INFORMATION SCIENCES **57–58**, 159–169 (1991)

A Unified Approach to Dynamic Estimation*

ROBERT KALABA

Departments of Electrical and Biomedical Engineering, University of Southern California, Los Angeles, California 90089

and

LEIGH TESFATSION

Departments of Economics and Mathematics, Iowa State University, Ames, Iowa 50011-1070

ABSTRACT

Discrepancies between assumed dynamical models and observations are often handled by making further probabilistic assumptions, a tactic which has both strengths and weaknesses. A re-examination of filtering and smoothing is conducted, and an alternative multicriteria approach, which is probability free, is advanced. This approach involves vector minimization as a key ingredient, and it specializes to the well-known Kalman, Viterbi, Larson-Peschon, and Swerling filters.

1. INTRODUCTION

Since World War II, probabilistic methods have held a dominant position in filtering and smoothing theory [1]. These methods, leading to likelihood and posterior distribution functions, have the great advantage that they produce scalar measures of theory and data incompatibility. Recently, various other methods for incorporating disparate sources of information into a single scalar incompatibility measure have attracted increasing attention, e.g., Bellman and Zadeh's fuzzy set approach [2] and Salukvadze's ideal point theory [11].

For many processes, however, model discrepancies arise from conceptually distinct sources, e.g., imperfect measurement devices versus mis-spec-

* Partially supported by NIH grant DK 33729. Please address correspondence to Leigh Tesfatsion.

655 Avenue of the Americas, New York, NY 10010 0020-0255/91/$03.50

ified dynamic laws of motion. It therefore can be difficult to achieve a scalarization of the incompatibility measure in a publicly credible way. In decision theory, this type of incommensurability is handled by multicriteria optimization techniques [9]; but, to date, such techniques have not been exploited systematically in state estimation theory.

In this paper we present a multicriteria framework for dynamic state estimation which encompasses a wide range of views concerning the appropriate specification of incompatibility measures. If available, probability assessments can be used to provide a single scalar measure of incompatibility, as illustrated by the well-known Kalman [8], Viterbi [3, 14], Larson-Peschon [10], and Swerling [12] filters. Alternatively, disparate sources of information can be systematically considered without forced scalarization, as illustrated by the "flexible least squares" approach [4, 5, 6, 7].

The following two sections use illustrative examples to compare and contrast the standard scalar-criterion approach to state estimation with the multicriteria approach. Section 2 discusses the standard approach to state estimation for a time-varying linear system in which probability relations for discrepancy terms are used to obtain a scalar measure of theory and data incompatibility, namely, a posterior probability density function for the sequence of state vectors. Section 3 discusses an alternative multicriteria approach to this problem which could be used by a data analyst who is either unable or unwilling to provide probability assessments for discrepancy terms, at least in the preliminary stages of his study. A multicriteria framework for more generally specified state estimation problems is outlined in Section 4 and concluding comments are given in Section 5.

2. *STANDARD TREATMENT OF A STATE ESTIMATION PROBLEM*

Suppose scalar observations $y_1, \ldots, y_T$ obtained on a process are postulated to be linearly related to a sequence of state vectors $x_1, \ldots, x_T$, as follows:

Measurement Relations:

$$y_t = h_t' x_t + v_t, \qquad t = 1, \ldots, T, \tag{1}$$

where

$h_t' = (h_{t1}, \ldots, h_{tN}) = 1 \times N$ row vector of known state coefficients;

$x_t = (x_{t1}, \ldots, x_{tN})' = N \times 1$ column vector of unknown state variables;

$v_t =$ a scalar measurement discrepancy term.

If no restrictions are placed on the discrepancy term v_t, then equation (1) is simply a defining relation for v_t. That is, v_t is a slack variable, and equation (1) is true by definition whether or not a linear relation between y_t and x_t exists in actuality. In particular, the equality sign in equation (1) really does mean equality in the usual exact mathematical sense. Introduced in this way, there is nothing controversial about v_t.

What will v_t depend on? Everything affecting y_t which is not captured by the term $h_t'x_t$, i.e., everything unknown, or not presumed to be known, about how y_t might depend on higher order terms in x_t, on missing variables, and so forth.

