Nonlinear Analysis, Theory, Methods & Applications, Vol. 12, No. 6, pp. 599–615, 1988.
Printed in Great Britain

0362-546X/88 $3.00 + .00

EXACT SEQUENTIAL FILTERING, SMOOTHING AND PREDICTION FOR NONLINEAR SYSTEMS

ROBERT KALABA and LEIGH TESFATSION
University of Southern California, University Park, Los Angeles, California 90089, U.S.A.

(*Received in revised form* 25 *February* 1987; *received for publication* 18 *June* 1987)

Key words and phrases: Nonlinear filtering, sequential state estimation, exact updating equations, invariant imbedding.

1. INTRODUCTION

CONSIDER a dynamical process described by an n-dimensional system of nonlinear difference equations

$$x_{t+1} = F(x_t) + \varepsilon_t, \qquad t = 0, \ldots, T-1, \tag{1a}$$

where x_t is an n-dimensional column vector of state variables and ε_t is an n-dimensional column vector of unknown modelling errors, $n \geq 1$. Observations on the state vectors are obtained in the nonlinear form

$$y_t = H(x_t) + \eta_t, \qquad t = 0, \ldots, T, \tag{1b}$$

where y_t is an m-dimensional column vector of observations and η_t is an m-dimensional column vector of unknown observational errors, $m \geq 1$. The modelling and observational errors are believed *a priori* to be close to zero, i.e.

$$\varepsilon_t \approx \mathbf{0}, \qquad t = 0, \ldots, T-1; \tag{1c}$$

$$\eta_t \approx \mathbf{0}, \qquad t = 0, \ldots, T. \tag{1d}$$

The basic problem at hand is assumed to be the reconciliation of theory with observations. Specifically, at each time T the modeller must determine whether or not there exists *any* sequence of state vectors $x_0, \ldots, x_T$ which satisfies the theoretical specifications (1) in an acceptable approximate sense for the realized sequence of observation vectors $y_0, \ldots y_T$. For simplicity, we suppose that the incompatibility of the theoretical specifications (1) with the observation vectors $y_0, \ldots, y_T$ is measured by minimizing a weighted sum of squared residual error terms of the form

$$\sum_{t=0}^{T} |y_t - H(x_t)|^2 + k \sum_{t=0}^{T-1} |x_{t+1} - F(x_t)|^2 \tag{2}$$

with respect to $x_0, \ldots, x_T$, where k is a fixed positive scalar weight and $|\;|$ denotes the usual Euclidean norm.

Minimization of the least-squares cost function (2) can be attempted in a variety of ways. In Sections 2 and 3, below, two algorithms are developed for the exact sequential minimization of the cost function (2) as the duration of the process increases and additional observation vectors are obtained. The first algorithm proceeds via an imbedding on the process length and the final *state* vector. The second algorithm proceeds via an imbedding on the process length and the final *observation* vector. Each algorithm generates optimal (least cost) filtered and smoothed state estimates, together with optimal one-step-ahead state predictions.

Section 4 clarifies the relationship between the exact sequential filtering, smoothing, and prediction (FSP) procedures developed in Sections 2 and 3 and the exact sequential FSP procedure previously developed in Kalaba and Tesfatsion [1]. Briefly, the present procedures focus on the direct updating of the minimized least-squares cost function (2) as the duration of the process increases and additional observations are obtained. Apart from the basic requirement that a unique cost-minimizing solution $(x_0, \ldots, x_T)$ exists, the Section 2 FSP procedure imposes no restrictions on the state function F and the measurement function H, and the Section 3 FSP procedure requires only that $n = m$. In contrast, the exact sequential FSP procedure developed in [1] takes as its starting point the first-order necessary conditions for the minimization of (2), and assumes $n = m = 1$, F is differentiable, and H is the identity function. In Section 4 it is shown that the exact sequential FSP procedure developed in [1] is obtained as a corollary of the present procedures for suitably restricted F and H functions.

In Section 5 the exact sequential FSP procedures developed in Sections 2 and 3 are compared with the Kalman–Bucy and extended Kalman filtering procedures. For *linear* dynamical processes (1), the optimal filter state estimates generated by the Section 2 and Section 3 procedures are shown to be analogous in form to Kalman–Bucy filter state estimates. However, for *nonlinear* dynamical processes (1), the optimal filter state estimates generated by the Section 2 and Section 3 procedures are shown *not* to be analogous in form to extended Kalman filter state estimates.

The final Section 6 discusses the extension of the Section 2 and Section 3 sequential FSP procedures to more generally specified dynamical processes (1) and incompatibility measures (2). Also, the basic conceptual idea underlying the paper—using a cost-of-estimation function to measure the incompatibility of *all* theoretical model specifications with observed data simultaneously—is elaborated. In particular, it is shown how stochastic restrictions on error terms might be incorporated into the cost function in addition to dynamical and observational constraints.

As a historical note, the idea of forming a cost-of-estimation function as a suitably weighted sum of squared observational and dynamical modelling errors was stressed by Sridhar, Bellman, and other associates in a series of studies ([2–4]) focusing on a class of continuous-time nonlinear filtering problems arising in rigid body dynamics. The calculus of variations was originally applied to obtain minimal cost solutions. However, emphasis soon shifted to sequential solutions, and use was made of parallel ideas developed in the theory of invariant imbedding for obtaining solutions of nonlinear two-point boundary value problems as functions of interval length and boundary condition parameters. Since the resulting sequential filtering equations were typically difficult to implement, Sridhar proposed a quadratic approximation technique for solving these equations. Subsequently, Sugisaka and Sagara [5] showed that significant improvements could be obtained using higher-order approximations.

Building on this work, Kalaba and Tesfatsion [1] obtain exact sequential filtering equations for a nonlinear two-point boundary value problem associated with a scalar, discrete-time analog of the basic Sridhar continuous-time filtering problem. A subsequent paper [6] introduces a tabular method for the numerical solution of these filtering equations. A companion paper [7] generalizes the results of [1] to a class of multidimensional discrete-time nonlinear estimation problems.

