat{x} with small perturbations in the data $\hat{\eta}$ might carry important (and possibly useful) information about the ‘size’ and ‘shape’ of the uncertainty region. For a particular selection to be desirable, one would have to show that it is significant — e.g., selecting a ‘most probable’ or ‘central’ element, if such exists. Certainly, if we do make such a selection then we wish to know that the result — here identified with what one gets in the application by using this approximation, i.e., $\mathbf{A}(\tilde{\varphi}(\hat{\eta}))$ — is a good approximation to $z := \mathbf{A}(x)$ but there seems to be no *intrinsic* condition to be imposed on $\tilde{\varphi}$ itself. Compare the phenomenon noted from [3] in Example 3.

If we identify the ‘application’, for example, with specification both of \mathbf{A} and of some region \mathcal{N} in \mathcal{R} (say, a specified neighborhood of some \bar{z}), then a legitimate ‘question’ is: Can we know (by our computations) that $\mathbf{A}(\mathcal{U}) \subset \mathcal{N}$ (or, conversely, that $\mathbf{A}(\mathcal{U})$ and \mathcal{N} are disjoint) so that we can say definitively⁶ that \bar{z} is the correct result (at the level of accuracy specified by \mathcal{N}) or that it is *not*? This assumes \mathbf{M} is given; for an *a priori* analysis we would wish to find some⁷ $[\mathbf{M}, \varepsilon]$ in a set \mathcal{M} of ‘admissible measurement operators’ for which such questions are answerable.

We next assert that the model itself may also be treated as part of the data, to be exploited in the computation — with the advantage that *this* part of the data is assumed to have no associated ‘measurement error’. For the considerations of Example 2 we note that we were precisely able to exploit the assumption that the the model itself was given analytically, exactly. Similarly, the assumptions and constraints built into the model often provide

⁶One would correspondingly look to answer such a question at some specified confidence level in the stochastic case. In this form we have reduced the problem to the Statistical formulation of Hypothesis Testing.

⁷We may, then, also wish to select this measurement operator ‘optimally’, to answer the question at minimal ‘cost’.

the important regularizing estimates which permit effective computation; see the (brief) discussion of this point in Section 5.

We wish also to emphasize at this point that our situation is extremely sensitive to perturbations of the data — this, after all, is the defining aspect of ill-posed problems — but also is extremely sensitive to specification of the model. Certainly, one has the responsibility to consider the effect of this irreducible model uncertainty.⁸ We observe that this may (and usually is) subsumed in the analysis of data uncertainty but that the relevant perturbations of the model correspond to rather *structured* uncertainties in the data and this structuring should be noted and exploited. The compositional framework introduced here relieves this burden somewhat but, as it is itself posed in the framework of the model as paradigm, remains dependent on this.

4. COMPLEXITY

Before we can reasonably ask how complex it may be to do what we want to do, we first must ascertain what it is that we do want to do. We will now be using as our answer to the question “What is a solution?” that a solution is (an algorithm to provide) an approximand $\hat{z} \in \mathcal{R}$ with an estimate of its accuracy. In preparation for this we note the basic paradigm of information-based complexity; see [31] or, more briefly, cf. the review [17].

Given a map $\mathbf{D} : \mathcal{X} \rightarrow \mathcal{Y}$ with \mathcal{X}, \mathcal{Y} Banach spaces, suppose one wishes to ‘solve’ equations of the form $\mathbf{D}(x) = y$ for given $y \in \mathcal{Y}$. For computation, of course, one only works with finite sets of numbers (i.e., in \mathbb{R}^n). Further, one cannot suppose that y is actually ‘given’ but only that we can ‘extract information’ about y using an ‘information operator’ $\mathbf{N} : \mathcal{Y} \rightarrow \mathbb{R}^n$. Then, using $\mathbf{N}(y)$ as input, one will construct some computationally feasible map

⁸We point out that one of the primary uses of our analysis may be to *reject the model entirely* if one would find that the uncertainty region \mathcal{U} is actually empty. This possibility, after all, is part of what ‘scientific method’ is about!

