Flexible Least Squares for Approximately Linear Systems

ROBERT KALABA AND LEIGH TESFATSION

***Abstract*—The problem of filtering and smoothing for a system described by approximately linear dynamic and measurement relations has been studied for many decades. Yet the potential problem of misspecified dynamics, which makes the usual probabilistic assumptions involving normality and independence questionable at best, has not received the attention it merits. A probability-free multicriteria "flexible least squares" filter that meets this misspecification problem head on is proposed. A Fortran program implementation is provided for this filter, and references to simulation and empirical results are given. Although there are close connections with the standard Kalman filter, there are also important conceptual and computational distinctions. The Kalman filter, relying on probability assumptions for model discrepancy terms, provides a unique estimate for the state sequence. In contrast, the flexible least squares filter provides a family of state sequence estimates, each of which is vector-minimally incompatible with the prior dynamical and measurement specifications.**

I. Introduction

FOLLOWING World War II, probabilistic methods attained a dominant position in filtering and smoothing theory [1]. Early studies focused on linear system identification problems arising in radar and communications for which the theoretical specifications were essentially correct, with for which model discrepancy terms were reasonably modeled as random quantities with known distributions. For such problems, probabilistic methods could credibly be used to construct scalar measures for theory and data incompatibility in the form of likelihood or posterior distribution functions.

More recently, however, the social and biological sciences have presented filtering and smoothing problems of critical importance for which the processes of interest are highly nonlinear and poorly understood. In attempting to apply standard filtering and smoothing techniques to such a problem, a data analyst typically has to replace the unknown nonlinear process relations with an approximate system of linear relations. The resulting model discrepancy terms then incorporate model specification errors from various conceptually distinct sources—e.g., imperfectly specified measurements versus imperfectly specified state dynamics; hence it is questionable whether these discrepancy terms are either jointly or separately governed by meaningful probability relations. More generally, it is difficult to provide any credible way to scale and weigh the discrepancy terms relative to one another.

In decision theory, incommensurability of this type is typically handled by multicriteria optimization techniques [2]. However, such techniques have not yet been exploited systematically in state estimation theory. Rather, currently available filtering and smoothing techniques *require* the data analyst to provide probability assessments for all discrepancy terms. In consequence, social and biological scientists attempting to apply these techniques are often forced to resort to conventional probability specifications such as normality and independence that may have little public credibility.

This paper proposes a probability-free multicriteria filter for the estimation of approximately linear dynamical systems. Briefly stated, this "flexible least squares" (FLS) filter solves the following multicriteria optimization problem: Characterize the set of all state sequence estimates which achieve vector-minimal incompatibility between imperfectly specified linear theoretical relations and process observations.

The FLS filtering and smoothing problem for approximately linear dynamical systems is set out in Section II. The FLS recurrence relations for the solution of this problem are derived in Section III. Section IV considers the relationship between FLS and Kalman filtering. Concluding remarks are given in Section V. A Fortran program *GFLS* which implements the FLS recurrence relations for this application is provided in an appendix.

II. The Basic Problem

Consider a system whose state at time t, $t=1,2,\cdots$, is an n-dimensional vector x_t. It is believed that the state transition equations for the system take the approximately linear form

$$x_{t+1} \approx F(t)x_t + a(t), \qquad t=1,2,\cdots, \tag{1}$$

where $F(t)$ is a known $n\times n$ square matrix, and $a(t)$ is a known n-dimensional column vector. At each time t, an m-dimensional vector y_t of observations is obtained. The measurement relations are assumed to take the approximately linear form

$$y_t \approx H(t)x_t + b(t), \qquad t=1,2,\cdots, \tag{2}$$

Manuscript received November 6, 1989; revised April 1, 1990. This work was supported in part by NIH Grant #DK 33729.

R. Kalaba is with the Departments of Biomedical and Electrical Engineering, University of Southern California, Los Angeles, CA 90089-1451.

L. Tesfatsion is with the Department of Economics, University of Southern California, Los Angeles, CA 90089-0253.

IEEE Log Number 9036361.

0018-9472/90/0900-0978$01.00

where $H(t)$ is a known $m \times n$ rectangular matrix and $b(t)$ is a known m-dimensional column vector.

Each possible sequence of estimates $\hat{x}_1, \hat{x}_2, \ldots$ for the state vectors entails two conceptually distinct types of model specification errors: namely, measurement errors consisting of the discrepancies $[y_t - H(t)\hat{x}_t - b(t)]$ between the actual and the estimated observation at each time t; and dynamic errors consisting of the discrepancies $[\hat{x}_{t+1} - F(t)\hat{x}_t - a(t)]$ that arise due to misspecification of the state transition equations. The basic filtering and smoothing problem then involves *multicriteria* optimization. Given a sequence of observation vectors $y_1, y_2, \cdots, y_T$ up to time T with $T \geqslant 1$, determine the state sequence estimates $\hat{X}_T = (\hat{x}_1, \cdots, \hat{x}_T)$, which in some sense make both types of specification error as small as possible.

Suppose a dynamic cost $c_D(\hat{X}_T, T)$ and a measurement cost $c_M(\hat{X}_T, T)$ 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 dynamic cost associated with $\hat{X}_T$ is taken to be

$$c_D(\hat{X}_T, T) = \sum_{t=1}^{T-1} [\hat{x}_{t+1} - (F(t)\hat{x}_t + a(t))]' \cdot D(t)[\hat{x}_{t+1} - (F(t)\hat{x}_t + a(t))] \quad (3)$$

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

$$c_M(\hat{X}_T, T) = \sum_{t=1}^{T} [y_t - (H(t)\hat{x}_t + b(t))]' \cdot M(t)[y_t - (H(t)\hat{x}_t + b(t))]. \quad (4)$$

Here $D(t)$ and $M(t)$ are square, symmetric, positive definite scaling matrices of orders n and m, respectively. Having nonzero off-diagonal terms in these matrices would presume knowledge about the relative signs of the discrepancy terms, a presumption that is not very reasonable when discrepancy terms result from model misspecification. Nevertheless, these matrices are left in general form because it does not impede the analytical treatment presented as follows.

If the prior beliefs (1) and (2) concerning the dynamic and measurement relations are absolutely true, then the actual state sequence $X_T = (x_1, \cdots, x_T)$ would result in zero values for both c_D and c_M. In any real-world application, we would of course expect to see positive dynamic and measurement costs associated with each potential state sequence estimate $\hat{X}_T$. Nevertheless, not all of these state sequence 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.

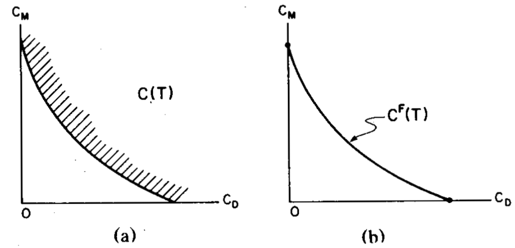

Fig. 1. Trade-offs between dynamic and measurement costs. (a) Cost possibility set. (b) Cost-efficient frontier.

We therefore focus attention on the set of state sequence estimates that 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 dynamic and measurement specifications (1) and (2). Without additional model criteria to augment (1) and (2), 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 commonalities and divergencies displayed by these potential state trajectories.

The collection $C^F(T)$ of cost vectors (c_D, c_M) associated with the FLS estimates is referred to as the *cost-efficient frontier*. Given the cost specifications (3) and (4), the frontier is a downward sloping strictly convex curve in the $c_D - c_M$ plane. (See Fig. 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 (1) and (2) with the observation vectors $y_1, \cdots, y_T$. First, the frontier can be examined to determine the efficient trade-offs between the dynamic and measurement costs c_D and c_M. For example, one can determine the minimum measurement cost that would have to be paid in order to achieve zero dynamic cost, i.e., an exact fit of the state transition equations (1). Second, descriptive summary statistics (e.g., average values and standard deviations) can be constructed for the trajectories traced out by the FLS estimates along the frontier. Finally, the trajectories traced out by the FLS estimates can be directly examined from left to right along the frontier to assess the effects of decreasing the implicit penalty imposed for dynamic versus measurement cost.

Reference [3] applies this three-stage FLS analysis to a time-varying linear regression problem, a special case of (1) and (2) with scalar observations ($m = 1$), no forcing terms, and state transition matrices $F(t)$ set identically equal to the identity matrix. For this application the components of the $1 \times n$ vectors $H(t)$ are interpreted as explanatory variables for the scalar observations y_t, the state vectors x_t are interpreted as coefficient vectors for the "linear regression" relations (2), and the state transition equations (1) with $F(t) \equiv I$ and $a(t) \equiv 0$ are interpreted as smoothness relations governing the evolution of the coefficient vectors over time.

An empirical FLS study of coefficient stability for a well-known log-linear regression model of U.S. money demand over the volatile period 1959–1985 is undertaken in [4]. Interesting insights are obtained concerning shifts in the coefficients at economically reasonable points in time. In [5], the FLS approach is used to develop a new measure of productivity change; the coefficients characterizing the production process are allowed to evolve slowly over time. The new measure compared favorably with more traditional measures when tested for U.S. agricultural data.