In another series of studies ([8–10]), the direct updating of criterion functions ("criterion filtering") is stressed as a conceptually and computationally attractive alternative to the updating of state probability distributions for certain classes of adaptive control problems. However, in contrast to the present paper, the criterion filters proposed in these studies are designed to generate (approximately) optimal controls while by-passing explicit state estimation altogether.

2. EXACT SEQUENTIAL FSP WITH IMBEDDING ON TIME AND FINAL STATE

In the present section it is shown how a sequence of auxiliary functions $\phi_0(x_1)$, $\phi_1(x_2), \ldots$ can be associated with any filtering, smoothing, and prediction problem characterized by the model equations (1) and the least-squares cost function (2). For each $t \geq 0$, the auxiliary function $\phi_t(x_{t+1})$ measures the cost of the estimation process as a function of the process duration time t and the state estimate x_{t+1} for time $t+1$. The derivation of the auxiliary functions $\phi_0, \ldots, \phi_{T-1}$ for times 0 through $T-1$ yields, as a by-product, an optimal (least cost) filter estimate for the state vector x_T, optimal smoothed estimates for the state vectors $x_0,, \ldots, x_{T-1}$, and an optimal one-step-ahead prediction for the state vector x_{T+1}.

Let $C_{t-1}(x_0, \ldots, x_t)$ denote the cost of the estimation process starting at time 0 and extending through time $t-1$, conditioned on the state estimates $x_0, \ldots, x_t$. Thus,

$$C_{t-1}(x_0, \ldots, x_t) \equiv \sum_{s=0}^{t-1} |y_s - H(x_s)|^2 + k \sum_{s=0}^{t-1} |x_{s+1} - F(x_s)|^2 \tag{3}$$

for $t \geq 1$. Also, let $\phi_{t-1}(x_t)$ denote the *smallest* cost of the estimation process starting at time 0 and extending through time $t-1$, conditioned on the state estimate x_t. Thus,

$$\phi_{t-1}(x_t) \equiv \inf_{x_0, \ldots, x_{t-1}} C_{t-1}(x_0, \ldots, x_t) \tag{4}$$

for $t \geq 1$.

By construction, each cost function C_{t-1} in (3) is bounded below over its domain $R^{n[t+1]}$. It follows from the principle of iterated infima [11, p. 42] that the functions ϕ_{t-1} in (4) satisfy the recurrence relation

$$\phi_t(x_{t+1}) = \inf_{x_t} \{|y_t - H(x_t)|^2 + k|x_{t+1} - F(x_t)|^2 + \phi_{t-1}(x_t)\} \tag{5a}$$

for all x_{t+1} in R^n and $t \geq 0$, where the recurrence relation is initialized by defining

$$\phi_{-1}(x_0) \equiv 0 \tag{5b}$$

for all x_0 in R^n. Thus, the functions $\phi_0(x_1)$, $\phi_1(x_2), \ldots$ are determined one after the other. In practice, a grid could be introduced for $x_1, x_2, \ldots$, and the infima could be found by direct search. Alternatively, use might be made of the first-order necessary conditions associated with the successive minimizations. [See Section 4, below.]

The recurrence relation (5) can be given a dynamic programming interpretation. The bracketed term on the right-hand side of (5a) gives the total cost of the estimation process at time t, conditioned on the state estimates x_t and x_{t+1}. This cost is decomposed into three parts.

The first part, $|y_t - H(x_t)|^2$, denotes the penalty imposed if the time-t state estimate x_t fails to satisfy the basic observational equations (1b) and (1d). The second part, $k|x_{t+1} - F(x_t)|^2$, denotes the penalty imposed if the time-t state estimate x_t and the time-$(t + 1)$ state estimate x_{t+1} fail to satisfy the basic dynamical equations (1a) and (1c). The third part, $\phi_{t-1}(x_t)$, denotes the minimum total penalty incurred over times 0 through $t - 1$ for all deviations from equations (1), given that the time-t state estimate is x_t. Thus, $\phi_t(x_{t+1})$, the infimum of the bracketed term with respect to x_t, yields the minimum total penalty incurred over times 0 through t for all deviations from equations (1), given the time-$(t + 1)$ state estimate is x_{t+1}.

Suppose the least-squares cost function (2) is minimized by a unique sequence of state vectors for each process length $T \geqslant 0$. Then the following exact sequential procedure can be used to generate the minimizing value $\hat{x}(T|T)$ for the state vector x_T at time T for each process length $T \geqslant 0$.

In storage at time $T = 0$ is the function $\phi_{-1}(x_0) \equiv 0$. A first observation vector y_0 is obtained. The optimal filter estimate $\hat{x}(0|0)$ for the state vector x_0 at time 0, based on the single observation vector y_0, is found by minimizing the expression

$$\{|y_0 - H(x_0)|^2 + \phi_{-1}(x_0)\} \tag{6}$$

with respect to x_0. Formally,

$$\hat{x}(0|0) = \arg\left[\min_{x_0}\{|y_0 - H(x_0)|^2 + \phi_{-1}(x_0)\}\right]. \tag{7}$$

In preparation for the next time 1, determine and store for each x_1 in R^n the function value

$$\phi_0(x_1) \equiv \min_{x_0}\{|y_0 - H(x_0)|^2 + k|x_1 - F(x_0)|^2 + \phi_{-1}(x_0)\}. \tag{8}$$

In storage at time $T \geqslant 0$ is the function $\phi_{T-1}(x_T)$, defined for all x_T in R^n. An additional observation vector y_T is obtained. The optimal filter estimate $\hat{x}(T|T)$ for the state vector x_T at time T, based on the observation vectors $y_0, \ldots, y_T$, is then given by

$$\hat{x}(T|T) = \arg\left[\min_{x_T}\{|y_T - H(x_T)|^2 + \phi_{T-1}(x_T)\}\right]. \tag{9}$$