$\tilde{\varphi} : \mathbb{R}^n \rightarrow \mathcal{X}$. Thus we consider the problem:

$$(4.1) \quad \begin{array}{ccc} \mathcal{X} & \xrightarrow{\mathbf{D}} & \mathcal{Y} \\ & \searrow \tilde{\varphi} & \downarrow \mathbf{N} \\ & & \mathbb{R}^m \end{array} \quad \begin{array}{l} \text{Find } n, \mathbf{N}, \tilde{\varphi} \text{ so} \\ \tilde{\varphi} \circ \mathbf{N} \approx \mathbf{S} = \mathbf{D}^{-1} \end{array}$$

On explicitly introducing the finite dimensional spaces $\mathbb{R}^n \equiv \mathcal{Y}_n$ (codomain of \mathbf{N}) and $\mathbb{R}^m \equiv \mathcal{X}_m$ (corresponding to the range of $\tilde{\varphi}$), this diagram becomes

$$(4.2) \quad \begin{array}{ccc} \mathcal{X} & \xrightarrow{\mathbf{D}} & \mathcal{Y} \\ \uparrow \mathbf{I}_m & & \downarrow \mathbf{N} \\ \mathcal{X}_m & \xleftarrow{\varphi} & \mathcal{Y}_n \end{array}$$

where the map $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ represents the actual computation and $\mathbf{I}_m : \mathbb{R}^m = \mathcal{X}_m \rightarrow \mathcal{X}$ provides the ‘interpretation’ of $\varphi(\eta)$ as an element $\tilde{\varphi}(\eta) = \tilde{x} \in \mathcal{X}$. In this paradigm we may refer to the ‘upper level’ as the ‘model’ and to the ‘lower level’ as a computational ‘stage’ with \mathbf{N} and \mathbf{I}_m relating these through suitable parametric ‘representations’ in the infinite dimensional spaces \mathcal{X}, \mathcal{Y} . Typically, the components of \mathbf{N} in \mathbb{R}^n are obtained by function evaluations which may be taken to be ‘expensive’ compared to the rest of the computation so n dominates a cost analysis. Clearly, \mathbf{N} cannot be injective and \mathbf{I}_m cannot be surjective whence, inevitably, there is an error of approximation (truncation error). We ignore round-off, modeling the computation itself as infinitely precise, but one should include the possibility of some corrupting ‘noise’ $\hat{\nu}$ in the action of \mathbf{N} ; compare [19].

Thus the problem of (4.1) is to make $\mu(y)$ small where

$$(4.3) \quad \mu(y) := \|\mathbf{S}(y) - \varphi(\mathbf{N}[y + \hat{\nu}])\| \quad \text{with } \|\hat{\nu}\| \leq \hat{\varepsilon}.$$

Following [31], one may consider, among other possibilities, the *worst case error* ($\mu := \sup \{\mu(y) : \|y\| \leq c, \|\hat{\nu}\| \leq \hat{\varepsilon}\}$) or an *average case error* ($\mu := \text{expectation of } \mu(y) \text{ for some probability distribution of } y \in \mathcal{Y} \text{ and } \hat{\nu} \in$

\mathcal{Y}_n). The parameters of the analysis are then $\hat{\varepsilon}, \varepsilon, n$ — one estimates the minimal ε (best choice of \mathbf{N}, φ) for given n (and given $\hat{\varepsilon}$) and then inverts this to determine the $n, \hat{\varepsilon}$ required to attain some given output accuracy ε . This optimization problem is the core of information-based complexity analysis but later in this section we will modify it somewhat. We remark that the typical formulation is to view this whole setting asymptotically, i.e., embedded in a family of such problems parametrized by $\varepsilon \rightarrow 0$; see Section 5.

Our paradigm for ill-posed problems is initially parallel to the above. The model involves a ‘direct map’ $\mathbf{F} : \mathcal{P} \rightarrow \mathcal{D}$ where we view \mathcal{P} as a ‘parameter space’ in which we are to do identification and \mathcal{D} as the space of (observable) data. Rather than an ‘information operator’ \mathbf{N} , we here have a ‘measurement operator’ \mathbf{M} from \mathcal{D} . Just as the complexity formulation in (4.1) noted that computation *must* take place in a finite dimensional context, we here note that in the real world any actual implementation of measurement *must* lead to a finite set of numbers, with finite precision ε . Parallel to (4.1), then, we have

$$(4.4) \quad \begin{array}{ccc} \mathcal{D} & \xleftarrow{\mathbf{F}} & \mathcal{P} \\ \downarrow \mathbf{M}_k & & \uparrow \mathbf{I}_k \\ \mathcal{D}_k & \xleftarrow{\varphi_k} & \mathcal{P}_k \end{array}$$