Suppose, in addition to (1), that the state vector x_t is assumed to evolve over time as follows:

Dynamic Relations:

$$x_{t+1} = x_t + w_t, \qquad t = 1, \ldots, T\text{-}1, \tag{2}$$

where

$$w_t = \text{a dynamic discrepancy term.}$$

As before, if no restrictions are placed on the discrepancy term w_t, then equation (2) defines w_t to be a slack variable incorporating everything unknown, or not presumed to be known, about how the state vector x_{t+1} depends on x_t and on missing variables. Equation (2) is thus true regardless of the actual relation between x_{t+1} and x_t, and the equality sign in equation (2) again means equality in the usual exact mathematical sense.

If no additional theoretical relations are introduced at this point, the problem of estimating the state vectors x_t would seem intrinsically to be a *multicriteria* optimization problem. Each possible estimate for the state sequence $(x_1, \ldots, x_T)$ entails two conceptually distinct apple-and-orange types of discrepancy terms—measurement and dynamic—and a data analyst undertaking this estimation would presumably want *each* type of discrepancy to be small.

However, standard state estimation techniques invariably introduce a third type of theoretical relation in addition to (1) and (2): namely, probability relations governing the discrepancy terms v_t and w_t and the initial state vector x_1. Consider, for example, the following commonly assumed relations:

Probability Relations:

$$[\text{PDF for } v_t] = P(v_t),\ t = 1, \ldots, T; \tag{3a}$$

$$[\text{PDF for } w_t] = P(w_t),\ t = 1, \ldots, T\text{-}1; \tag{3b}$$

$$(v_t) \text{ and } (w_t) \text{ mutually and serially independent processes;} \tag{3c}$$

$$[\text{PDF for } x_1] = P(x_1); \tag{3d}$$

$$x_1 \text{ distributed independently of } v_t \text{ and } w_t \text{ for each } t. \tag{3e}$$

Under relations (3), the discrepancy terms v_t and w_t are interpreted as random quantities with known probability density functions (PDF's) governing both their individual and joint behavior. The equality signs in (1) and (2) are still interpreted as equalities in the usual exact mathematical sense, hence v_t and w_t now appear in (1) and (2) as commensurable disturbance terms impinging on correctly specified theoretical relations. The previous interpretation for v_t and w_t as conceptually distinct apple-and-orange discrepancy terms incorporating everything unknown about the measurement and dynamic aspects of the process is thus dramatically altered.

Specifically, combining the measurement relations (1) with the probability relations (3) permits the derivation of a probability density function $P(Y_T \mid X_T)$ for the observation sequence $Y_T = (y_1, \ldots, y_T)$ conditional on the state sequence $X_T = (x_1, \ldots, x_T)$. Combining the dynamic relations (2) with the probability relations (3) permits the derivation of a "prior" probability density function $P(X_T)$ for X_T. The multiplication of these two derived probability density functions then yields the joint probability density function for X_T and Y_T,

$$P(Y_T \mid X_T) \cdot P(X_T) = P(X_T, Y_T). \tag{4}$$

The joint probability density function (4) elegantly combines the two distinct sources of theory and data incompatibility—measurement and dynamic—into a single *scalar* measure of incompatibility for any considered state sequence X_T.

The usual objective assumed for problem (1) through (3) is to determine the state sequence X_T which maximizes the posterior probability density function $P(X_T \mid Y_T)$. Since the observation sequence Y_T is assumed to be given, this objective is equivalent to determining the state sequence X_T which maximizes the product of $P(X_T \mid Y_T)$ and $P(Y_T)$. By the agreed-upon rules of probability theory,

$$P(X_T \mid Y_T) \cdot P(Y_T) = P(Y_T \mid X_T) \cdot p(X_T), \tag{5}$$

where, as earlier explained, the right-hand expression in (5) can be evaluated using relations (1), (2), and (3). Determining the state sequence with greatest posterior probability is thus equivalent to determining the state sequence which minimizes the scalar incompatibility cost function

$$c(X_T, Y_T, T) = -\log[P(Y_T \mid X_T) \cdot P(X_T)]. \tag{6}$$

In summary, what ultimately has been accomplished by the augmentation of the measurement and dynamic relations (1) and (2) with the probability relations (3)? *The multicriteria problem of achieving vector-minimal incompatibility between imperfectly specified theoretical relations and process observations has been transformed into the scalar optimization problem of determining the most probable state sequence for a stochastic model assumed to be correctly and completely specified.*