In preparation for the next time $T + 1$, determine and store for each x_{T+1} in R^n the function value

$$\phi_T(x_{T+1}) \equiv \min_{x_t}\{|y_T - H(x_T)|^2 + k|x_{T+1} - F(x_T)|^2 + \phi_{T-1}(x_T)\}. \tag{10}$$

It is easily established that (9) does yield the optimal (least cost) filter estimate for the time-T state vector x_T. By assumption, the total cost of the estimation process at time T, conditioned on the state estimates $x_0, \ldots, x_T$, is measured by the cost function (2). This cost function can equivalently be written as

$$|y_T - H(x_T)|^2 + C_{T-1}(x_0, \ldots, x_T), \tag{11}$$

where the function C_{T-1} is defined as in (3) for $t = T$. The simultaneous minimization of the cost function (2) with respect to the state vectors $x_0, \ldots, x_T$ thus yields

$$\begin{aligned}\min_{x_0,\ldots,x_T} &[|y_T - H(x_T)|^2 + C_{T-1}(x_0,\ldots,x_T)]\\ &= \min_{x_T}\left[|y_T - H(x_T)|^2 + \min_{x_0,\ldots,x_{T-1}} C_{T-1}(x_0,\ldots,x_T)\right]\\ &= \min_{x_T}\{|y_T - H(x_T)|^2 + \phi_{T-1}(x_T)\},\end{aligned} \tag{12}$$

which is precisely the term in large square brackets in (9).

The recurrence relation (10) can also be used to generate optimal smoothed state estimates. For example, consider the problem of obtaining the optimal smoothed estimate for the time-T state vector x_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 function value $\phi_T(x_{T+1})$ in (10) is calculated and stored for each x_{T+1} in R^n. As a by-product of this calculation, one obtains the minimizing x_T as a function of x_{T+1}. Suppose this functional relationship is also stored. Let this functional relationship be denoted by

$$x_T = S_T(x_{T+1}). \tag{13}$$

From (9), the optimal filter estimate $\hat{x}(T+1|T+1)$ for the state vector x_{T+1} at time $T+1$ is given by

$$\hat{x}(T+1|T+1) = \arg\left[\min_{x_{T+1}}\{|y_{T+1} - H(x_{T+1})|^2 + \phi_T(x_{T+1})\}\right]. \tag{14}$$

As detailed in Section A.1 of the appendix, the optimal smoothed estimate for the time-T state vector x_T, based on the observation vectors $y_0,\ldots,y_{T+1}$ for times 0 through $T+1$, is then given by

$$\hat{x}(T|T+1) = S_T(\hat{x}(T+1|T+1)). \tag{15}$$

More generally, given any fixed time t, $0 \le t \le T$, the optimal smoothed estimate $\hat{x}(t|T+1)$ for the time-t state vector x_t based on the observation vectors $y_0,\ldots,y_{T+1}$ for times 0 through $T+1$ is found as follows. In storage at time $T+1$ are the previously calculated functions

$$\begin{aligned} x_t &= S_t(x_{t+1});\\ \vdots\ &\quad \vdots\\ x_T &= S_T(x_{T+1}),\end{aligned} \tag{16}$$

together with the function $\phi_T(x_{T+1})$. The optimal filter estimate $\hat{x}(T+1|T+1)$ for the state vector x_{T+1} at time $T+1$ is determined as in (14). The optimal smoothed estimate $\hat{x}(t|T+1)$ for the state vector x_t at time t is then found by solving equations (16) for x_t in reverse order, starting from the initial condition $x_{T+1} = \hat{x}(T+1|T+1)$.

Finally, the optimal one-step-ahead prediction $\hat{x}(T+1|T)$ for the state vector x_{T+1} at time $T+1$ based on the observation vectors $y_0,\ldots,y_T$ for times 0 through T is determined as follows. Since no observation vector for time $T+1$ is yet available, the selection of a predicted value for x_{T+1} only incurs a cost if this predicted value deviates from the value generated directly through the dynamical equations (1a) and (1c). By assumption, this cost takes the form

$$|x_{T+1} - F(\hat{x}(T|T))|^2. \tag{17}$$

Thus, the optimal one-step-ahead prediction for x_{T+1} is given by

$$\hat{x}(T+1|T) = F(\hat{x}(T|T)). \tag{18}$$

3. EXACT SEQUENTIAL FSP WITH IMBEDDING ON TIME AND FINAL OBSERVATION

In the present section it is shown how an alternate sequence of auxiliary functions $\psi_0(\beta_0)$, $\psi_1(\beta_1), \ldots$ can be associated with any filtering, smoothing, and prediction problem characterized by the model equations (1) and the least-squares cost function (2) for the special case $n = m$. For each $T \geq 0$, the auxiliary function $\psi_T(\beta_T)$ measures the cost of the estimation process as a function of the process duration time T and the parameterized time-T observation vector β_T. As in the previous section, these auxiliary functions can be used to generate optimal (least cost) filtered and smoothed state estimates, together with optimal one-step-ahead state predictions.

Let $J_T(x_0, \ldots, x_T, \beta_T)$ denote the cost of the estimation process starting at time 0 and extending through time T, conditioned on the state estimates $x_0, \ldots, x_T$ and the parameterized time-T observation vector β_T. Thus,

$$J_T(x_0, \ldots, x_T, \beta_T) \equiv \sum_{t=0}^{T-1} |y_t - H(x_t)|^2 + |\beta_T - H(x_T)|^2 + k \sum_{t=0}^{T-1} |x_{t+1} - F(x_t)|^2 \tag{19a}$$

for $T \geq 1$, with

$$J_0(x_0, \beta_0) \equiv |\beta_0 - H(x_0)|^2. \tag{19b}$$

Note that $J_T(x_0, \ldots, x_T, \beta_T)$ reduces to the least-squares cost function (2) when $\beta_T = y_T$.

Also, let $\psi_T(\beta_T)$ denote the *smallest* cost of the estimation process starting at time 0 and extending through time T, conditioned on the time-T observation vector β_T. Thus,

$$\psi_T(\beta_T) \equiv \inf_{x_0, \ldots, x_T} J_T(x_0, \ldots, x_T, \beta_T) \tag{20}$$

for $T \geq 0$.