As earlier, we identify \mathcal{D}_k with $\mathbb{R}^{n(k)}$ and \mathcal{P}_k with $\mathbb{R}^{m(k)}$, i.e., using $m(k)$ numbers as parameters in some appropriate representation for (an approximation to) the desired infinite dimensional variable x . This notion of ‘representation’ then induces *some* notion as to what we mean by asking that $\varphi_k \approx \mathbf{F}_k^{-1}$ with $\mathbf{F}_k \approx \mathbf{F}$. We refer to the lower level of (4.4) as a **stage** — again invoking an embedding of the problem in an asymptotic formulation in which, in principle, one can make more and more increasingly accurate measurements, giving $n(k) \rightarrow \infty, \varepsilon_k \rightarrow 0$ as $k \rightarrow \infty$ (where ε_k estimates the magnitude of the measurement error ν intrinsically associated with the measurement operator \mathbf{M}_k).

In isolation, the application problem treated as in (4.1), (4.2), becomes:

$$(4.5) \quad \begin{array}{ccc} \mathcal{P} & \xrightarrow{\mathbf{A}} & \mathcal{R} \\ \downarrow \mathbf{N}_k & & \uparrow \\ \hat{\mathcal{P}}_k & \xrightarrow{\mathbf{A}_k} & \mathcal{R}_k \end{array}$$

and setting the diagrams (4.4) and (4.5) side by side we have:

$$(4.6) \quad \begin{array}{ccccccc} \mathcal{D} & & \xleftarrow{\mathbf{F}} & \mathcal{P} & \mathcal{P} & \xrightarrow{\mathbf{A}} & \mathcal{R} \\ \downarrow \mathbf{M}_k & & & \uparrow \mathbf{I}_k & \downarrow \mathbf{N}_k & & \uparrow \\ \mathcal{D}_k & \xrightarrow{\mathbf{F}_k^{-1}} & \mathcal{P}_k & \hat{\mathcal{P}}_k & \xrightarrow{\mathbf{A}_k} & \mathcal{R}_k \end{array}$$

Observe that these can be concatenated to obtain

$$(4.7) \quad \begin{array}{ccccccc} \mathbf{G} : & \mathcal{D} & \xleftarrow{\mathbf{F}} & \mathcal{P} & \xrightarrow{\mathbf{A}} & \mathcal{R} \\ & \downarrow \mathbf{M}_k & & & & \uparrow \\ \mathbf{G}_k : & \mathcal{D}_k & \xrightarrow{\mathbf{F}_k^{-1}} & \mathcal{P}_k & \xrightarrow{\mathbf{A}_k} & \mathcal{R}_k \end{array}$$

if the spaces $\mathcal{P}_k, \hat{\mathcal{P}}_k$ in (4.6) are actually the same with the ‘compatibility⁹ condition’:

$$(4.8) \quad \mathbf{N}_k \circ \mathbf{I}_k \approx \text{id}(\mathcal{P}_k).$$

With a little diagram chasing, we note that in applying (4.3) to the application \mathbf{A} we have

$$(4.9) \quad \hat{\nu} = \hat{\nu}_k = [\mathbf{F}_k^{-1}[\mathbf{M}_k(\mathbf{F}(x)) + \nu_k] - \mathbf{N}_k(x)].$$

⁹Essentially, this corresponds to the use of the same (approximating) representation. To the extent that one might alternatively be working with different representations, the conversion from one to the other may be subsumed either in $[\mathbf{F}_k]^{-1}$ or in \mathbf{A}_k — with attendant costs of computation.

This, of course, induces a relation between a bound $\varepsilon = \varepsilon_k$ on (some norm of) ν_k and a bound $\hat{\varepsilon}$ on (some norm of) $\hat{\nu}$.

This construction shows us that for the complexity analysis of such problems it is necessary to augment the paradigm (4.1) to consider a notion of **compositional complexity** corresponding to the transition from (4.6) to (4.7) via (4.8).