How are the cost-efficient frontier and the FLS estimates actually generated? Section III suggests what might be done.

III. The Flexible Least Squares Filter

In view of the strict convexity of the cost-efficient frontier, each point on this frontier solves a problem of the form "minimize c_M subject to $c_D =$ constant." Consequently, each FLS state sequence estimate $\hat{X}_T = (\hat{x}_1, \cdots, \hat{x}_T)$ can be generated as the solution to a problem of the form

$$\min_{X_T} [\mu c_D(X_T, T) + c_M(X_T, T)], \tag{5}$$

where μ is a suitably chosen Lagrange multiplier lying between 0 and $+\infty$. Hereafter the bracketed expression in (5) will be referred to as the *incompatibility cost* associated with X_T, conditional on μ and T. The multiplier μ, multiplied by -1, gives the slope of the cost-efficient frontier at the solution point for (5); thus μ parameterizes the trade-offs attainable between dynamic and measurement cost along the cost-efficient frontier.

The FLS approach envisions the generation of the entire cost-efficient frontier, together with the corresponding FLS state sequence estimates. Numerical experiments (e.g., [3]) have shown that the cost-efficient frontier can be adequately sketched out by solving the minimization problem (5) over a rough grid of μ-points increasing by powers of ten.

How is this minimization to be done? The solution of (5) appears to be a formidable problem. Since each state vector x_t is n-dimensional, the first-order necessary conditions for the solution of (5) constitute a linear two-point boundary value problem in nT scalar unknowns. Fortunately, as will now be shown, problem (5) can be reduced to its proper dimensionality, n, through the use of a dynamic programming technique.

A. The Basic FLS Filter

Let $\mu > 0$ be given. A recursive procedure will now be developed for the exact sequential solution of the incompatibility cost minimization problem (5) as the duration T of the process increases and additional observation vectors are obtained.

Suppose that the time is $T \geqslant 2$. Observation vectors have previously been obtained for times $1, \cdots, T-1$, and a new observation vector y_T has just become available. Any choice of an estimate x_T for the current time-T state vector incurs two costs. First, a measurement cost is incurred if there is a discrepancy between the actual observation vector y_T and the estimated observation vector $[H(t)x_T + b(T)]$. Second, consideration must also be given to the minimum achievable incompatibility cost over the earlier part of the process, conditional on the state estimate for time T being x_T. The time-separability of the cost functions (3) and (4) implies that this latter cost depends only on x_T and the observation vectors through time $T-1$.

Let a function be introduced to represent the minimum incompatibility cost that can be achieved through time $T-1$, conditional on any given time-T state vector x_T:

$$\phi(x_T; \mu, T-1) = \text{the minimum incompatibility cost attainable through choice of } x_1, x_2, \cdots, x_{T-1}, \text{ conditional on the state vector at time } T \text{ being } x_T. \tag{6}$$

The FLS estimate for the time-T state vector, conditional on μ and the observation vectors obtained through time T, is then found by solving the minimization problem

$$\min_{x_T} \{ [y_T - (H(T)x_T + b(T))]'M(T) \cdot [y_T - (H(T)x_T + b(T))] + \phi(x_T; \mu, T-1) \}. \tag{7}$$

Let this FLS estimate be denoted by

$$x_T^{FLS}(\mu, T) = \arg\min_{x_T} \{ \cdots \}. \tag{8}$$

At time T it is necessary to prepare for the appearance of an observation vector at time $T+1$. To do this, one needs to know the cost function $\phi(x_{T+1}; \mu, T)$. This cost function is given by

$$\begin{aligned}\phi(x_{T+1}; \mu, T) = \min_{x_T} \{ & \mu[x_{T+1} - (F(T)x_T + a(T))]' \\ & \cdot D(T)[x_{T+1} - (F(T)x_T + a(T))] \\ & + [y_T - (H(T)x_T + b(T))]' \\ & \cdot M(T)[y_T - (H(T)x_T + b(T))] \\ & + \phi(x_T; \mu, T-1) \}.\end{aligned} \tag{9}$$

The recursive relationship (9) can be given a dynamic programming interpretation. Conditional on any possible state vector x_{T+1} for time $T+1$, the choice of a state estimate x_T for time T incurs three types of cost. First, there is a dynamic cost associated with the estimated state transition from time T to time $T+1$. Second, there is a measurement cost associated with the discrepancy between the estimated and the actual time-T observation vector. And third, there is a minimum achievable incompatibility cost based on everything that is known about the process through time $T-1$, conditional on the time-T state vector being x_T. Selecting x_T to minimize the sum of these three costs yields the minimum achievable incompatibility cost based on everything that is known about the process through time T, conditional on the time-$(T+1)$ state vector being x_{T+1}.

Using (9), the cost functions $\phi(x_2;\mu,1)$, $\phi(x_3;\mu,2),\ldots$ can be determined one after the other. At time T, assume that the function $\phi(x_T;\mu,T-1)$ is known. An observation vector y_T then becomes available, and the function $\phi(x_{T+1};\mu,T)$ can be determined. To start matters off, it is assumed that an initial cost function $\phi(x_1;\mu,0)$ is given. For the particular cost specifications (3) and (4), this initial cost is identically zero. More generally, however, the initial cost could summarize whatever beliefs one has concerning the cost of estimating that the system is in state x_1 at time $T=1$ before an observation vector at time $T=1$ has been received.

The connection between the minimization problems (5) and (7) is straightforward. Using relationship (9) with $\phi(x_1;\mu,0)\equiv 0$, the cost function $\phi(x_T;\mu,T-1)$ can be expanded in the form

$$\begin{aligned}\phi(x_T;\mu,T-1) &= \min_{x_1,x_2,\cdots,x_{T-1}} \Bigg\{\mu \sum_{t=1}^{T-1} [x_{t+1}-F(t)x_t-a(t)]'D(t) \\ &\quad \cdot [x_{t+1}-F(t)x_t-a(t)] \\ &\quad + \sum_{t=1}^{T-1}[y_t-H(t)x_t-b(t)]' \\ &\quad \cdot M(t)[y_t-H(t)x_t-b(t)]\Bigg\}. \qquad (10)\end{aligned}$$

Recalling definitions (3) and (4) for c_D and c_M, it is then immediately seen that the minimization problem (7) is an alternative representation for the incompatibility cost minimization problem (5).

The recurrence relation (9) is a special case of a multicriteria filter shown elsewhere [6] to generalize various well-known filters such as those of Kalman [7], Viterbi [8], Larson–Peschon [9], and Swerling [10]. It illustrates how one might formulate and update a cost-of-estimation function for a dynamic process when discrepancy terms are not given a probabilistic interpretation. The recurrence relation (9) thus replaces the use of Bayes' rule, which would be employed if discrepancy terms were interpreted as random quantities having known probability distributions and satisfying various independence restrictions. This point will be elaborated in Section IV, below.

B. A More Concrete Representation for the FLS Filter

It will now be shown how the basic recurrence relation (9) can be more concretely represented in terms of recurrence relations for an $n\times n$ matrix $Q_T(\mu)$, an $n\times 1$ vector $p_T(\mu)$, and a scalar $r_T(\mu)$.

From general considerations in linear-quadratic control theory, it is known that if the cost function appearing in the righthand side expression in (9) is given by

$$\begin{aligned}\phi(x_T;\mu,T-1) &= x_T'Q_{T-1}(\mu)x_T \\ &\quad -2p_{T-1}(\mu)'x_T+r_{T-1}(\mu), \qquad (11)\end{aligned}$$

where $Q_{T-1}(\mu)$ is a real $n\times n$ symmetric matrix, then the cost function appearing on the lefthand side has the form

$$\begin{aligned}\phi(x_{T+1};\mu,T) &= x_{T+1}'Q_T(\mu)x_{T+1} \\ &\quad -2p_T(\mu)'x_{T+1}+r_T(\mu). \qquad (12)\end{aligned}$$

We shall show this below in detail.

First, suppose the initial cost function takes the quadratic form

$$\phi(x_1;\mu,0)=x_1'Q_0(\mu)x_1-2p_0(\mu)'x_1+r_0(\mu), \qquad (13)$$

where the $n\times n$ matrix $Q_0(\mu)$ is symmetric and positive semidefinite. As earlier noted, this function summarizes our knowledge of the cost of estimating that the system is in state x_1 at time $T=1$ before an observation vector at time $T=1$ has been received. For the particular cost specifications (3) and (4), the coefficient terms $Q_0(\mu)$, $p_0(\mu)$, and $r_0(\mu)$ are all zero.