Suppose $n = m$, and the infimum in (20) is achieved by a unique sequence of state vectors $x_0, \ldots, x_T$ for each β_T in R^n and $T \geq 0$. Let this minimizing sequence of state vectors be denoted by

$$x(0, \beta_T, T), x(1, \beta_T, T), \ldots, x(T, \beta_T, T). \tag{21}$$

By construction, $x(t, \beta_T, T)$ denotes the optimal estimate for the time-t state vector x_t when the duration of the estimation process is T and the time-T observation vector takes on the value β_T, $0 \leq t \leq T$. Suppose, in addition, that $x(T, \beta_T, T)$ is a one-to-one onto function of β_T for each $T \geq 0$.

As detailed in Section A.2 of the appendix, it now follows by the principle of iterated infima [11, p. 42] that the cost functions $\psi_T(\beta_T)$ defined by (20) satisfy the recurrence relation

$$\begin{aligned}\psi_{T+1}(\beta_{T+1}) = \min_{\beta_T} \Big[&\psi_T(\beta_T) + 2[\beta_T - H(x(T, \beta_T, T))]'[y_T - \beta_T] + |y_T - \beta_T|^2 \\ &+ \min_{x_{T+1}} \left(|\beta_{T+1} - H(x_{T+1})|^2 + k|x_{T+1} - F(x(T, \beta_T, T))|^2 \right) \Big]\end{aligned} \tag{22}$$

for $T \geq 0$, where the prime denotes transpose. Using (22), the following exact sequential procedure can be constructed for updating the optimal filter estimate $\hat{x}(T|T)$ for the time-T state vector x_T as the length of the process increases from T to $T+1$ and an additional observation vector y_{T+1} is obtained.

At time $T = 0$, calculate and store the function values $x(0, \beta_0, 0)$ and $\psi_0(\beta_0)$ for each β_0 in R^n. A first observation vector y_0 is obtained. Output the optimal filter estimate for the state vector x_0 at time 0, given by

$$\hat{x}(0|0) = x(0, y_0, 0). \tag{23}$$

At time $T \geq 0$, in storage are the function values $x(T, \beta_T, T)$ and $\psi_T(\beta_T)$ for each β_T in R^n. For each β_{T+1} in R^n, calculate and store the function value $\psi_{T+1}(\beta_{T+1})$ determined by the cost recurrence relation (22). As a by-product of this calculation, one obtains the minimizing β_T as a function of β_{T+1}. Let this functional relationship be denoted by

$$\beta_T = R_T(\beta_{T+1}). \tag{24}$$

For each β_{T+1} in R^n, calculate and store the optimal estimate $x(T+1, \beta_{T+1}, T+1)$ for x_{T+1} as a function of β_{T+1}, found by solving the inner minimization with respect to x_{T+1} in (22) after setting β_T equal to $R_T(\beta_{T+1})$. An additional observation vector y_{T+1} is now obtained. Output the optimal filter estimate for the state vector x_{T+1} at time $T+1$, given by

$$\hat{x}(T+1|T+1) = x(T+1, y_{T+1}, T+1). \tag{25}$$

Optimal smoothed state estimates can be derived as a direct by-product of this algorithm. For example, consider the problem of obtaining the optimal smoothed estimate for the time-T state vector x_T as the duration of the process increases from T to $T+1$ and an additional observation vector is obtained. As previously noted, in storage at time T is the function value $x(T, \beta_T, T)$ for each β_T in R^n. In preparation for time $T+1$, the function value $\psi_{T+1}(\beta_{T+1})$ in (22) is calculated and stored for each β_{T+1} in R^n. As a by-product of this calculation, one obtains the minimizing β_T as a function $R_T(\beta_{T+1})$ of β_{T+1}. An additional observation vector y_{T+1} is obtained. Then, as shown in Section A.2 of the appendix, the optimal smoothed estimate $\hat{x}(T|T+1)$ for the time-T state vector x_T based on the observation vectors $y_0, \ldots, y_{T+1}$ for times 0 through $T+1$ is given by

$$\hat{x}(T|T+1) = x(T, \hat{\beta}_T, T), \tag{26a}$$

where

More generally, given any fixed time t satisfying $0 \leq t \leq T$, the optimal smoothed estimate for the time-t state vector x_t based on the observation vectors $y_0, \ldots, y_{T+1}$ for times 0 through $T+1$ is found as follows. In storage at time $T+1$ are the functions

$$\begin{aligned} \beta_t &= R_t(\beta_{t+1}); \\ \vdots \quad & \quad \vdots \\ \beta_T &= R_T(\beta_{T+1}), \end{aligned} \tag{27}$$

calculated over the previous times t through T. Also in storage at time $T+1$ is the function value $x(t, \beta_t, t)$ for each β_t in R^n, calculated at time t. An additional observation vector y_{T+1} is obtained. Let $\hat{\beta}_t$ denote the value determined for β_t by solving equations (27) in reverse order, starting from the initial condition $\beta_{T+1} = y_{T+1}$. The optimal smoothed estimate $\hat{x}(t|T+1)$ for the time-t state vector x_t based on the observation vectors $y_0, \ldots, y_{T+1}$ is then given by

$$\hat{x}(t|T+1) = x(t, \hat{\beta}_t, t). \tag{28}$$

Finally, as in the previous section, the optimal one-step-ahead prediction $\hat{x}(T+1|T)$ for the state vector x_{T+1} at time $T+1$ based on the observation vectors $y_0, \ldots, y_T$ for times 0 through T is given by

$$\hat{x}(T+1|T) = F(\hat{x}(T|T)). \tag{29}$$

4. RELATIONSHIP TO PREVIOUS RESULTS

The present section clarifies the relationship between the exact sequential FSP procedures developed in Sections 2 and 3 and the exact sequential FSP procedure previously developed in Kalaba and Tesfatsion [1] and numerically tested in Kalaba, Spingarn, and Tesfatsion [6].