We asserted earlier that the core of information-based complexity analysis is an optimization for the choice of $[n, \tilde{\varphi} \circ \mathbf{N}]$ in (4.1). The notion of ‘optimization’ is usually taken as minimizing n over ‘all’ possible (feasible) choices of \mathbf{N} , $\tilde{\varphi}$, corresponding to a computational model in which, e.g., each component of $\mathbf{N}(y)$ corresponds to a function evaluation which is quite ‘expensive’ compared to computation; more generally, one can introduce a set \mathcal{N} of ‘admissible information operators’ and a *cost functional* $\hat{\mathcal{J}}$ (with the possibility that simply $\hat{\mathcal{J}}(\mathbf{N}, \tilde{\varphi}) := n$). Here $\hat{\mathcal{J}}$ can take into account all three types of cost: (i) the cost of ‘evaluations’ to obtain $\mathbf{N}(y)$ as above, (ii) the computational costs in using φ , and (iii) the (economic) cost with respect to the application of uncertainty/error in the determination of $z \in \mathcal{R}$. The optimization problem associated with (4.1), (4.3), etc., then becomes

$$(4.10) \quad \text{Minimize } \{\hat{\mathcal{J}}(\mathbf{N}, \tilde{\varphi}) : \mathbf{N} \in \mathcal{N}, \mu \leq \bar{\mu}\}.$$

Here $\bar{\mu}$ is viewed as a given condition on the acceptability of the computation for the application. We note that this may be omitted, subsumed in the definition of $\hat{\mathcal{J}}$ by taking $\hat{\mathcal{J}}$ to be infinite when $\mu > \bar{\mu}$. This formulation would also permit more general criteria for admissible accuracy.

In this form we have a closely corresponding formulation for the inverse problem (4.4). We introduce a set \mathcal{M} of ‘admissible measurement operators’ (or, at each stage, a set \mathcal{M}_k) — with the convention that specification of $\mathbf{M} \in \mathcal{M}$ includes specification of n , giving the codomain \mathbb{R}^n , and of the measurement accuracy (corresponding to a suitable ‘uncertainty region’ in the codomain), typically indicated by a bound ε for some \mathbb{R}^n -norm. We then introduce a cost functional \mathcal{J}_0 and, in view of our teleological formulation, associate with (4.4) the optimization problem:

$$(4.11) \quad \text{Minimize } \{\mathcal{J}_0(\mathbf{M}_k, \varphi_k) : \mathbf{M}_k \in \mathcal{M}_k, \varphi_k \circ \mathbf{M}_k = \mathbf{N}\},$$

where we have (temporarily) fixed some information operator $\mathbf{N} \in \mathcal{N}$ for the application problem (4.5) rather than a bound on the ‘output error’ — but noting (4.9).

We note that Example 2 provides another way of looking at (4.11): at each such stage one may simply replace the original \mathbf{A} appearing in (4.5) by the information operator \mathbf{N}_k so

$$(4.12) \quad \begin{array}{ccccc} \mathcal{D} & \xleftarrow{\mathbf{F}} & \mathcal{P} & \xrightarrow{\mathbf{N}_k} & \mathcal{P}_k \\ \downarrow \mathbf{M}_k & & & \nearrow \text{id} & \\ \mathcal{D}_k & \xrightarrow{\varphi_k} & \mathcal{P}_k & & \end{array}$$

and our concern is then to find feasible \mathbf{M}_k and φ_k so that: $\varphi_k \circ \mathbf{M}_k \approx \mathbf{N}_k \circ \mathbf{F}^{-1}$.

The infimum for (4.11), say, for a fixed $x \in \mathcal{P}$, can then be denoted by $\mathcal{J}(\mathbf{N}) = \mathcal{J}(\mathbf{N}; x)$. Note that in this context the relevant set of ‘admissible information operators’ for (4.5) is just given by

$$(4.13) \quad \mathcal{N} = \mathcal{N}_k := \{\varphi_k \circ \mathbf{M}_k : \mathbf{M}_k \in \mathcal{M}_k\}.$$

and the relevant cost functional is

$$(4.14) \quad \hat{\mathcal{J}}(\mathbf{N}, \varphi) := \hat{\mathcal{J}}_0(\mathbf{N}, \varphi) + \mathcal{J}(\mathbf{N})$$

where $\hat{\mathcal{J}}_0$ is the cost functional to be associated with information-based complexity analysis of the application, considered as a problem ‘in isolation’. Note that in specifying $\mathbf{N} \in \mathcal{N}$ and defining $\hat{\mathcal{J}}_0$, etc., we again use the convention that this includes a specification of the associated ‘noise’ level $\hat{\varepsilon}$ and here obtain this through (4.3).