Let us now determine the recurrence relations connecting $Q_T(\mu)$, $p_T(\mu)$, and $r_T(\mu)$ with $Q_{T-1}(\mu)$, $p_{T-1}(\mu)$, and $r_{T-1}(\mu)$ for an arbitrary time $T\geqslant 1$, where the $n\times n$ matrix $Q_{T-1}(\mu)$ is symmetric and positive semidefinite. Consider (9) for any given x_{T+1}. The large curly bracketed term in (9) breaks down into quadratic, linear, and constant parts with respect to x_T, as follows:

$$\begin{aligned}\{\cdots\} &= x_T'[\mu F(T)'D(T)F(T) \\ &\quad +H(T)'M(T)H(T)+Q_{T-1}(\mu)]x_T \\ &\quad +(2\mu[x_{T+1}-a(T)]'D(T)[-F(T)] \\ &\quad +2[y_T-b(T)]'M(T)[-H(T)] \\ &\quad -2p_{T-1}(\mu)')x_T+\mu[x_{T+1}-a(T)]'D(T) \\ &\quad \cdot[x_{T+1}-a(T)]+[y_T-b(T)]'M(T)[y_T-b(T)] \\ &\quad +r_{T-1}(\mu). \qquad (14)\end{aligned}$$

To do the minimization called for in (9), the derivative with respect to x_T of the right-hand side of (14) is set equal to the null vector, which yields

$$\begin{aligned}0 &= [\mu F(T)'D(T)F(T)+H(T)'M(T)H(T) \\ &\quad +Q_{T-1}(\mu)]x_T \\ &\quad -(\mu[x_{T+1}-a(T)]'D(T)F(T) \\ &\quad +[y_T-b(T)]'M(T)H(T)+p_{T-1}(\mu)')'. \qquad (15)\end{aligned}$$

Assuming the bracketed term in (15) is invertible (e.g., assuming the positive semidefinite matrix $Q_{T-1}(\mu)$ is positive definite, or that either $F(T)$ or $H(T)$ has rank n), the optimizing vector x_T is given by

$$\begin{aligned}x_T &= [\mu F(T)'D(T)F(T)+H(T)'M(T)H(T) \\ &\quad +Q_{T-1}(\mu)]^{-1} \\ &\quad \times(\mu F(T)'D(T)[x_{T+1}-a(T)] \\ &\quad +H(T)'M(T)[y_T-b(T)]+p_{T-1}(\mu)). \qquad (16)\end{aligned}$$

To simplify the notation, let us now introduce the symmetric matrix $V_T(\mu)$ as

$$V_T(\mu)=[\mu F(T)'D(T)F(T)+H(T)'M(T)H(T)+Q_{T-1}(\mu)]^{-1}. \quad (17)$$

Then we may write the optimizing vector x_T in the form

$$x_T=s_T(\mu)+G_T(\mu)x_{T+1}, \quad (18)$$

where

$$s_T(\mu)=V_T(\mu)(H(T)'M(T)[y_T-b(T)]+p_{T-1}(\mu)-\mu F(T)'D(T)a(T)) \quad (19)$$

and

$$G_T(\mu)=V_T(\mu)\mu F(T)'D(T). \quad (20)$$

Now we are ready to find $\phi(x_{T+1};\mu,T)$. Substituting (18) into (9), the quadratic terms in x_{T+1} have the matrix $Q_T(\mu)$ given by

$$\begin{aligned}&\mu[I-F(T)G_T(\mu)]'D(T)[I-F(T)G_T(\mu)]\\&\quad+(H(T)G_T(\mu))'M(T)H(T)G_T(\mu)\\&\quad+G_T(\mu)'Q_{T-1}(\mu)G_T(\mu)\\&=G_T(\mu)'V_T(\mu)^{-1}G_T(\mu)\\&\quad+2\mu D(T)[-F(T)]G_T(\mu)+\mu D(T). \quad (21)\end{aligned}$$

But

$$G_T(\mu)'=\mu D(T)F(T)V_T(\mu), \quad (22)$$

so that

$$G_T(\mu)'V_T(\mu)^{-1}=\mu D(T)F(T). \quad (23)$$

It follows that

$$\begin{aligned}Q_T(\mu)&=\mu D(T)F(T)G_T(\mu)\\&\quad-2\mu D(T)F(T)G_T(\mu)+\mu D(T)\\&=\mu D(T)[I-F(T)G_T(\mu)]. \quad (24)\end{aligned}$$

By standard matrix manipulations (see, e.g., [11, p. 7]), it can be shown that $Q_T(\mu)$ in (24) is positive semidefinite given the positive semidefiniteness of $Q_{T-1}(\mu)$ and the positive definiteness of the weight matrices $D(T)$ and $M(T)$ as assumed in Section II.

Next we shall determine the vector $p_T(\mu)$. Consider, again, the substitution of (18) into (9). The linear terms in x_{T+1} have the coefficient vector $-2p_T(\mu)$ given by

$$\begin{aligned}&2G_T(\mu)'V_T(\mu)^{-1}s_T(\mu)+2\mu D(T)[-F(T)]s_T(\mu)\\&\quad+G_T(\mu)'\{2\mu F(T)'D(T)a(T)+2[-H(T)]'M(T)\\&\quad\cdot[y_T-b(T)]-2p_{T-1}(\mu)\}\\&\quad+2\mu D(T)[-a(T)]. \quad (25)\end{aligned}$$

It follows, after some simplification, that

$$p_T(\mu)=G_T(\mu)'[H(T)'M(T)[y_T-b(T)]+p_{T-1}(\mu)]+Q_T(\mu)'a(T). \quad (26)$$

In a similar manner, we find for $r_T(\mu)$ that

$$\begin{aligned}r_T(\mu)&=r_{T-1}(\mu)+[y_T-b(T)]'M(T)[y_T-b(T)]\\&\quad+\mu a(T)'D(T)a(T)\\&\quad-s_T(\mu)'[V_T(\mu)']^{-1}s_T(\mu). \quad (27)\end{aligned}$$

The relations (24), (26), and (27) constitute the desired recurrence relations for $Q_T(\mu)$, $p_T(\mu)$, and $r_T(\mu)$.

Finally, using these recurrence relations, the FLS filter estimate (8) for the state vector at time $T\geqslant 1$ can also be given a more concrete representation. Let

$$U_T(\mu)=H(T)'M(T)H(T)+Q_{T-1}(\mu), \quad (28)$$

and let

$$z_T(\mu)=H(T)'M(T)[y_T-b(T)]+p_{T-1}(\mu). \quad (29)$$

Then

$$x_T^{FLS}(\mu,T)=[U_T(\mu)]^{-1}z_T(\mu). \quad (30)$$

C. FLS Smoothed State Estimates

Consider the problem of obtaining the FLS smoothed estimate for the state vector x_T at time T as the length of the process increases from T to $T+1$ and an additional observation vector y_{T+1} is obtained.

In preparation for time $T+1$, the quadratic, linear, and constant terms $Q_T(\mu)$, $p_T(\mu)$, and $r_T(\mu)$ characterizing the cost function in (12) have been calculated and stored. As a byproduct of this calculation, the unique cost-minimizing x_T as a function of x_{T+1} has been determined in accordance with (18) to be $x_T=s_T(\mu)+G_T(\mu)x_{T+1}$. Using (30) updated to time $T+1$, the FLS filter estimate for the state vector at time $T+1$ is given by

$$x_{T+1}^{FLS}(\mu,T+1)=[U_{T+1}(\mu)]^{-1}z_{T+1}(\mu). \quad (31)$$

The FLS smoothed estimate for the time-T state vector x_T, based on the observation vectors $y_1,\cdots,y_{T+1}$ for times 1 through $T+1$, is then given by

$$x_T^{FLS}(\mu,T+1)=s_T(\mu)+G_T(\mu)x_{T+1}^{FLS}(\mu,T+1). \quad (32)$$

More generally, given any fixed time t, $0\leqslant t\leqslant T$, the FLS smoothed estimate $x_t^{FLS}(\mu,T+1)$ for the state vector x_t at time t, based on the observation vectors $y_1,\cdots,y_{T+1}$ for times 1 through $T+1$, is found by solving the system of equations

$$\begin{aligned}x_t&=s_t(\mu)+G_t(\mu)x_{t+1}\\&\vdots\\x_T&=s_T(\mu)+G_T(\mu)x_{T+1} \quad (33a)\end{aligned}$$

in reverse order, starting with the initial condition

$$x_{T+1}=x_{T+1}^{FLS}(\mu,T+1). \quad (33b)$$

Relations (30) and (33) for generating the FLS filtered and smoothed state estimates result naturally from the dynamic programming procedure used to update incompatibility cost. Alternative formulas for generating these state estimates could be obtained from (30) and (31) using appropriate matrix manipulations (see [11]). Based on past numerical experience, however, we elected to adhere closely to the dynamic programming formulation.

A Fortran program *GFLS* for generating the FLS filtered and smoothed state estimates by means of the relations (30) and (33) is provided in an appendix to this paper. In simulation experiments conducted to date with *GFLS* on an IBM Model 3090, the generated FLS estimates have satisfied the first-order necessary conditions for the cost-minimization problem (5) up to the maximum degree of accuracy (fourteen to sixteen digits) permitted by the double-precision word length employed. Our empirically based belief, then, is that the suggested procedure for determining the FLS filtered and smoothed state estimates is numerically stable and highly accurate.