The previous paper [1] investigates a special scalar case of the FSP problem outlined in Section 1; namely, the dynamical process (1) is assumed to be one-dimensional ($n = m = 1$), the state function F is assumed to have a well-defined nonvanishing derivative at each point along the optimal (least cost) state sequence, and the measurement function H is assumed to be the identity function. The minimization of the least-squares cost function (2) is considered for an arbitrary process length $T \geqslant 0$. The Euler–Lagrange first-order necessary conditions for this minimization problem are represented as a two-point boundary value problem. The two-point boundary value problem is then converted into an initial value problem via an imbedding on the process duration time T and the time-T observation y_T. Optimal filtered and smoothed state estimates are obtained as a direct by-product in the course of solving this initial value problem.

The sequential FSP procedures developed in Sections 2 and 3 for general dynamical processes (1) do *not* require F to be differentiable or H to be the identity function, and only the Section 3 procedure requires $n = m$. Suppose, however, that $n = m \geqslant 1$, that F is twice continuously differentiable, and that $H(z) = z$ for each z in R^n. Then, given certain additional regularity conditions, the following relationships can be established: (A) The first-order necessary conditions for the sequence of minimizations required by the Section 2 FSP procedure yield, in vector form, the two-point boundary value problem investigated in [1, pp. 1145–1146]; and (B) the first-order necessary conditions for the sequence of minimizations required by the Section 3 FSP procedure yield, in vector form, the initial value problem obtained in [1, pp. 1146–1148].

4.1. *Derivation of relationship (A)*

Consider a dynamical process (1) with dimension $n = m = 1$, differentiable state function F, and measurement function H given by $H(z) = z$ for each z in R. Let $(\hat{x}_0, \ldots, \hat{x}_T)$ denote a state sequence which minimizes the least-squares cost function (2) for this process. As established in [1, (8), p. 1145], the first-order necessary conditions which must be satisfied by this minimizing state sequence have the following two-point boundary value problem representation:

$$0 = 2[\hat{x}_t - y_t] - [\mathrm{d}F(\hat{x}_t)/\mathrm{d}x]\hat{\mu}_{t+1} + \hat{\mu}_t, \quad t = 0, \ldots, T-1; \tag{30a}$$

$$0 = \hat{x}_{t+1} - F(\hat{x}_t) - [1/(2k)]\hat{\mu}_{t+1}, \quad t = 0, \ldots, T-1; \tag{30b}$$

$$0 = \hat{\mu}_0; \tag{30c}$$

$$y_T = \hat{x}_T + [1/2]\hat{\mu}_T, \tag{30d}$$

where $(\hat{\mu}_0, \ldots, \hat{\mu}_T)$ denotes a vector of Lagrange multipliers. It will now be shown that equations (30) can be obtained, in vector form, from the first-order necessary conditions for the sequence of minimizations required by the Section 2 FSP procedure.

Consider a dynamical process (1) with dimension $n = m \geq 1$, and with $H(z) = z$ for all z in R^n. Let $(\hat{x}_0, \ldots, \hat{x}_T)$ denote a state vector sequence which minimizes the least-squares cost function (2) for this process. For any $t = 0, \ldots, T-1$, consider the cost function $C_t(x, x_{t+1})$ defined as in (3) with $x \equiv (x_0, \ldots, x_t)$. By definition (4),

$$\phi_t(x_{t+1}) \equiv \inf_x C_t(x, x_{t+1}) \tag{31}$$

for each x_{t+1} in R^n. Suppose the state function F is twice continuously differentiable at each point $\hat{x}_t$, $t = 0, \ldots, T-1$, so that C_t is twice continuously differentiable at $(\hat{x}, \hat{x}_{t+1})$, where $\hat{x} \equiv (\hat{x}_0, \ldots, \hat{x}_t)$. By assumption, the infimum in (31) is achieved by the vector $\hat{x}$ when $x_{t+1} = \hat{x}_{t+1}$, hence,

$$\partial C_t(\hat{x}, \hat{x}_{t+1})/\partial x = \mathbf{0}'. \tag{32}$$

Finally, suppose the Hessian matrix of C_t with respect to x is nonsingular at $(\hat{x}, \hat{x}_{t+1})$.

It now follows immediately from the envelope theorem [12, p. 160] that

$$\begin{aligned} \partial\phi_t(\hat{x}_{t+1})/\partial x_{t+1} &= \partial C_t(\hat{x}, \hat{x}_{t+1})/\partial x_{t+1} \\ &= 2k[\hat{x}_{t+1} - F(\hat{x}_t)]' \end{aligned} \tag{33}$$

for $t = 0, \ldots, T-1$. Define

$$\begin{aligned} \hat{\mu}_{t+1} &= [\partial\phi_t(\hat{x}_{t+1})/\partial x_{t+1}]' \\ &= 2k[\hat{x}_{t+1} - F(\hat{x}_t)], \quad t = 0, \ldots, T-1. \end{aligned} \tag{34}$$

Recalling that $\phi_{-1}(x_0) \equiv 0$ for all x_0 in R^n, also define

$$\hat{\mu}_0 \equiv [\partial\phi_{-1}(\hat{x}_0)/\partial x_0]' = \mathbf{0}. \tag{35}$$

Equations (34) and (35) are the exact vector analogues of the equations (30b) and the initial boundary condition (30c) appearing in the two-point boundary value problem (30). In particular, the Lagrange multipliers $\hat{\mu}_t$ appearing in (30) are the *derivatives* of the cost functions ϕ_t appearing in the cost recurrence relation (5).