Remark 1: Most of the current analysis of ill-posed problems concentrates on the relation between ε and $\hat{\varepsilon}$ corresponding to (4.3). This amounts to taking $\mathcal{J}_0 \equiv 0$ and to considering the inverse problem in isolation. The first of these seems somewhat ingenuous — even neglecting computational costs, the costs of measurement are often considerable for these problems (especially as one asks for greater accuracy) and this practical consideration is likely to dominate the treatment. The effect of the second is, of course, the principal point of this paper; see also the discussion in the next section.

Remark 2: These considerations seem forced on us in the context of ill-posed problems in order to see how one may have problems with finite

computational complexity in which there is involved an ill-posed problem, having infinite complexity (when considered in isolation) by [33]. On the other hand, it is by no means uncommon in ‘real world’ contexts to have the output of one problem become the input of another (a data ‘pipeline’, to use an image from UNIX) and these considerations seem to be more generally applicable than merely to our present analysis. Using (4.13) and (4.14), given by (4.11) with (4.3), in (4.10) is the essence of our new notion of ‘compositional complexity’.

5. ASYMPTOTIC ANALYSIS

Earlier, we remarked that the typical formulation for complexity (or, for that matter, numerical analysis in general) is to view the whole setting asymptotically, i.e., embedded as a family of ‘stages’ for which one will have increasing accuracy for the result: $\varepsilon \rightarrow 0$.

In the context of (4.4) we are here talking about constructing $\mathbf{I}_k \circ \varphi_k$ with \mathbf{M}_k given although, more precisely, we are in the context of (4.11) constructing φ_k with both the input and output operators (\mathbf{M}_k and \mathbf{N}_k) given — these provide specification of the ‘representation’ above except that we take the specification of ‘error level’ associated with \mathbf{M}_k as part of the ‘data’, along with the value of $\mathbf{M}_k(x)$. The ‘teleology’ corresponds to (4.14) with (4.11) defining \mathcal{J} .

One has both a *local* and a *global* analysis on the basis of (4.11), (4.14), etc. The preceding discussion has had a primarily ‘local’ relevance: Assuming $\hat{\mathcal{J}}_0$ in (4.14) is already ‘known’, we have been concerned with the analysis of the function: $\mathbf{N} \mapsto \mathcal{J}(\mathbf{N})$ which defines the ‘teleology’ through (4.14). Note that we actually have $\mathcal{J}(\mathbf{N}) = \mathcal{J}(\mathbf{N}; x)$ and a significant part of our concern is with the effect of this dependence on the unknown x .

A ‘global’ treatment corresponds, essentially, to the usual asymptotic analysis, in which we embed this problem as a ‘stage’ in a family of similar problems, envisioning potential requests for answers to similarly structured questions at increasing levels of precision (\mathcal{N} shrinking to $\{\bar{z}\}$) and wish simultaneously to treat the whole family. For simplicity, one takes \mathcal{N} to be a ball in \mathcal{R} , characterized by its radius $\bar{\mu}$ which we take as a bound on $\mu(y)$ as in (4.3). One then considers only the asymptotic relation between $\bar{\mu}$ and parameters n, ε characterizing some \mathbf{M}_* . In the formulation of the last

section, this¹⁰ corresponds to taking $\mathcal{J}_0 \equiv 0$ and letting $\hat{\mathcal{J}}_0$ (depending on $\bar{\mu}$ as a parameter, $\bar{\mu} \rightarrow 0$) be 0 when $\mu \leq \bar{\mu}$ and infinity when $\mu > \bar{\mu}$.

For well-posed problems the usual analysis leads to the result that:

One can find the desired result to any degree of specified precision — provided one is prepared to work hard enough.

Here, what we mean by ‘working hard enough’ is: making enough measurements with enough accuracy and then doing enough computation. The history of analysis of ill-posed problems has consisted primarily of demonstrations that this remains true for ill-posed problems even without considering the composed framework introduced here. The principal difference is that the ill-posedness shows up in an explosive growth of the accuracy requirements for the measurements as one seeks more and more precision in the result. We observe that *ill-posedness resides only in the asymptotics* so that for any specific level of desired precision it is quite possible that the problem may be entirely reasonable; compare Example 4.

It is not at all clear, however, that our concern for convergence of an approximation scheme¹¹ has much relevance to actual computation. In practice one never actually approaches the asymptotic limit so these results are significant only as a guide when used to provide error estimates at any stage so as to select appropriate computational parameters. Note, too, that one is only *guessing* the magnitudes of any ‘constants’ appearing in an asymptotic expression unless it is reasonable to view these as available *a priori* from the model — there is a big difference between knowing that some constant exists and knowing (even a rough estimate of) its value.