IV. Relationship with Kalman Filtering

FLS and Kalman filtering address conceptually distinct problems. FLS treats a multicriteria model specification problem that does not require probability assumptions either for its motivation or for its solution: the characterization of the set of all state sequence estimates that achieve vector-minimal incompatibility between imperfectly specified theoretical relations and process observations. Kalman filtering is a point estimation technique that determines the most probable state sequence for a stochastic model assumed to be correctly and completely specified. Nevertheless, when applied to approximately linear systems, the two approaches satisfy duality relations which generalize the well-known duality [7, p. 42] between the noise-free regulator problem and maximum a posteriori probability estimation.

Conceptual differences between FLS and Kalman filtering are examined in Section IV-A. In Section IV-B the Kalman filter recurrence equations are derived by means of simple cost-function arguments that mimic the steps outlined in Section III-B for the derivation of the FLS recurrence relations. Probabilistic arguments (e.g., Bayes' Rule or iterated expectations) are not required. Conversely, in Section IV-C it is seen that the FLS recurrence relations for generating any particular state sequence estimate along the cost-efficient frontier reduce to information filter equations, the "inverse" of Kalman filter equations, if the model discrepancy terms are assumed to satisfy various independence and normality restrictions. Implications of these duality relations are discussed in Section IV-D.

A. Conceptual Differences Between FLS and Kalman Filtering

Previous sections of this paper investigate how filtering and smoothing might be undertaken for the approximately linear system (1) and (2) when the dynamic and measurement discrepancy terms $w_t \equiv [x_{t+1} - F(t)x_t - a(t)]$ and $v_t \equiv [y_t - H(t)x_t - b(t)]$ are incommensurable model specification errors. A multicriteria FLS solution is proposed for this problem. As seen in Section III, this multicriteria solution can be implemented by means of a *family* of Riccati-type recurrence relations. The Riccati-equation form of these recurrence relations is not surprising; it has been known for decades [12] that linear-quadratic minimization leads to recurrence relations of this type. What is new is the probability-free motivation provided for why one should be interested in this entire family of recurrence relations.

Suppose, instead, that the following probability relations, commonly assumed in Kalman filtering studies, are introduced for the discrepancy terms w_t and v_t and for the initial state vector x_1:

- [PDF for w_t] $= N(0, S(t))$;
- [PDF for v_t] $= N(0, R(t))$;
- (w_t) and (v_t) are mutually and serially independent processes;
- [PDF for x_1] $= N(x_1^*, \Sigma_1)$;
- x_1 is distributed independently of v_t and w_t for each t.

(34)

Under assumptions (34), the discrepancy terms w_t and v_t are interpreted as white noise random vectors with known Gaussian probability density functions (PDF's) governing both their individual and joint behavior. In particular, w_t and v_t are now supposed to be perfectly commensurable quantities that can be scaled and weighed relative to one another. The FLS interpretation for w_t and v_t as conceptually distinct apple-and-orange model specification errors incorporating everything unknown about the dynamic and measurement aspects of the process in thus dramatically altered.

Combining the measurement relations (2) with the probability relations (34) permits the derivation of a probability density function $P(Y_T|X_T)$ for the observation sequence $Y_T = (y_1, \cdots, y_T)$ conditional on the state sequence $X_T = (x_1, \cdots, x_T)$. Combining the dynamic relations (1) with the probability relations (34) permits the derivation of a "prior" probability density function $P(X_T)$ for X_T. The multiplication of these two derived probability density functions yields the joint probability density function for X_T and Y_T,

$$P(Y_T|X_T) \cdot P(X_T) = P(X_T, Y_T). \qquad (35)$$

The joint probability density function (35) 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.

Given the probability relations (34), the usual Kalman filter objective is to determine the maximum *a posteriori* (MAP) state sequence, i.e., the state sequence which maximizes the posterior probability density function $P(X_T|Y_T)$. Since the observation sequence Y_T is assumed to be given, this objective is equivalent to determining the state sequence which maximizes the product of $P(X_T|Y_T)$ and $P(Y_T)$. By the agreed upon rules of probability theory,

$$P(X_T|Y_T) \cdot P(Y_T) = P(Y_T|X_T) \cdot P(X_T), \qquad (36)$$

where, as earlier explained, the right-hand expression in

(36) can be evaluated using (1), (2), and the probability relations (34). Determining the MAP state sequence is thus equivalent to determining the state sequence that minimizes the scalar "incompatibility cost function"

$$c(X_T,T) = -\log\left[P(Y_T|X_T)\cdot P(X_T)\right]. \quad (37)$$

What has been achieved by the introduction of the probability relations (34)? Without relations such as (34), the dynamic and measurement discrepancy terms cannot be scaled and weighed relative to one another. The filtering and smoothing problem is thus intrinsically a multicriteria optimization problem: Conditional on the given observations, determine the state sequence estimates which are in some sense minimally incompatible with each of the imperfectly specified theoretical relations (1) and (2). Given the probability relations (34), however, the discrepancy terms are transformed into perfectly commensurable "disturbance terms" impinging on correctly specified theoretical relations in accordance with known probability distributions. In this case, MAP estimation seems an emminently reasonable way to proceed. The multicriteria optimization problem is thus transformed into the scalar optimization problem of determining the most probable state sequence for a stochastic model assumed to be correctly and completely specified.

Making use of Bayes' rule, Larson and Peschon [9] develop a recurrence relation for the sequential updating of the posterior density function $P(X_T|Y_T)$ as the duration T of the process increases and additional observation vectors are obtained. This recurrence relation is used to determine recursively the MAP state sequence for each time T. The Larson–Peschon filter is derived under assumptions (34) without the requirement that the PDF's be Gaussian; nonlinearity of the dynamic and measurement relations is also permitted. Larson and Peschon show that their filter reduces to the Kalman filter when Gaussian distributions and linear dynamic and measurement relations are assumed.

For example, suppose for simplicity that the forcing terms $a(t)$ and $b(t)$ in the dynamic and measurement relations (1) and (2) are identically zero. For this case, Larson and Peschon obtain the relations

$$\begin{aligned}\Sigma^{-1}(T+1|T+1) &= H(T+1)'R(T+1)^{-1}H(T+1)\\ &\quad + \left[F(T)\Sigma(T|T)F(T)'+S(T)\right]^{-1};\\ x(T+1|T+1) &= F(T)x(T|T)\\ &\quad + \Sigma(T+1|T+1)H(T+1)'\\ &\quad \cdot R(T+1)^{-1}\left[y_{T+1}-H(T+1)F(T)x(T|T)\right].\end{aligned} \quad (38)$$

In (38), $x(T+1|T+1)$ is the MAP estimate for the state vector at time $T+1$, conditional on the observation vectors obtained through time $T+1$; and $\Sigma(T+1|T+1)$ is the error covariance matrix for $x(T+1|T+1)$. By use of appropriate matrix inversion formulas, the relations (38) can be transformed into a pair of recurrence relations either for the error covariance matrix $\Sigma(T|T)$ and the state estimate $x(T|T)$—the standard Kalman filter equations (see [7] and [13, pp. 105–120])—or for the inverse "information matrix" $\Sigma^{-1}(T|T)$ and the modified state estimate $\Sigma^{-1}(T|T)x(T|T)$, yielding the "information filter equations" (see [13, pp. 139–142]).

B. Cost Derivation of the Kalman Filter Recurrence Relations

It will now be shown that the recursive relations (38) can alternatively be derived by means of simple intuitive cost considerations, without reliance on probabilistic arguments.

As in Section IV-A, suppose for simplicity that the forcing terms $a(t)$ and $b(t)$ in (1) an (2) are identically zero. For any time $T>1$, let X_T denote the T-length state trajectory $(x_1,\cdots,x_T)$; and let the time-T incompatibility cost function be specified by

$$\begin{aligned}c(X_T,T) = \Bigg\{ &\sum_{t=1}^{T-1}\left[x_{t+1}-F(t)x_t\right]'S(t)^{-1}\left[x_{t+1}-F(t)x_t\right]\\ &+\sum_{t=1}^{T}\left[y_t-H(t)x_t\right]'R(t)^{-1}\left[y_t-H(t)x_t\right]\\ &+\left[x_1-x_1^*\right]'\Sigma_1^{-1}\left[x_1-x_1^*\right]\Bigg\}.\end{aligned} \quad (39)$$

Also, let the time-1 incompatibility cost function be specified by

$$c(X_1,1) = \left[x_1-x_1^*\right]'\Sigma_1^{-1}\left[x_1-x_1^*\right]. \quad (40)$$

Given the probability relations (34), the time-T incompatibility cost function (39) coincides with the previously defined incompatibility cost function (37) apart from a nonessential constant term. Finally, for any time $T\geqslant 1$, let $C^F(x_T,T)$ denote the minimum cost (39) attainable at time T, conditional on the time-T state vector being x_T.