By assumption, the time-t state vector $\hat{x}_t$ satisfies the minimization with respect to $\hat{x}_t$ required by the cost recurrence relation (5) when x_{t+1} is set equal to $\hat{x}_{t+1}$, $t = 0, \ldots, T-1$. Consider the first-order necessary conditions for this sequence of minimizations:

$$\begin{aligned} \mathbf{0}' = {} & 2[\hat{x}_t - y_t]' - 2k[\hat{x}_{t+1} - F(\hat{x}_t)]'[\partial F(\hat{x}_t)/\partial x_t] \\ & + \partial\phi_{t-1}(\hat{x}_t)/\partial x_t, \quad t = 0, \ldots, T-1. \end{aligned} \tag{36}$$

Substituting (34) and (35) into (36), and transposing, one obtains

$$\mathbf{0} = 2[\hat{x}_t - y_t] - [\partial F(\hat{x}_t)/\partial x_t]'\hat{\mu}_{t+1} + \hat{\mu}_t, \quad t = 0, \ldots, T-1. \tag{37}$$

Equations (37) are the exact vector analogue of the equations (30a) appearing in the two-point boundary value problem (30).

Finally, by assumption, the time-T state vector $\hat{x}_T$ satisfies the minimization problem (9). Consider the first-order necessary condition for this minimization:

$$\begin{aligned}\mathbf{0}' &= 2[\hat{x}_T - y_T]' + \partial\phi_{T-1}(\hat{x}_T)/\partial x_T \\ &= 2[\hat{x}_T - y_T]' + \hat{\mu}'_T, \end{aligned} \tag{38}$$

or

$$y_T = \hat{x}_T + [1/2]\hat{\mu}_T. \tag{39}$$

Equation (39) is the exact vector analogue of the terminal boundary condition (30d) appearing in the two-point boundary value problem (30).

Thus, the first-order necessary conditions for the sequence of minimizations required by the Section 2 FSP procedure have generated, in vector form, all of the equations appearing in the two-point boundary value problem (30).

4.2. *Derivation of relationship (B)*

Suppose the time-T observation y_T appearing in the scalar two-point boundary value problem (30) is replaced by an arbitrary parameter β_T in R. Let the state sequence $(x_0, \ldots, x_T)$ which solves this parameterized problem be denoted by

$$(x(0, \beta_T, T), \ldots, x(T, \beta_T, T)). \tag{40}$$

For $T = 0$, set $x(0, \beta_0, 0) = \beta_0$ for each β_0 in R. Finally, let the solution for the time-T state x_T when $\beta_T = y_T$ be denoted by

$$\hat{x}(T|T) \equiv x(T, y_T, T) \tag{41}$$

for $T \geqq 0$.

The following exact sequential procedure is established in [1] for updating the solution $\hat{x}(T|T)$ for the time-T state x_T as the duration of the process increases from T to $T+1$ and an additional observation vector y_{T+1} is obtained.

At time $T = 0$, in storage is the function value $x(0, \beta_0, 0) = \beta_0$ for each β_0 in R. A first observation y_0 is obtained. Output the solution for the state x_0 at time 0, given by $\hat{x}(0|0) \equiv x(0, y_0, 0) = y_0$.

At time $T \geqslant 0$, in storage is $x(T, \beta_T, T)$ for each β_T in R, together with the time-T observation y_T. For each β_T in R, calculate and store:

$$\beta_{T+1} \equiv F(x(T, \beta_T, T)) + [(1+k)/k][\mathrm{d}F(x(T, \beta_T, T))/\mathrm{d}x]^{-1}[\beta_T - y_T]; \tag{42}$$

$$x(T+1, \beta_{T+1}, T+1) \equiv [k/(1+k)]F(x(T, \beta_T, T)) + [1/(1+k)]\beta_{T+1}. \tag{43}$$

An additional observation y_{T+1} is obtained. Output the solution for the state x_{T+1} at time $T+1$, given by $\hat{x}(T+1|T+1) \equiv x(T+1, y_{T+1}, T+1)$.

The β recurrence relation (42) and the state recurrence relation (43) constitute an initial value problem representation for the first-order necessary conditions for minimizing the least-squares cost function (2), given the special assumptions of [1]. It will now be shown that these recurrence relations can both be obtained, in vector form, as by-products of the Section 3 FSP procedure.

Consider a dynamical process (1) with dimension $n = m \geq 1$, and with $H(z) = z$ for each z in R^n. Suppose the time-T observation vector y_T is replaced by an arbitrary vector β_T in R^n. As in Section 3, equations (19) through (21), let $J_T(x, \beta_T)$ denote the cost of the estimation process conditioned on the sequence of state estimates $x = (x_0, \ldots, x_T)$ and the parameterized observation vector β_T; and let the sequence of state estimates which minimizes this cost conditioned on β_T be denoted by

$$x(\beta_T) \equiv (x(0, \beta_T, T), \ldots, x(T, \beta_T, T)). \tag{44}$$

The cost recurrence relation (22) is obtained as in Section 3, with $H(x_{T+1})$ set equal to x_{T+1} by assumption. Carrying out the inner minimization with respect to x_{T+1} required by (22), it is easily verified that one obtains the exact vector analogue of the state recurrence relation (43). Substituting this state recurrence relation back into (22) then yields the reduced form cost recurrence relation

$$\begin{aligned}\psi_{T+1}(\beta_{T+1}) = \min_{\beta_T} [\psi_T(\beta_T) &+ 2[x(T, \beta_T, T) - \beta_T]'[\beta_T - y_T] + |\beta_T - y_T|^2 \\ &+ [k/(1+k)]|\beta_{T+1} - F(x(T, \beta_T, T))|^2].\end{aligned} \tag{45}$$

Suppose the state function F is twice continuously differentiable. Then J_T is twice continuously differentiable at $(x(\beta_T), \beta_T)$, with

$$\partial J_T(x(\beta_T), \beta_T)/\partial x = \mathbf{0}'. \tag{46}$$

Suppose in addition that the Hessian matrix of J_T with respect to x is nonsingular at $(x(\beta_T), \beta_T)$. It then follows by the implicit function theorem that each component of $x(\beta_T)$ is continuously differentiable in a neighborhood of β_T.