The traditional way (in numerical analysis) of using *a priori* bounds on the solution or on data uncertainty is to work with algorithms in which these bounds do not appear explicitly but to have them appear in the subsequent error estimates. For ill-posed problems, on the other hand, relevant bounds appear in two ways: (i) as ‘stabilization’, defining compact sets for which the considerations of Example 1 apply¹² and (ii) in ‘controlling’ the computation

¹⁰For the more general formulation envisioned in (4.14), one could similarly consider letting \mathcal{J}_0 and especially $\hat{\mathcal{J}}_0$ depend on *some* parameter and ask for the asymptotic behavior of $\hat{\mathcal{J}}$ (optimized as in (4.11), (4.10)) with respect to this parameter. For the present we do not pursue this greater generality.

¹¹Cf. [26] or, for that matter, any of the myriad of papers on regularization for ill-posed problems — or most of the ‘standard’ numerical analysis literature!

¹²This seems to lie behind many of the known results here, typically through compactness of some embedding.

— e.g., determining the optimal value of the parameter in Tikhonov regularization [30]. Without *a priori* information on such bounds, one can only show convergence, using some asymptotic determination of this parameter; *with* the appropriate choice of the regularization parameter (made computationally on the basis of model-provided information) one obtains explicit optimal convergence rates as in [7] and [20].

EXAMPLE 5: For a variety of problems the *method of Optimal Filtering* [27] uses such bounds explicitly in the formulation of algorithms for which one then may have optimal error estimates. Implementation of this has been worked out specifically for two well-known ill-posed problems: the sideways [6] and the backwards [8] heat equations; we consider the latter problem.

For a bounded region $\Omega \subset \mathbb{R}^m$ and $0 < t < T$, consider the heat equation

$$(5.1) \quad \dot{u} = \Delta u \quad \text{with } u|_{\partial\Omega} = 0; \quad u|_{t=0} = x.$$

We assume that $x \in L^2(\Omega)$ is unknown but that we have (as a given constraint) the fixed bound $\|x\| \leq M$. We have observation/measurement of $y := u|_{t=1}$ and seek $z := u|_{\tau}$ for some fixed $\tau \in (0, 1)$; This is the ‘backward heat equation’.¹³ It has long been known [13] that this classical ill-posed problem is stabilized by the presence of the *a priori* bound on x . Our data for computation are then: M , the approximate observation \hat{y} , and a bound ε on the measurement error $\|y - \hat{y}\|$; from these (and the equation itself) we wish to obtain an ‘optimal’ approximation \hat{z} to z . Note that this formulation explicitly omits direct consideration of the unknown x , which appears only as the ‘subject’ of the stabilizing constraint. Without giving details, we note that one can construct, by ‘optimal filtering’, a solution estimate \hat{z} with the guaranteed accuracy

$$(5.2) \quad \|\hat{z} - z\| \leq M^{1-\tau} \varepsilon^{\tau}.$$

It can also be shown that this estimate is optimal in the sense that one can find examples (for a sequence $\varepsilon \rightarrow 0$) for which the region of uncertainty contains points z, z' with twice this separation. Certainly (5.2) gives an asymptotic convergence rate: the accuracy of the computed solution is $\mathcal{O}(\varepsilon^{\tau})$

¹³We recall a talk on this problem in the 50’s by F. John under the title, “*Recherche de la temperature perdu*”.

as $\varepsilon \rightarrow 0$ but, of course, it also provides an assured estimate (in terms of M, ε) for any fixed level of measurement accuracy.

Remark: Suppose the actual value of M were *not* known but we take $\hat{M} = \hat{M}(\varepsilon)$ with $\hat{M} \rightarrow \infty$ as $\varepsilon \rightarrow 0$. If we apply the method above replacing M by \hat{M} in the computation, then the bound $\|x\| \leq \hat{M}$ is certainly valid for ε small enough so (5.2) is then true! This gives, asymptotically, *almost* as good a convergence rate as if M were known since one can take the growth of $\hat{M}(\cdot)$ to be arbitrarily slow. BUT: this is entirely useless *for any fixed* ε since, without the information about the ‘true’ M , one cannot *know* when ε is actually ‘small enough’ to validate the estimate.¹⁴ ■