By definition, the state-conditioned cost function $C^F(x_1,1)$ for time 1 coincides with the time-1 cost function $c(X_1,1)$; hence it has the quadratic form

$$C^F(x_1,1) = \left[x_1-x(1|1)\right]'\Sigma^{-1}(1|1)\left[x_1-x(1|1)\right], \quad (41a)$$

where

$$\Sigma^{-1}(1|1) \equiv \Sigma_1^{-1}; \quad (41b)$$

$$x(1|1) \equiv x_1^*. \quad (41c)$$

Note that $x(1|1)$ is the state vector x_1 which minimizes the state-conditioned cost function $C^F(x_1,1)$.

Suppose the state-conditioned cost function $C^F(x_T,T)$ for some time $T\geqslant 1$ has the quadratic form

$$C^F(x_T,T) = \left[x_T-x(T|T)\right]'\Sigma^{-1}(T|T)\cdot\left[x_T-x(T|T)\right]+k_T, \quad (42)$$

where k_T is independent of x_T. As shown in [6, Section 4.3], the state-conditioned cost function for time $T+1$

satisfies the recurrence relation

$$C^F(x_{T+1},T+1) = \min_{x_T}\{\Delta c(x_T,x_{T+1},T+1) + C^F(x_T,T)\}, \quad (43a)$$

where

$$\Delta c(x_T,x_{T+1},T+1) \equiv [x_{T+1}-F(T)x_T]'S(T)^{-1}[x_{T+1}-F(T)x_T] + [y_T-H(T)x_T]'R(T)^{-1}[y_T-H(T)x_T] \quad (43b)$$

denotes the total change in cost associated with the transition from T to $T+1$. Substituting (42) into (43a), it follows by straightforward calculations (analogous to those in Section III-B) that the state-conditioned cost function for time $T+1$ has the quadratic form

$$C^F(x_{T+1},T+1) = [x_{T+1}-x(T+1|T+1)]'\Sigma^{-1}(T+1|T+1) \cdot [x_{T+1}-x(T+1|T+1)] + k_{T+1}, \quad (44)$$

where $\Sigma(T+1|T+1)$ and $x(T+1|T+1)$ satisfy the recursive relations (38). As is clear from (44), $x(T+1|T+1)$ is the state vector x_{T+1} that minimizes the state-conditioned cost function $C^F(x_{T+1},T+1)$.

The terms $\Sigma(T+1|T+1)$ and $x(T+1|T+1)$ appearing in the cost expression (44) thus coincide with the error covariance matrix and state estimate generated by the Kalman filter recurrence relations derived from (38). Note, also, that the quadratic and linear coefficient terms $\Sigma^{-1}(T+1|T+1)$ and $\Sigma^{-1}(T+1|T+1)x(T+1|T+1)$ for the cost expression (44), considered as a function of x_{T+1}, coincide with the information matrix and modified state estimate generated by the information filter equations. It is not surprising, then, that the cost arguments used to derive the recursive relations (38) for these terms are entirely analogous to the cost arguments used in Section III-B to determine recursive relations for the quadratic and linear coefficient terms $Q_T(\mu)$ and $p_T(\mu)$ for the cost expression $\phi(x_{T+1};\mu,T)$.

In summary, the Kalman and information filter recurrence relations can be derived for approximately linear systems using simple cost arguments, without recourse to probabilistic arguments such as Bayes' rule or iterated expectations. All that is needed is that the basic cost function used to measure theory and data incompatibility be a quadratic function exhibiting time-separability.

C. The FLS Recurrence Relations as Information Filter Equations

Conversely, the FLS recurrence relations associated with any given point μ on the cost-efficient frontier reduce to a variant of the information filter equations if the theoretical relations (1) and (2) are augmented by probability relations of the form (34).

Specifically, suppose the dynamic weight matrix $\mu D(t)$ is taken to be the inverse of the covariance matrix $S(t)$ for w_t, and the measurement weight matrix $M(t)$ is taken to be the inverse of the covariance matrix $R(t)$ for v_t, for each time t; and suppose also that the initial cost matrix $Q_0(\mu)$ is taken to be the inverse of the covariance matrix Σ_1 for the initial state vector x_1. In this case the matrix $U_T(\mu)$ in (28) corresponds to the inverse of the "measurement-update" error covariance matrix $\Sigma(T|T)$ and the vector $z_T(\mu)$ in (29) corresponds to the modified state estimate $\Sigma^{-1}(T|T)x(T|T)$. Moreover, the matrix $Q_T(\mu)$ corresponds to the inverse of the "time-update" error covariance matrix $\Sigma(T+1|T)$, defined [13, ch. 3] to be the error covariance matrix for the MAP estimate of x_{T+1} based on observations through time T.

D. Duality Implications

If the probability relations (34) are justified for a given filtering and smoothing application, they should of course be incorporated in the estimation procedure. However, for many important applications—particularly in the social sciences—obtaining agreement among researchers regarding probability relations such as (34) can be difficult.

For example, the process observations may be the outcome of a nonreplicable experiment, so that no objective test of these relations can be carried out. Also, the theoretical relations may represent tentatively held conjectures concerning a poorly understood process; or they may be a linearized set of relations obtained for an analytically intractable nonlinear process, as in many aerospace filtering and smoothing problems. In these cases it is doubtful whether the discrepancy terms are governed by *any* meaningful probability relations. Independence restrictions, in particular, are questionable and troublesome.

For these reasons, the FLS procedure, with its minimal assumptions concerning discrepancy terms, appears to offer a useful complement to existing filtering and smoothing techniques. Moreover, the FLS duality relations discussed in previous sections may shed some light on the robustness properties of the Kalman filter.

It is now conventional to interpret any quadratic criterion function representing sums of squared dynamic and measurement errors—e.g., the Kalman filter criterion function (39)—as a log-likelihood expression arising from some underlying stochastic model in which model discrepancy terms are interpreted as independent and normally distributed random variables. Yet it is also known that Kalman filtering works remarkably well in some contexts in which these strong stochastic assumptions are not even remotely satisfied. A partial explanation for this robustness is that the Kalman filter criterion function can be given an alternative interpretation: namely, as a cost function embodying the criterion that model discrepancy terms be *small*.

"Smallness" should not be confused with "randomness." Postulating that x_{t+1} is close to $[F(t)x_t + a(t)]$

does not mean that the discrepancy term $[x_{t+1} - F(t)x_t - a(t)]$ is necessarily a random vector. As numerous experiments with FLS have shown (see, e.g., [3]), the postulate of small dynamic and measurement discrepancy terms is a powerful assumption that allows state trajectories to be tracked and recovered with surprising qualitative accuracy at each point along the cost-efficient frontier.

V. Conclusion

The main purpose of this paper is to present a probability-free multicriteria approach to the problem of filtering and smoothing when prior beliefs concerning dynamics and measurements take an approximately linear form. In particular, model discrepancy terms are treated as model specification errors that may not have any meaningful probabilistic description. Applications are envisioned in various fields, particularly in the social and biological sciences, where obtaining agreement among researchers regarding probability relations for discrepancy terms is difficult.

The essence of the proposed FLS procedure is the cost-efficient frontier. This frontier, a curve in a two-dimensional cost plane, provides an explicit and systematic way to determine the efficient trade-offs between the separate costs incurred for dynamic and measurement specification errors.

The estimated state sequences whose associated cost vectors attain the cost-efficient frontier, referred to as FLS estimates, show how the state vector could have evolved over time in a manner minimally incompatible with the prior dynamic and measurement specifications. Each FLS estimate has the property that it is not possible simultaneously to reduce both the dynamic and the measurement cost by choice of an alternative state sequence estimate. The similarities displayed by the FLS estimates suggest working hypotheses regarding the evolution of the actual state vector. The divergencies displayed by these estimates reflect the residual uncertainty inherent in the problem specifications regarding the exact nature of this evolution. Without additional prior information, restricting attention to any proper subset of the FLS estimates is an arbitrary decision.

A Fortran program *GFLS* for implementing the FLS filtering and smoothing procedure for approximately linear systems is provided in the appendix. This program has been used in both simulation and empirical studies of time-varying linear regression ([3]–[5]).

Nonlinear systems are studied from the multicriteria FLS point of view in [6].

Appendix

This appendix provides a Fortran program *GFLS* that implements the sequential FLS solution of the bicriteria filtering and smoothing problem posed in Section II. The program has received extensive testing. In addition, the program incorporates a check of the sequential FLS solution based upon using the standard first-order conditions for the solution of the incompatibility cost minimization problem (5).

The variable names used in the *GFLS* program adhere strictly to those used in the body of the paper. Moreover, numerous comment statements are interspersed throughout the program that are geared to the equation numbers used in the paper.

User inputs are required in a subroutine INPUT. This subroutine initializes the penalty weight μ, the total number of observation vectors *TCAP*, the state vector dimension n, the observation vector dimension m, and the initial cost function coefficient terms $Q_0(\mu)$, $p_0(\mu)$, and $r_0(\mu)$. The program is currently dimensioned for $TCAP \leqslant 110$, $n \leqslant 15$, and $m \leqslant 15$.