This by now conventional series of modeling steps would not be open to question if it were any easy task to specify probabilistic properties for the discrepancy terms v_t and w_t in a credible manner. However, for many applications—particularly in the fields of economics and biomedical engineering—this is not the case. For example, the observations $y_1, \ldots, y_T$ may be the outcome of a nonreplicable experiment, so that agreement among data analysts concerning probabilistic properties for the discrepancy terms is difficult to achieve. Alternatively, the theoretical relations (1) and (2) may represent tentatively held conjectures concerning a poorly understood process, or a linearized set of relations obtained for an analytically intractable nonlinear process. In this case it is questionable whether the discrepancy terms are governed by any well-defined probability relationships. A data analyst may then have to resort to specifications determined largely by convention if he is forced to provide a probabilistic characterization for the discrepancy terms.

How might a data analyst determine the degree to which the theoretical relations (1) and (2) are incompatible with the observations $y_1, \ldots, y_T$ when he is either unable or unwilling to provide a probabilistic characterization for the discrepancy terms v_t and w_t?

The next section illustrates what might be done.

3. A MULTICRITERIA APPROACH

Suppose scalar observations $y_1, \ldots, y_T$ have been obtained on a process which is not yet well understood. The following linear relation is postulated between the observation y_t and an $N \times 1$ vector x_t of unknown state variables at each time t:

Measurement Relations [Approximately Linear Measurement]

$$[y_t - h_t' x_t] \approx 0, \quad t = 1, \ldots, T, \tag{7a}$$

where h_t' is a given $1 \times N$ vector of coefficients. It is recognized that some systematic time-variation in the state vectors x_t might have occurred over the observation period. However, it is anticipated that any such evolution will have been gradual, so that successive state vectors do not differ too much from one observation time to the next.

Dynamic Relations [Slowly Evolving State Vector]

$$[x_{t+1} - x_t] \approx 0, \quad t = 1, \ldots, T\text{-}1. \tag{7b}$$

The problem of filtering and smoothing is to try to determine the state sequence estimates which are in some sense minimally incompatible with given theoretical relations, conditional on a given set of observations. This problem is essentially a *multicriteria* optimization problem which will be presented in general terms in Section 4.

For the case at hand, the multicriteria nature of the filtering and smoothing problem is seen as follows. Two conceptually distinct types of model specification error can be associated with each possible state sequence estimate $\hat{X}_T = (\hat{x}_1, \ldots, \hat{x}_T)$. First, the choice of $\hat{X}_T$ could result in measurement specification errors consisting of non-zero discrepancy terms $[y_t - h_t'\hat{x}_t]$ in (7A). Second, the choice of $\hat{X}_T$ could result in dynamic specification errors consisting of non-zero discrepancy terms $[\hat{x}_{t+1} - \hat{x}_t]$ in (7b). In order to conclude that the theoretical relations (7) are in reasonable agreement with the observations, *each* type of discrepancy would have to be small.

Suppose a measurement cost $c_M(\hat{X}_T, Y_T, T)$ and a dynamic cost $c_D(\hat{X}_T, Y_T, T)$ are separately assessed for the two disparate types of model specification errors entailed by the choice of a state sequence estimate $\hat{X}_T$. On the basis of both tractability and general intuitive appeal, these costs are taken to be sums of squared discrepancy terms. More precisely, for any given state sequence estimate $\hat{X}_T$, the measurement cost associated with $\hat{X}_T$ is taken to be

$$c_M(\hat{X}_T, Y_T, T) = \sum_{t=1}^{T} [y_t - h_t'\hat{x}_t]^2 \tag{8}$$

and the dynamic cost associated with $\hat{X}_T$ is taken to be

$$c_D(\hat{X}_T, Y_T, T) = \sum_{t=1}^{T-1} [\hat{x}_{t+1} - \hat{x}_t]'D[\hat{x}_{t+1} - \hat{x}_t], \tag{9}$$

where D is a suitably selected positive definite scaling matrix.[1]

If the prior beliefs (7) concerning the measurement and dynamic relations are absolutely true, then the actual state sequence X_T would result in zero

[1] The scaling matrix D can be specified so that the "FLS estimates" obtained below for the state vectors x_t are essentially invariant to the choice of units for the components of the coefficient vectors h_t. See Tesfatsion and Veitch [13, Footnote 3].

values for both c_M and c_D. In any real-world application, we would of course expect to see positive measurement and dynamic costs associated with each potential state sequence estimate $\hat{X}_T$. Nevertheless, not all of these estimates are equally interesting. Specifically, we would not be interested in a state sequence estimate $\hat{X}_T$ if it were cost-subordinated by another estimate $\hat{X}_T^*$ in the sense that $\hat{X}_T^*$ yielded a lower value for one type of cost without increasing the value of the other.