The first-order necessary condition for the minimization with respect to β_T required by the cost recurrence relation (45) thus reduces, after some cancellation of terms, to

$$\begin{aligned}\mathbf{0}' = \partial\psi_T(\beta_T)/\partial\beta_T &+ 2[\beta_T - y_T]'[\partial x(\beta_T)/\partial\beta_T] + 2[x(T, \beta_T, T) - \beta_T]' \\ &+ 2[k/(1+k)][\beta_{T+1} - F(x(T, \beta_T, T))]'[-\partial F(x(T, \beta_T, T))/\partial x_T][\partial x(\beta_T)/\partial\beta_T].\end{aligned} \tag{47}$$

By the envelope theorem [12, p. 160],

$$\partial\psi_T(\beta_T)/\partial\beta_T = \partial J_T(x(\beta_T), \beta_T)/\partial\beta_T = -2[x(T, \beta_T, T) - \beta_T]'. \tag{48}$$

Substituting (48) into (47), and transposing, the first-order necessary condition (47) further reduces to

$$\begin{aligned}\mathbf{0} = {} & 2[\partial x(\beta_T)/\partial\beta_T]'[\beta_T - y_T] \\ &+ 2[k/(1+k)][\partial x(\beta_T)/\partial\beta_T]'[-\partial F(x(T, \beta_T, T))/\partial x_T]'[\beta_{T+1} - F(x(T, \beta_T, T))].\end{aligned} \tag{49}$$

Finally, assuming

$$\det[\partial x(T, \beta_T, T)/\partial\beta_T] \neq 0 \quad \text{and} \quad \det[\partial F(x(T, \beta_T, T)/\partial x_T)] \neq 0, \tag{50}$$

the first-order necessary condition (49) can equivalently be expressed as

$$\beta_{T+1} = F(x(T, \beta_T, T)) + [(1+k)/k][\partial F(x(T, \beta_T, T))/\partial x_T]'^{-1}[\beta_T - y_T]. \tag{51}$$

Equation (51) is the exact vector analogue of the β recurrence relation (42). Note, by construction, that equation (51) is an explicit *inverse* representation for the function $\beta_T = R_T(\beta_{T+1})$ described in Section 3, equation (24).

In summary, an exact vector analogue of the state recurrence relation (43) is obtained from the first-order necessary condition for the inner minimization with respect to x_{T+1} required by the basic Section 3 cost recurrence relation (22). Substituting this state recurrence relation back into (22) yields the reduced form cost recurrence relation (45). Given certain additional regularity conditions, the first-order necessary condition for the minimization with respect to β_T required by (45) then yields an exact vector analogue of the β recurrence relation (42).

5. RELATIONSHIP TO KALMAN–BUCY AND EXTENDED KALMAN FILTERING

For explicit comparison of the present sequential filtering procedures with the Kalman–Bucy filtering procedure, suppose the dynamical process (1) takes the simple linear form

$$x_{t+1} = Bx_t + \varepsilon_t, \quad t = 0, \ldots, T-1; \tag{52a}$$

$$y_t = x_t + \eta_t, \quad t = 0, \ldots, T, \tag{52b}$$

where B is a nonsingular $n \times n$ matrix. The Kalman–Bucy filter estimate for the state vector x_{T+1} at time $T+1$ then takes the form

$$\begin{aligned} x^*(T+1|T+1) &= x^*(T+1|T) + K_{T+1}[y_{T+1} - x^*(T+1|T)] \\ &= Bx^*(T|T) + K_{T+1}[y_{T+1} - Bx^*(T|T)] \\ &= [I - K_{t+1}]Bx^*(T|T) + K_{T+1}y_{T+1}, \end{aligned} \tag{53}$$

where the filter gain K_{T+1} is a function of the statistical characteristics assumed for the error terms. (See, e.g., [13, pp. 120–123].)

On the other hand, for the special linear case (52), it can be shown that the optimal estimate $x(T, \beta_T, T)$ for the state vector x_T at time T generated by the Section 3 sequential filtering procedure reduces to

$$x(T, \beta_T, T) = D_T\beta_T + E_T \tag{54a}$$

for all β_T in R^n and $T \geqslant 0$, where

$$D_{T+1} \equiv [1/(1+k)][k^2BD_T[kBBD_T + (1+k)I]^{-1}B + I], \; T \geqslant 0; \tag{54b}$$

$$D_0 \equiv I, \tag{54c}$$

and E_T is a certain time-dependent vector. In addition, equation (51) provides an explicit inverse representation for the function $\beta_T = R_T(\beta_{T+1})$ appearing in (24). Thus, combining (51) and the vector analogue of (43) with (24) through (26), it can be shown that the optimal filter estimate for the state vector x_{T+1} at time $T+1$ is given by

$$\begin{aligned}\hat{x}(T+1|T+1) &= x(T+1, y_{T+1}, T+1)\\ &= [k/(1+k)]Bx(T, \hat{\beta}_T, T) + [1/(1+k)]y_{T+1}\\ &= [k/(1+k)]B\hat{x}(T|T+1) + [1/(1+k)]y_{T+1}\\ &= [I - D_{T+1}]Bx(T, y_T, T) + D_{T+1}y_{T+1}\\ &= [I - D_{T+1}]B\hat{x}(T|T) + D_{T+1}y_{T+1}\end{aligned} \tag{55a}$$

for $T \geqslant 0$, with

$$\hat{x}(0|0) = y_0. \tag{55b}$$

Comparing (53) with (55), it is clear for the linear case that the Kalman–Bucy filter state estimates have the same general structural form as the optimal filter state estimates generated by the Section 3 (equivalently, Section 2) sequential filtering procedure.

However, for a dynamical process (52) with a *nonlinear* state function $F(x_t)$ in place of Bx_t, the extended Kalman filter state estimates do *not* have the same general structural form as the optimal filter state estimates generated by the Section 3 procedure.