Subroutine INPUT also requires the user to set two flags. The first flag, IFLAGR, is set equal to 1 if the user wishes to generate evaluations for the constant terms $r_T(\mu)$ in the cost functions (12), and is set equal to 0 otherwise. The second flag, IFLAGS, is set equal to 1 if the user wishes to generate smoothed state estimates in addition to filtered state estimates, and is otherwise set equal to 0. If the user sets IFLAGS = 1, the program automatically carries out a test of the first-order conditions for the incompatibility cost minimization problem (5).

User inputs are also required in a subroutine MODEL. For each current time T, subroutine MODEL generates the $n \times n$ state transition matrix $F(T)$, the $n \times 1$ dynamic forcing term $a(T)$, the $m \times n$ measurement matrix $H(T)$, the $m \times 1$ measurement forcing term $b(T)$, the $n \times n$ dynamic weight matrix $D(T)$, the $m \times m$ measurement weight matrix $M(T)$, and the $m \times 1$ observation vector y_T. For simulation studies, the observation vector y_T is generated in accordance with the relation $y_T = H(T)x_T + b(T) + v_T$, where x_T is an $n \times 1$ user-specified state vector and v_T is an $m \times 1$ user-specified discrepancy term. The user-specified state vector x_T is stored in an array TRUEX for later comparison with the numerically generated FLS smoothed estimate for x_T.