We therefore focus attention on the set of state sequence estimates which are not cost-subordinated by any other state sequence estimate. Such estimates are referred to as *flexible least squares (FLS) estimates*. Each FLS estimate shows how the state vector could have evolved over time in a manner minimally incompatible with the prior measurement and dynamic relations (7). Without additional prior information or additional modeling criteria, restricting attention to any proper subset of the FLS estimates is a purely arbitrary decision. Consequently, the FLS approach envisions the generation and consideration of all of the FLS estimates in order to determine the commonalities and divergencies displayed by these potential state sequences.

Define the *cost possibility set* to be the collection

$$C(T) = \{c_D(\hat{X}_T, Y_T, T), c_M(\hat{X}_T, Y_T, T) \mid \hat{X} \in E^{TN}\} \tag{10}$$

of all possible configurations of dynamic and measurement costs attainable at time T, conditional on the given observation sequence Y_T. The *cost-efficient frontier* C^F (T) is then defined to be the collection of all cost vectors $c = (c_D, c_M)$ in $C(T)$ which are not subordinated by any other cost vector c^* in $C(T)$ in the sense that $c^* \leq c$. Formally, letting "vmin" denote vector-minimization,

$$C^F(T) = \text{vmin } C(T). \tag{11}$$

By construction, then, the cost-efficient frontier is the collection of all cost vectors associated with the FLS state sequence estimates.

If the $N \times T$ matrix $[h_1, \ldots, h_T]$ has full rank N, the cost-efficient frontier $C^F(T)$ is a strictly convex curve in the c_D–c_M plane giving the locus of vector-minimal costs attainable at time T, conditional on the given observations. In particular, $C^F(T)$ reveals the measurement cost c_M that must be paid in order to achieve the zero dynamic cost (time-constant state vector estimates) required by OLS estimation. [See Figure 1.]

Once the FLS estimates and the cost-efficient frontier are determined, three different levels of analysis can be used to investigate the incompatibility of the theoretical relations (7) and the observations $y_1, \ldots, y_T$. First, the frontier can be examined to determine the efficient attainable trade-offs be-

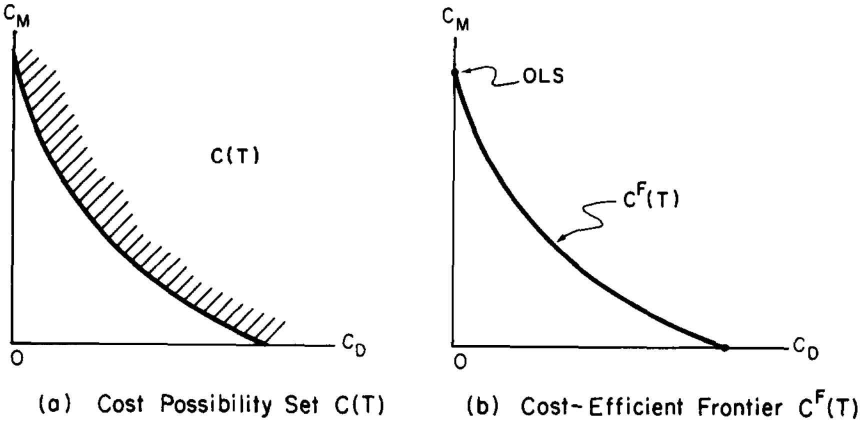

Fig. 1(a). Cost possibility set $C(T)$
Fig. 1(b). Cost-efficient frontier $C^F(T)$

tween the measurement and dynamic costs c_M and c_D. Second, descriptive summary statistics, e.g., average value and standard deviation, can be constructed for the time-paths traced out by the FLS estimates at each point along the frontier. These summary statistics provide rough indicators of the extent to which the FLS estimates deviate from the OLS solution associated with the extreme point of the frontier where dynamic cost is zero. Finally, the time-paths traced out by the FLS estimates can be directly examined for evidence of systematic movements in individual state variables, e.g., unanticipated jumps at dispersed points in time. These movements might be difficult to discern from summary statistical characterizations of the time-paths.