Finally, we note that the earlier view of model uncertainty applies only at a particular stage of analysis, corresponding to a particular level of precision, since *no* model, as used, is ever really exact — models ‘as used’ always *necessarily* involve simplifications. For asymptotic analysis, however, the model uncertainty would also have to disappear as the measurement error is made to disappear so one would be forced to deal with a succession of increasingly accurate and comprehensive models; in principle one would need an exact ‘limit model’. The history of science assures us that we may hope to have such a succession but it is far from clear that one could anticipate any such thing as an exact limit model. Indeed, the very context of analysis may disappear in such a limit — e.g., at increasingly smaller scales continuum mechanics becomes statistical mechanics which, in turn, becomes quantum mechanics. As another example of this, consider the old question of determining precisely (*sic!*) the ‘total length of the coastline of the United States’ — a figure which continues to increase unboundedly as one insists on refining the measurement by tracking the ins and outs of first every bay, then every cove, then every clod of dirt, . . .

Note that we asserted, just above, that ill-posedness resides in the asymptotics and now claim that the nature of the real world does not admit of any

¹⁴One might then ask: *How useful is an asymptotic estimate?*

We also note a related computation: if our problem were to approximate x itself — an ill-posed problem even with the bound M — then we could similarly apply the method replacing τ by $\hat{\tau} = \hat{\tau}(\varepsilon)$ with $\hat{\tau} \rightarrow 0$ as $\varepsilon \rightarrow 0$. This actually provides an estimate \hat{z}_ε for $u|_{t=\hat{\tau}}$ and we note that, if this $\hat{\tau}(\cdot)$ is chosen so that $\varepsilon^{\hat{\tau}(\varepsilon)}$ as $\varepsilon \rightarrow 0$, then the estimate (5.2) together with the continuity of u as $t \rightarrow 0$, given by the usual theory for the heat equation, ensure convergence of \hat{z}_ε , although without any convergence *rate* (unless one can bound x in some higher order norm to provide a suitable modulus of continuity for the solution u).

such asymptotics. Perhaps this paradox, too, can be addressed — at a later time!

6. SUMMARY

In summary, we merely list some of our conclusions:

- Ill-posedness is often the result of a *factorization* of the ‘genuine’ problem.
- Our successes (counterargument **1**’ of the Introduction) do not really contradict the negative arguments **1**, **1+** asserted in [10], [33]. The occasions when we have ‘succeeded’ with an ill-posed problem seem only to occur when this has appeared as merely one component of a genuine problem which, altogether, is actually well-posed.
- Any useful analysis of the complexity of such problems must emphasize the *context*: it is the problem *as a whole* which gives the appropriate notion of ‘complexity’ as a balance of costs.
- It is useful for us to look at related modes of analysis (e.g., adaptive control theory and Statistics) for related paradigms which may be ‘translated’ or adapted.
- The model is part of the data, to be used for computation as well as for the analysis.
- Ill-posedness resides in the asymptotics but one must be very careful in the interpretation and application of asymptotic analyses.

REFERENCES

- [1] R.S. Anderssen, *The linear functional strategy for improperly posed problems*, in (J.R. Cannon and U. Horning, eds.) *Inverse Problems* [ISNM #77], Birkhäuser, Basel, 1987, pp. 11–30.
- [2] G.E. Backus and J.F. Gilbert, *Numerical applications of a formalism for geophysical inverse problems*, Geophys. J. R. Astr. Soc. **13** (1967), pp. 247–276; *The resolving power of gross Earth data*, *ibid.* **16** (1968), pp. 169–205.
- [3] A. Becker, P.R. Kumar, and C.Z. Wei, *Adaptive control with the*

stochastic approximation algorithm: geometry and convergence, IEEE Trans. Automatic Control **AC-30** (1985), pp. 330–338.

[4] G. Chen, *On the information-based complexity of optimal reconstruction for nonlinear systems*, Rice Univ. (Dept. Elect./Comput. Eng.) **TR 8905**, 1989.

[5] W.C. Chewning and T.I. Seidman, *A convergent scheme for boundary controls for the heat equation*, SIAM J. Control Opt. **15** (1977), pp. 64–72.

[6] L. Eldén and T.I. Seidman, *An ‘optimal filtering’ method for the side-ways heat equation*, in preparation.

[7] H. Engl and A. Neubauer, *Optimal discrepancy principle for the Tikhonov regularization of integral equations of the first kind*, in *Constructive Methods for the Practical Treatment of Integral Equations*, (ISNM #73), Birkhäuser, Basel, 1988.