The *GFLS* program contains subroutines for all needed matrix operations. Currently, these subroutines are dimensioned for 15×15 matrices. To keep the number of subroutines to a minimum, vector and scalar operations are carried out with these matrix subroutines by considering some vectors to lie in the first column of a 15×15 matrix, and some scalars to be the upper-left component of a 15×15 matrix.

```
// EXEC FORTVCLG,REGION=512K
//*
/*JOBPARM COPIES=4
C
C
C      GFLS: FLEXIBLE LEAST SQUARES FOR APPROXIMATELY LINEAR SYSTEMS
C            R. KALABA AND L. TESFATSION
C
C      FILENAME:  GFLS.CNTL
C      LAST UPDATED: 23 OCTOBER 1989
C
       IMPLICIT REAL*8(A-H,O-Z)
       INTEGER T,TCAP,TCAP1
       REAL*8 M
C
C      THIS PROGRAM IS CURRENTLY DIMENSIONED FOR A MAXIMUM OF TCAP=110
C      OBSERVATION VECTORS Y OF MAXIMUM DIMENSION MOBS = 15 WITH STATE
C      VECTORS X OF MAXIMUM DIMENSION N = 15.
```

```
C
      DIMENSION QO(15,15),PO(15,15),RO(15,15),QZERO(15,15)
      DIMENSION PZERO(15,15),RZERO(15,15)
      DIMENSION F(15,15),A(15,15),H(15,15),B(15,15),D(15,15)
      DIMENSION M(15,15),Y(15,15),TRUEX(15,110),YY(15,110)
      DIMENSION HT(15,15),U(15,15),C(15,15),W(15,15),V(15,15)
      DIMENSION E(15,15),Z(15,15),G(15,15),QNEW(15,15),PNEW(15,15)
      DIMENSION S(15,15),RNEW(15,15),XTCAP(15,15),X(15,110)
      DIMENSION AA(15,15),BB(15,15),CC(15,15),DD(15,15),EE(15,15)
      DIMENSION FF(15,15),HH(15,15),OO(15,15),PP(15,15),QQ(15,15)
      DIMENSION RR(15,15),TT(15,15)
C
C     ADDITIONAL ARRAYS IF SMOOTHED ESTIMATES ARE TO BE CALCULATED
C     FOR INTERMEDIATE X VALUES (I.E., IF IFLAGS IS SET AT 1)
C
      DIMENSION GG(15,15,110),SS(15,110)
C
C     THE FOLLOWING SUBROUTINE INITIALIZES THE PENALTY WEIGHT AMU,
C     THE NUMBER OF OBSERVATIONS TCAP, THE STATE VECTOR DIMENSION
C     N, THE OBSERVATION VECTOR DIMENSION MOBS, AND THE INITIAL COST
C     FUNCTION CHARACTERISTICS QZERO, PZERO, AND RZERO.  IT ALSO SETS
C     THE VALUE FOR A FLAG "IFLAGR" TO DETERMINE IF RNEW IS TO BE
C     CALCULATED (1) OR NOT (0) AND A FLAG "IFLAGS" TO DETERMINE IF
C     SMOOTHED ESTIMATES FOR INTERMEDIATE X VALUES ARE TO BE CALCULATED
C     (1) OR NOT (0).
C
      CALL INPUT(AMU,TCAP,N,MOBS,QZERO,PZERO,RZERO,IFLAGR,IFLAGS)
      CALL SHIFT(N,N,QZERO,QO)
      CALL SHIFT(N,1,PZERO,PO)
      CALL SHIFT(1,1,RZERO,RO)
C
C     ENTERING THE MAIN DO LOOP FOR GENERATING Q,P,AND R FOR
C     SUCCESSIVE TIMES T = 1,TCAP USING EQS.(24),(26), AND (27).
C
      DO 50 T=1,TCAP
      CALL MODEL(T,F,A,H,B,D,M,Y,TRUEX)
      DO 5 I=1,MOBS
      YY(I,T) = Y(I,1)
    5 CONTINUE
C
C     GETTING U=HT*M*H + QO IN EQ.(28)
C
      CALL MUL(MOBS,MOBS,N,M,H,AA)
      CALL TRANS(MOBS,N,H,HT)
      CALL MUL(N,MOBS,N,HT,AA,BB)
      CALL ADD(N,N,BB,QO,U)
C
C     GETTING C=FT*D
C
      CALL TRANS(N,N,F,AA)
      CALL MUL(N,N,N,AA,D,C)
C
C     GETTING W=AMU*C*F+U
C
      CALL MUL(N,N,N,C,F,AA)
      CALL MULCON(N,N,AMU,AA,BB)
      CALL ADD(N,N,BB,U,W)
C
C     GETTING V=WINV IN EQ.(17)
C
      CALL INV(N,W,V)
C
C     GETTING E = (Y-B)
C
      CALL SUB(MOBS,1,Y,B,E)
C
C     GETTING Z - HT*M*E + PO IN EQ.(29)
C
      CALL MUL(MOBS,MOBS,1,M,E,AA)
      CALL MUL(N,MOBS,1,HT,AA,BB)
      CALL ADD(N,1,BB,PO,Z)
C
C     GETTING G = AMU*V*C IN EQ.(20)
C
      CALL MUL(N,N,N,V,C,AA)
      CALL MULCON(N,N,AMU,AA,G)
      IF(IFLAGS.EQ.0) GO TO 110
C
C     STORE G FOR CALCULATION OF SMOOTHED ESTIMATES
C
      DO 10 I=1,N
      DO 20 J=1,N
      GG(I,J,T)=G(I,J)
   20 CONTINUE
   10 CONTINUE
  110 CONTINUE
C
C     GETTING QNEW - AMU*D*(I-F*G) IN EQ.(24)
C
      CALL MUL(N,N,N,F,G,AA)
      CALL IDEN(N,BB)
      CALL SUB(N,N,BB,AA,CC)
      CALL MUL(N,N,N,D,CC,DD)
      CALL MULCON(N,N,AMU,DD,QNEW)
C
C     GETTING PNEW - GT*Z+QNEWT*A IN EQ.(26)
C
      CALL TRANS(N,N,G,AA)
      CALL MUL(N,N,1,AA,Z,BB)
      CALL TRANS(N,N,QNEW,CC)
      CALL MUL(N,N,1,CC,A,DD)
      CALL ADD(N,1,BB,DD,PNEW)
C
C     GETTING S = V*(Z - AMU*C*A) IN EQ.(19)
C
      CALL MUL(N,N,1,C,A,BB)
      CALL MULCON(N,1,AMU,BB,CC)
      CALL SUB(N,1,Z,CC,DD)
      CALL MUL(N,N,1,V,DD,S)
      IF(IFLAGS.EQ.0) GO TO 210
C
C     STORE S FOR CALCULATION OF SMOOTHED ESTIMATES
C
      DO 30 I=1,N
      SS(I,T)=S(I,1)
   30 CONTINUE
  210 CONTINUE
      IF(IFLAGR.EQ.0) GO TO 310
C
C     GETTING RNEW - RO + ET*M*E + AMU*AT*D*A - ST*W*S IN EQ.(27)
C
      CALL MUL(MOBS,MOBS,1,M,E,AA)
      CALL TRANS(MOBS,1,E,BB)
      CALL MUL(1,MOBS,1,BB,AA,CC)
      CALL ADD(1,1,RO,CC,DD)
      CALL MUL(N,N,1,D,A,EE)
      CALL TRANS(N,1,A,FF)
      CALL MUL(1,N,1,FF,EE,HH)
      CALL MULCON(1,1,AMU,HH,OO)
      CALL ADD(1,1,DD,OO,PP)
      CALL MUL(N,N,1,W,S,QQ)
      CALL TRANS(N,1,S,RR)
      CALL MUL(1,N,1,RR,QQ,TT)
      CALL SUB(1,1,PP,TT,RNEW)
  310 CONTINUE
      IF(T.EQ.TCAP) GO TO 50
C
C     UPDATING QO,PO, AND RO
C
      CALL SHIFT(N,N,QNEW,QO)
      CALL SHIFT(N,1,PNEW,PO)
      IF(IFLAGR.EQ.0) GO TO 50
      CALL SHIFT(1,1,RNEW,RO)
   50 CONTINUE
C
C     GETTING THE FLS FILTER ESTIMATE FOR XTCAP - UINV*Z IN EQ.(30)
C
      CALL INV(N,U,AA)
      CALL MUL(N,N,1,AA,Z,XTCAP)
      DO 65 I=1,N
      X(I,TCAP)=XTCAP(I,1)
   65 CONTINUE
      IF (IFLAGS.EQ.1) GOTO 410
C
C     PRINTING OUT THE FLS FILTER ESTIMATE FOR XTCAP
C
      CALL OUTPUT(TCAP,N,X,TRUEX)
      IF(IFLAGS.EQ.0) GOTO 510
  410 CONTINUE
C
C     GETTING SMOOTHED ESTIMATES FOR X1,..., XTCAP-1 IN EQS.(33A)
C
      TCAP1=TCAP-1
      DO 70 T=1,TCAP1
      L=TCAP-T
      DO 80 I=1,N
      X(I,L)=SS(I,L)
      DO 90 J=1,N
      X(I,L)=X(I,L)+GG(I,J,L)*X(J,L+1)
   90 CONTINUE
   80 CONTINUE
   70 CONTINUE
C
C     PRINTING OUT THE FLS ESTIMATES FOR X1,...,XTCAP
C
      DO 150 T=1,TCAP
      CALL OUTPUT(T,N,X,TRUEX)
  150 CONTINUE
C     VALIDATION TEST: HOW WELL DO THE FLS ESTIMATES SATISFY THE
C     FIRST-ORDER CONDITIONS FOR THE COST MINIMIZATION PROBLEM (5)
      CALL FOCTST(X,YY)
  510 CONTINUE
      STOP
      END
C
C     MATRIX SUBROUTINES FOR ADDITION, MULTIPLICATION, TRANSPOSITION,
C     SUBTRACTION, INVERSION, MULTIPLICATION BY A SCALAR, SHIFT, AND
C     FORMATION OF AN IDENTITY MATRIX
C
C     OBTAINING THE SUM C=A+B OF TWO NROW X MCOL MATRICES A AND B
C
      SUBROUTINE ADD(NROW,MCOL,A,B,C)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),B(15,15),C(15,15)
      DO 10 I=1,NROW
      DO 20 J=1,MCOL
      C(I,J)=A(I,J)+B(I,J)
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
C
C     OBTAINING THE PRODUCT C=A*B OF AN NROW X L MATRIX A AND AN
C     L X MCOL MATRIX B
C
      SUBROUTINE MUL(NROW,L,MCOL,A,B,C)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),B(15,15),C(15,15)
      DO 10 I=1,NROW
      DO 20 J=1,MCOL
      SUM=0.0D+00
      DO 30 K=1,L
      SUM=SUM+A(I,K)*B(K,J)
   30 CONTINUE
      C(I,J)=SUM
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
C
```

```
C     OBTAINING THE TRANSPOSE B OF AN NROW X MCOL MATRIX A
C
      SUBROUTINE TRANS(NROW,MCOL,A,B)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),B(15,15)
      DO 10 I=1,NROW
      DO 20 J=1,MCOL
      B(J,I)=A(I,J)
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
C
C     OBTAINING THE DIFFERENCE C=A-B BETWEEN NROW X MCOL MATRICES
C     A AND B
C
      SUBROUTINE SUB(NROW,MCOL,A,B,C)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),B(15,15),C(15,15)
      DO 10 I=1,NROW
      DO 20 J=1,MCOL
      C(I,J)=A(I,J)-B(I,J)
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
C
C     OBTAINING THE INVERSE C OF A K X K MATRIX A
C
      SUBROUTINE INV(K,A,C)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),B(15,30),C(15,15)
      DO 5 J=1,K
      DO 6 I=1,K
      B(I,J)=A(I,J)
    6 CONTINUE
    5 CONTINUE
      K2=K*2
      DO 7 J=1,K
      DO 8 I=1,K
      B(I,K+J)=0.0D+00
      IF(I.EQ.J) B(I,K+J)=1.0D+00
    8 CONTINUE
    7 CONTINUE
C
C     THE PIVOT OPERATION STARTS HERE
C
      DO 9 L=1,K
      PIVOT = B(L,L)
      DO 13 J=L,K2
      B(L,J)=B(L,J)/PIVOT
   13 CONTINUE
C
C     TO IMPROVE THE ROWS
C
      DO 14 I=1,K
      IF(I.EQ.L) GO TO 14
      AIL=B(I,L)
      DO 15 J=L,K2
      B(I,J)=B(I,J)-AIL*B(L,J)
   15 CONTINUE
   14 CONTINUE
    9 CONTINUE
      DO 45 I=1,K
      DO 46 J=1,K
      C(I,J)=B(I,K+J)
   46 CONTINUE
   45 CONTINUE
      RETURN
      END
C
C     OBTAINING THE PRODUCT C*A OF A SCALAR C AND AN NROW X MCOL
C     MATRIX A
C
      SUBROUTINE MULCON(NROW,MCOL,C,A,CA)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),CA(15,15)
      DO 10 I=1,NROW
      DO 20 J=1,MCOL
      CA(I,J)=C*A(I,J)
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
C
C     PUTTING AN NROW X MCOL MATRIX A INTO AN NROW X MCOL MATRIX B
C
      SUBROUTINE SHIFT(NROW,MCOL,A,B)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(15,15),B(15,15)
      DO 10 I=1,NROW
      DO 20 J=1,MCOL
      B(I,J)=A(I,J)
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
C
C     FORMING THE N X N IDENTITY MATRIX E
C
      SUBROUTINE IDEN(N,E)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION E(15,15)
      ZERO=0.0D+00
      ONE=1.0D+00
      DO 10 I=1,N
      DO 20 J=1,N
      E(I,J)=ZERO
   20 CONTINUE
   10 CONTINUE
      DO 30 L=1,N
      E(L,L)=ONE
   30 CONTINUE
      RETURN
      END
C
      SUBROUTINE INPUT(AMU,TCAP,N,MOBS,QZERO,PZERO,RZERO,IFLAGR,IFLAGS)
      IMPLICIT REAL*8(A-H,O-Z)
      INTEGER TCAP
      DIMENSION QZERO(15,15),PZERO(15,15),RZERO(15,15)
      AMU = 1.0D+00
      TCAP = 30
      N = 2
      MOBS = 1
      DO 10 J = 1,N
      DO 20 I = 1,N
      QZERO(I,J) = 0.0D+00
      PZERO(I,J) = 0.0D+00
      RZERO(I,J) = 0.0D+00
   20 CONTINUE
   10 CONTINUE
      IFLAGR=1
      IFLAGS=1
      RETURN
      END
C
      SUBROUTINE MODEL(T,F,A,H,B,D,M,Y,TRUEX)
      IMPLICIT REAL*8(A-H,O-Z)
      REAL*8 M
      REAL*4 GNORM
      INTEGER T,TCAP
      DIMENSION F(15,15),A(15,15),H(15,15),B(15,15),D(15,15),M(15,15)
      DIMENSION Y(15,15),TRUEX(15,110),ZERO(15,15)
      DIMENSION QZERO(15,15),PZERO(15,15),RZERO(15,15)
C
C     TIME-VARYING LINEAR REGRESSION STUDY WITH A SHIFT IN THE COEFF.
C     VECTOR AT MIDPOINT OBSERVATION TIME T=15 (SEE SECTION 2).
      CALL INPUT(AMU,TCAP,N,MOBS,QZERO,PZERO,RZERO,IFLAGR,IFLAGS)
      SIGMA = 0.000D+00
      DO 10 I=1,15
      DO 20 J=1,15
      ZERO(I,J) = 0.0D+00
   20 CONTINUE
   10 CONTINUE
      CALL IDEN(N,F)
      CALL SHIFT(N,1,ZERO,A)
      H(1,1)=1.0D+00
      H(1,2)=1.0D+00
      AT=DFLOAT(T)
      IF(T.EQ.1) GO TO 200
      H(1,1)=DSIN(10.0D+00+(AT))+0.01D+00
      H(1,2)=DCOS(10.0D+00+(AT))
  200 CONTINUE
      CALL SHIFT(MOBS,1,ZERO,B)
      CALL IDEN(N,D)
      CALL IDEN(MOBS,M)
      IF (T.GT.15) GOTO 150
      TRUEX(1,T) = 2.0D+00
      TRUEX(2,T) = 3.0D+00
      GOTO 175
  150 TRUEX(1,T) = 4.0D+00
      TRUEX(2,T) = 5.0D+00
  175 CONTINUE
      UU = DBLE(GNORM(0))
      Y(1,1)=H(1,1)*TRUEX(1,T) + H(1,2)*TRUEX(2,T) + SIGMA*UU
      RETURN
      END
C
      SUBROUTINE OUTPUT(T,N,X,TRUEX)
      IMPLICIT REAL*8(A-H,O-Z)
      INTEGER T
      DIMENSION X(15,110),TRUEX(15,110)
      L = T
      WRITE(6,100) L,(X(I,L),I=1,N)
  100 FORMAT(1H0,'TIME EQUALS',I3/1X,'FLS ESTIMATES',7X,2D25.10)
      WRITE(6,200) (TRUEX(I,L),I=1,N)
  200 FORMAT(1X,'TRUE X VALUES',7X,2D25.10)
      RETURN
      END
C
C     VALIDATION TEST: HOW WELL DO THE FLS ESTIMATES SATISFY THE
C     FIRST-ORDER CONDITIONS FOR THE COST MINIMIZATION PROBLEM (5)
C
      SUBROUTINE FOCTST(X,YY)
      IMPLICIT REAL*8(A-H,O-Z)
      INTEGER T,TP1,TCAP,TCAP1
      REAL*8 M,MH
      DIMENSION QZERO(15,15),PZERO(15,15),RZERO(15,15)
      DIMENSION XT(15,15),X(15,110),XTT(15,15),E(15,15)
      DIMENSION PZEROT(15,15),EE(15,15),CO(15,15),YT(15,15),YY(15,110)
      DIMENSION F(15,15),A(15,15),H(15,15),B(15,15),D(15,15),M(15,15)
      DIMENSION Y(15,15),TRUEX(15,110)
      DIMENSION MH(15,15),EM(15,15),EMT(15,15),W(15,15),XTP1(15,15)
      DIMENSION ED(15,15),EDT(15,15),U(15,15),V(15,15),FOCD(15,15)
      C = -1.0D+00
C     FORM THE STATE VECTOR FOR TIME T = 1
      CALL INPUT(AMU,TCAP,N,MOBS,QZERO,PZERO,RZERO,IFLAGR,IFLAGS)
      DO 100 I=1,N
      XT(I,1) = X(I,1)
  100 CONTINUE
C     FORM THE INITIAL INCREMENTAL COST CO = -(X1'Q0 - P0')
      CALL TRANS(N,1,XT,XTT)
      CALL MUL(1,N,N,XTT,QZERO,E)
      CALL TRANS(N,1,PZERO,PZEROT)
      CALL SUB(1,N,E,PZEROT,EE)
      CALL MULCON(1,N,C,EE,CO)
C     DO LOOP FOR THE SEQUENTIAL CHECK OF THE FOC FOR T=1,TCAP
      DO 200 T=1,TCAP
C     FORM THE TIME-T STATE VECTOR XT
      DO 300 I=1,N
      XT(I,1) = X(I,T)
```

```
  300 CONTINUE
C     FORM THE TIME-T OBSERVATION VECTOR YT
      DO 400 J=1,MOBS
      YT(J,1) = YY(J,T)
  400 CONTINUE
      CALL MODEL(T,F,A,H,B,D,M,Y,TRUEX)
C     FORM W = (YT - H(T)XT - B(T))'M(T)H(T)
      CALL MUL(MOBS,MOBS,N,M,H,MH)
      CALL RME(N,MOBS,YT,XT,H,B,EM)
      CALL TRANS(MOBS,1,EM,EMT)
      CALL MUL(1,MOBS,N,EMT,MH,W)
      IF(T.EQ.TCAP) GOTO 600
C     FORM THE TIME-T+1 STATE VECTOR XTP1
      TP1 = T + 1
      DO 500 I=1,N
      XTP1(I,1) = X(I,TP1)
  500 CONTINUE
C     FORM U = AMU*(XTP1 - F(T)XT - A(T))'*D(T)
      CALL RDE(N,XTP1,XT,F,A,ED)
      CALL TRANS(N,1,ED,EDT)
      CALL MUL(1,N,N,EDT,D,E)
      CALL MULCON(1,N,AMU,E,U)
C     FORM V = U*F
      CALL MUL(1,N,N,U,F,V)
      GOTO 800
  600 CONTINUE
      DO 700 I=1,N
      V(1,I) = 0.0D+00
  700 CONTINUE
  800 CONTINUE
C     DETERMINE THE FOC DISCREPANCIES FOR TIME T
C     GIVEN BY FOCD = CO + V + W
      CALL ADD(1,N,CO,V,E)
      CALL ADD(1,N,E,W,FOCD)
C     PRINT OUT THE FOC DISCREPANCIES FOCD FOR TIME T
      WRITE (6,36) T
   36 FORMAT(1H0,'FOC DISCREPANCIES FOR TIME',I3)
      WRITE (6,37) (FOCD(1,I),I=1,N)
   37 FORMAT(1X,13D10.2)
C     UPDATE THE INITIAL INCREMENTAL COST CO
      CALL MULCON(1,N,C,U,CO)
  200 CONTINUE
      RETURN
      END
C
C     SUBROUTINE FOR EVALUATING THE MEASUREMENT SPECIFICATION ERROR
C     EM = (YT - H(T)XT - B(T)) FOR TIME T
C
      SUBROUTINE RME(N,MOBS,YT,XT,H,B,EM)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION YT(15,15),XT(15,15),H(15,15),B(15,15),EM(15,15)
      DIMENSION HX(15,15),HXPB(15,15)
      CALL MUL(MOBS,N,1,H,XT,HX)
      CALL ADD(MOBS,1,HX,B,HXPB)
      CALL SUB(MOBS,1,YT,HXPB,EM)
      RETURN
      END
C
C     SUBROUTINE FOR EVALUATING THE DYNAMIC SPECIFICATION ERROR
C     ED = (XTP1 - F(T)XT - A(T)) FOR TIME T
C
      SUBROUTINE RDE(N,XTP1,XT,F,A,ED)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION XTP1(15,15),XT(15,15),F(15,15),A(15,15),ED(15,15)
      DIMENSION FXT(15,15),FXTPA(15,15)
      CALL MUL(N,N,1,F,XT,FXT)
      CALL ADD(N,1,FXT,A,FXTPA)
      CALL SUB(N,1,XTP1,FXTPA,ED)
      RETURN
      END
```