A detailed theoretical discussion of the FLS technique is given in [4–6]. A Fortran program for generating the FLS estimates is provided in Kalaba and Tesfatsion [5]; and simulation experiments demonstrating the ability of the FLS estimates to track linear, quadratic, sinusoidal, and regime shift motions in the true state variables, despite noisy observations, are reported and graphically depicted. In Tefatsion and Veitch [13], the FLS technique is used to undertake an empirical investigation of a well-known log-linear regression model for U.S. money demand over the volatile period 1959:Q2–1985:Q3. Interesting insights are obtained concerning shifts in the money demand relation at economically reasonable points in time.

4. GENERALIZATIONS

In the previous section it is shown how a multicriteria approach might be used to investigate the basic incompatibility of theory and data for one type of filtering and smoothing problem. This multicriteria approach is generalized in Kalaba and Tesfatsion [7] to a much broader class of problems. The present section briefly reviews this work.

Consider a situation in which a sequence of observations $Y_T = (y_1, \ldots, y_T)$ has been obtained on a process over time periods $1, \ldots, T$. The basic problem is to learn about the sequence of states $X_T = (x_1, \ldots, x_T)$ through which the process has passed.

Suppose the degree to which each possible state sequence estimate X_T is incompatible with the given observation sequence Y_T is measured by a K-dimensional vector $c(\hat{X}_T, Y_T, T)$ of incompatibility costs. These costs may represent penalties imposed for failure to satisfy criteria *conjectured* to be true (theoretical relations), and also penalties imposed for failure to satisfy criteria *preferred* to be true (objectives). Let $C(T)$ denote the set of all incompatibility cost vectors $c = c(\hat{X}_T, Y_T, T)$ corresponding to possible state sequence estimates $\hat{X}_T$. The *cost-efficient frontier*, denoted by $C^F(T)$, is then defined to be the collection of cost vectors c in $C(T)$ which are not subordinated by any other cost vector c^* in $C(T)$ in the sense that $c^* \leq c$.

By construction, the state sequence estimates $\hat{X}_T$ whose cost vectors attain the cost-efficient frontier are characterized by a basic efficiency property: for the given observations, no other possible state sequence estimate yields lower incompatibility cost with respect to each of the K modeling criteria included in the incompatibility cost vector. Each of these state sequence estimates thus represents one possible way the actual process could have evolved over time in a manner minimally incompatible with the prior theoretical relations and objectives.

The basic multicriteria estimation problem can be summarized as follows:

The Basic Multicriteria Estimation Problem

Given a process length T, an observation sequence Y_T, and a multidimensional incompatibility cost function $c(\cdot, Y_T, T)$, determine all possible state sequence estimates $\hat{X}_T$ which vector-minimize the incompatibility cost $c(\hat{X}_T, Y_T, T)$. That is, determine all possible state sequence estimates $\hat{X}_T$ whose cost vectors $c(\hat{X}_T, Y_T, T)$ attain the cost-efficient frontier $C^F(T)$.

A vector-valued recurrence relation is established for the cost-efficient frontier in Kalaba and Tesfatsion [7]. This recurrence relation is readily recognizable as a multicriteria extension of the usual scalar dynamic programming equations.

To give a rough idea of this result, consider the estimation problem at any intermediate time t. Suppose a K-dimension vector $c(\hat{X}_t, Y_t, t)$ of incompatibility costs can be associated with each t-length state sequence estimate $\hat{X}_t = (\hat{x}_1, \ldots, \hat{x}_t)$, conditional on the sequence of observations $Y_t = (y_1, \ldots, y_t)$. Let $C(\hat{x}_t, t)$ denote the set of all cost vectors $c(\hat{X}_t, Y_t, t)$ attainable at time t, conditional on the time-t state estimate being $\hat{x}_t$; and let $C^F(\hat{x}_t, t)$ denote the cost-efficient frontier for $C(\hat{x}_t\, t)$.