[8] R. Ewing and T.I. Seidman, *An ‘optimal filtering’ method for the backwards heat equation*, in preparation.

[9] G. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), pp. 215–223.

[10] J. Hadamard, *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale Univ. Press, New Haven, 1923.

[11] A.J. Helmicki, C.A. Jacobson, and C.N. Nett, *Identification in H_∞ : a robustly convergent nonlinear algorithm*, preprint, Northeastern Univ., 1989.

[12] B. Hofmann, *Regularization for Applied Inverse and Ill-Posed Problems*, Teubner, Leipzig, 1986.

[13] F. John, *Continuous dependence on data for solutions of partial dif-*

ferential equations with a prescribed bound, Comm. Pure Appl. Math. **13** (1960), pp. 551–585.

[14] L.K. Jones, *Approximation-theoretic derivation of logarithmic entropy principles for inverse problems and unique extension of the maximum entropy method to incorporate prior knowledge*, SIAM J. Appl. Math. **49** (1989), pp. 650–661.

[15] I. Kant, *Prolegomena to all future Metaphysics*, 1783.

[16] A. Kirsch, *Inverse problems*, in *Trends in Mathematical Optimization* (ISNM #84), pp. 117–137, Birkhäuser, Basel, 1988.

[17] M.A. Kon, review of [31], Bull. Amer. Math. Soc. **21** (1989), pp. 332–339.

[18] H.A. Levine, *Continuous data dependence, regularization, and a three lines theorem for the heat equation with data in a space like direction*, Ann. Mat. Pura Appl. **CXXXIV** (1983), pp. 267–286.

[19] C.A. Micchelli and T.J. Rivlin, *A survey of optimal recovery*, in *Optimal Estimation in Approximation Theory*, (C.A. Micchelli and T.J. Rivlin, eds.) pp. 1–54, Plenum, New York, 1977.

[20] A. Neubauer, *Tikhonov regularization for the solution of nonlinear ill-posed problems: optimal convergence rates and finite-dimensional approximation*, Inverse Problems **5** (1989), pp. 541–557.

[21] F. O’Sullivan, *A statistical perspective on ill-posed inverse problems*, Statistical Sci. **1** (1986), pp. 502–527.

[22] R.L. Parker, *Understanding inverse theory*, Ann. Rev. Planetary Sci. (1977), pp. 35–64.

[23] W.W. Piegorsch and G. Casella, *The early use of matrix diagonal increments in Statistical problems*, SIAM Review **31** (1989), pp. 428–434.

- [24] T.I. Seidman, *The solution of singular equations, I: linear equations in Hilbert space*, Pac. J. Math. **61** (1975), pp. 513–520.
- [25] T.I. Seidman, *Computational approaches to ill-posed problems: general considerations*, in *Proc. 1976 Conf. on Inf. Sci., Systems*, pp. 258–262, Johns Hopkins Univ., 1976.
- [26] T.I. Seidman, *Convergent approximation methods for ill-posed problems, Part I: general theory*, Control and Cybernetics **10** (1981), pp. 31–49.
- [27] T.I. Seidman, *The method of ‘optimal filtering’ for a class of ill-posed problems*, in preparation.
- [28] T.I. Seidman and M.J. Munteanu, *Approaches to geophysical inverse problems: a methodological comparison*, Goddard Space Flight Center NASA TM **80301**, 1979.
- [29] A.M. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Sov. Math. Doklady **4**, (1963), pp. 1036–1038.
- [30] A.M. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington, 1977.
- [31] J.F. Traub, G.W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, NY, 1988.
- [32] G. Wahba, *Ill-posed problems: numerical and statistical methods for mildly, moderately, and severely ill-posed problems with noisy data*, Dept. Stat. **TR 595**, Univ. Wisconsin–Madison, 1980 [to have appeared in *Proc. Int. Conf. on Ill-Posed Problems (U. Delaware, 1979)*, unpublished].
- [33] A.G. Werschulz, *What is the complexity of ill-posed problems?* Numer. Funct. Anal. and Optim. **9**(1987), pp. 945–967.

- [34] A.G. Werschulz, *An information-based approach to ill-posed problems*, J. Complexity **3** (1987), pp. 270–301.
- [35] A.G. Werschulz, *Optimal residual algorithms for linear operator equations*, Theoretical Computer Science (China) (1989), to appear.
- [36] E.P. Wigner, *On the unreasonable effectiveness of Mathematics in the natural sciences*, Comm. Pure Appl. Math.