Acknowledgment

The authors are grateful to two anonymous referees for helpful comments.

References

[1] A. P. Sage and J. L. Melsa, *Estimation Theory With Applications to Communication and Control*. New York: McGraw-Hill, 1971.

[2] N. T. Koussoulas, "Multiobjective optimization in adaptive and stochastic control," in C. T. Leondes, Ed., *Control and Dynamic Systems*, vol. 25. New York: Academic Press, 1987, pp. 55–78.

[3] R. Kalaba and L. Tesfatsion, "Time-varying linear regression via flexible least squares," *Int. J. Comput. Math. with Appl.*, vol. 17, pp. 1215–1245, 1989.

[4] L. Tesfatsion and J. Veitch, "U.S. money demand instability: A flexible least squares approach," *J. Econ. Dyn. Contr.*, vol. 14, pp. 151–173, 1990.

[5] J. H. Dorfman and K. A. Foster, "Estimating productivity changes with flexible coefficients," Working Paper FS 89-41, Division of Agricultural Economics, University of Georgia, July 1989.

[6] R. Kalaba and L. Tesfatsion, "An Organizing Principle for Dynamic Estimation," *J. Optimization Theory and Appl.*, vol. 64, pp. 445–470, 1990.

[7] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans ASME: J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[8] A. J. Viterbi, "Error bounds of convolutional codes and an asymptotically optimal decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 268–269, 1973.

[9] R. E. Larson and J. Peschon, "A dynamic programming approach to trajectory estimation," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 537–540, 1966.

[10] P. Swerling, "First order error propagation in a stagewise smoothing procedure for satellite observations," *J. Astronautical Sci.*, vol. VI, pp. 46–52, 1959.

[11] G. Bierman, *Factorization Methods for Discrete Sequential Estimation*. New York: Academic Press, 1977.

[12] R. Bellman and R. Kalaba, *Dynamic Programming and Modern Control Theory*. New York: Academic Press, 1965.

[13] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. New York: Prentice-Hall, 1979.

Robert Kalaba received the B.A. and Ph.D. degrees from New York University, New York, in 1948 and 1958, respectively.

He has been a Professor of Electrical Engineering at the University of Southern California, Los Angeles, since 1967. He was an electronics technician's mate in the U.S. Navy during World War II, where he was involved in radar and communications. He has cowritten eleven books and more than 500 papers, mostly concerning applied computation.

Leigh Tesfatsion received her Ph.D. in economics, with a minor in mathematics, from the University of Minnesota, Minneapolis, in 1975.

She joined the Department of Economics at the University of Southern California, Los Angeles, in 1975, where she currently holds the rank of full professor. In September 1990 she will join the faculty of the Iowa State University of Science and Technology, Ames, IA, as a professor of economics and mathematics. She has more than 50 publications on topics that include microbased macroeconomic modeling, brokerage and financial intermediation, nonlinear analysis, artificial neural networks and associative memories for parameter estimation, and multicriteria estimation.