Given certain regularity conditions, it is shown that state-conditional frontiers at time t are mapped into state-conditional frontiers at time $t + 1$ in accordance with a vector-valued recurrence relation having the form

$$C^F(\hat{x}_{t+1}, t + 1) = \text{vmin}\ (\cup_{\hat{x}_t}[C^F(\hat{x}_t, t) + \delta c(\hat{x}_t, \hat{x}_{t+1}, y_{t+1}, t + 1)]), \quad (12)$$

where "vmin" denotes vector-minimization and $\delta c(\cdot)$ denotes a vector of incremental costs associated with the state transition $(\hat{x}_t, \hat{x}_{t+1})$. Moreover, the cost-efficient frontier at the final time T satisfies

$$C^F(T) = \text{vmin}\ [\cup_{\hat{x}_T} C^F(\hat{x}_T, T)]. \quad (13)$$

Three well-known state estimation algorithms are derived in Kalaba and Tesfatsion [7] as scalar-criterion special cases of the multicriteria recurrence relations (12) and (13): namely, the Kalman [8], Viterbi [3, 14], and Larson-Peschon [10] filters for sequentially generating maximum a *posteriori* probability estimates. In addition, an algorithm for sequentially generating the FLS estimates for the problem discussed above in Section 3 is derived as a bicriteria special case of (12) and (13).

5. CONCLUDING REMARKS

The specification of appropriate criteria for measuring the incompatibility of theory and data is a key issue for state estimation. The general multicriteria framework outlined in Section 4 provides an organizing principle for state estimation which accommodates a broad range of perspectives on this issue. If available, probability assessments can be used to provide a single scalar measure of incompatibility, as illustrated in Section 2. Alternatively, disparate sources of information can be systematically considered without forced scalarization, as illustrated in Section 3.

Future studies will stress both theoretical developments and practical applications. For example, the state sequence estimates whose costs attain the cost-efficient frontier constitute a "population" of estimates characterized by a basic efficiency property: For the given observations, these are the state sequence estimates which are minimally incompatible with the prior theoretical relations and objectives. Systematic procedures need to be developed for interpreting and reporting the behavior displayed by these estimates in both simulation and empirical studies. A second related issue concerns the use of the posterior information embodied in the cost-efficient frontier for adaptive model respecification.

For many years filtering and smoothing studies have primarily dealt with situations where theoretical specifications are essentially correct and model discrepancy terms are reasonably modeled as random quantities with known distributions. More recently, however, the social and biological sciences have presented filtering and smoothing problems of critical importance for which the underlying relations are not well understood. In such areas, model misspecification is an endemic problem, and procedures are needed for coping with this reality. As suggested in this paper, the explicit recognition of model specification errors raises a number of new and interesting challenges for filtering and smoothing theory.

REFERENCES

1. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.
2. R. Bellman and L. Zadeh, Decision-making in a fuzzy environment, *Management Science* 17:141–164 (1970).
3. G. D. Forney, Jr., The Viterbi algorithm, *Proceedings of the IEEE*, March:268–278 (1973).
4. R. Kalaba and L. Tesfatsion, Sequential nonlinear estimation with nonaugmented priors, *Journal of Optimization Theory and Applications* 60:421–438 (1989).
5. R. Kalaba and L. Tesfatsion, Time-varying linear regression via flexible least squares, *International Journal of Computers and Mathematics with Applications* 17:1215–1245 (1989).
6. R. Kalaba and L. Tesfatsion, Flexible least squares for approximately linear systems, *IEEE Transactions on Systems, Man, and Cybernetics* 20: (1990).
7. R. Kalaba and L. Tesfatsion, An organizing principle for dynamic estimation, *Journal of Optimization Theory and Applications* 64:445–470 (1990).
8. R. E. Kalman, A new approach to linear filtering and prediction problems, *Transactions of the ASME: Journal of Basic Engineering* 82:35–45 (1960).
9. N. T. Koussoulas, Multiobjective optimization in adaptive and stochastic control, in *Control and Dynamic Systems* 25 (C. T. Leondes, Ed.), Academic Press, 1987, pp. 55–78.
10. R. E. Larson and J. Peschon, A dynamic programming approach to trajectory estimation, *IEEE Transactions on Automatic Control* 11:537–540 (1966).
11. M. Salukvadze, *Vector-valued Optimization Problems in Control Theory*, Academic Press, 1979.
12. P. Swerling, First order error propagation in a stagewise smoothing procedure for satellite observations, *Journal of the Astronautical Sciences* VI:46–52 (1959).
13. L. Tesfatsion and J. Veitch, U.S. money demand instability: a flexible least squares approach, *Journal of Economic Dynamics and Control* 14:151–173 (1990).
14. A. J. Viterbi, Error bounds of convolutional codes and an asymptotically optimal decoding algorithm, *IEEE Transactions on Information Theory* 13:260–269 (1967).

Received 18 September 1989; revised 24 July 1990