The extended Kalman filter estimate for the state vector x_{T+1} at time $T+1$ takes the form

$$\begin{aligned}x^*(T+1|T+1) &= x^*(T+1|T) + N_{T+1}[y_{T+1} - x^*(T+1|T))]\\ &= F(x^*(T|T)) + N_{T+1}[y_{T+1} - F(x^*(T|T))]\\ &= [I - N_{T+1}]F(x^*(T|T)) + N_{T+1}y_{T+1},\end{aligned} \tag{56}$$

where the filter gain N_{T+1} is a function of the statistical properties assumed for the error terms and the nominal state trajectory selected for the state function linearization. (See, e.g., [13, pp. 130–132].) In contrast, again combining (51) and the vector analogue of (43) with (24) through (26), it can be shown that the optimal filter estimate for the state vector x_{T+1} at time $T+1$ generated by the Section 3 procedure takes the form

$$\begin{aligned}\hat{x}(T+1|T+1) &= x(T+1, y_{T+1}, T+1)\\ &= [k/(1+k)]F(x(T, \hat{\beta}_T, T)) + [1/(1+k)]y_{T+1}\\ &= [k/(1+k)]F(\hat{x}(T|T+1)) + [1/(1+k)]y_{T+1}.\end{aligned} \tag{57}$$

Thus, a *filtered* estimate for x_T appears in (56) as the state function argument, whereas a *smoothed* estimate for x_T appears in (57) as the state function argument.

Judging on the basis of (56) and (57), alone, the more complicated structure of the optimal filter (57) would seem to preclude the possibility of a sequential solution, one of the key attractions of the extended Kalman filter (56). However, as demonstrated in Section 3, this is not the case.

6. GENERALIZATIONS

The present paper develops two sequential FSP procedures for measuring the incompatibility of the theoretical dynamical process (1) with observations as the duration of the process increases and additional observations are obtained. For simplicity, incompatibility is measured in terms of the least-squares cost function (2).

The sequential FSP procedure developed in Section 2 permits the observation vectors y_t in (1) to have arbitrary dimension $m \geq 1$, and the state function F to be nonlinear. It follows that parameter estimation problems can be handled by augmentation of the state vectors. The Section 3 FSP procedure also permits the state function F to be nonlinear, but requires $n = m$. Nevertheless, parameter estimation problems can be handled by state augmentation if in addition the observations on the state vectors are suitably augmented by prior information on the parameters, so that the augmented observation vectors have the same dimension as the augmented state vectors. Thus, the Section 2 and Section 3 FSP procedures can be used for both state and parameter estimation.

Also, it is easily verified that the two sequential FSP procedures are formally valid for more general types of dynamical processes and for more general incompatibility measures. For example, the dynamical process (1) can be replaced by a more general dynamical process of the form:

$$F_t(x_{t+1}, x_t) = \varepsilon_t, \quad t = 0, \ldots, T-1; \tag{58a}$$

$$H_t(y_t, x_t) = \eta_t, \quad t = 0, \ldots, T; \tag{58b}$$

$$\varepsilon_t \approx \mathbf{0}, \quad t = 0, \ldots, T-1; \tag{58c}$$

$$\eta_t \approx \mathbf{0}, \quad t = 0, \ldots, T. \tag{58d}$$

The least-squares cost function (2) can be correspondingly generalized to

$$\sum_{t=0}^{T} V_t(H_t(y_t, x_t)) + \sum_{t=0}^{T-1} W_t(F_t(x_{t+1}, x_t)), \tag{59}$$

where $V_t(H)$ and $W_t(F)$ are real-valued functions, strictly monotone increasing with respect to $|H|$ and $|F|$, and satisfying $V_t(\mathbf{0}) = \mathbf{0}$ and $W_t(\mathbf{0}) = \mathbf{0}$. Of course, difficulties with nonexistent or multiple cost-minimizing solutions might limit the usefulness of such generalizations in practice.

The present paper stresses sequential FSP procedures. Nevertheless, the basic conceptual idea underlying the paper concerns initial model design. It is proposed that a modeller form a cost function which imposes a penalty for any deviation away from his theoretical model specifications, whatever form these theoretical model specifications might take. As elaborated in Kalaba and Tesfatsion [7], this basic conceptual idea can be applied to any model design problem for which the basic model specifications are representable as a system of equality and inequality constraints. In particular, stochastic restrictions on error terms can be incorporated into the cost function along with dynamical and observational model specifications.

For example, a prior belief that the error terms ε_t are independent drawings from a distribution with zero mean and finite variance might be handled by replacing restriction (1c) with a restriction on the sample mean of ε_t of the form

$$\left[\sum_{t=0}^{T-1} \varepsilon_t\right] / T \approx \mathbf{0}. \tag{60}$$

The cost function (2) should then be modified to take into account the new model specification (60) together with the previous model specifications (1a), (1b), and (1d). For example, the modified cost function might take the form

$$k_0 \sum_{t=0}^{T} |y_t - H(x_t)|^2 + k_1 \sum_{t=0}^{T-1} |x_{t+1} - F(x_t) - \varepsilon_t|^2 + k_2 \left| \left[\sum_{t=0}^{T-1} \varepsilon_t \right] \Big/ T \right|^2, \tag{61}$$

where k_0, k_1, and k_2 are positive scalar weights, normalized to sum to one.

The cost function (61) is then minimized with respect to both the state variables $(x_0, \ldots, x_T)$ and the error variables $(\varepsilon_0, \ldots, \varepsilon_{T-1})$. How this minimization is to be accomplished, whether by sequential or other methods, is a secondary issue. If all of the model specifications are correct, then, by the strong law of large numbers, the minimized value of this cost function will almost surely be close to zero for sufficiently large T. If some of the model specifications are not correct, it should be possible in principle to locate the misspecified equations by experimentation with the normalized weight factors. For example, if the value of (61) can only be brought close to zero when k_1 is relatively small in relation to k_0 and k_2, the indication is that the dynamical equation (1a) is misspecified. Although it is not possible to say with precision how small is small, one can at least make cost comparisons for competing model specifications. Once the basic plausibility of a model has been established, statistical techniques can be used to further refine estimates.

REFERENCES

1. Kalaba R. & Tesfatsion L., An exact sequential solution procedure for a class of discrete-time nonlinear estimation problems, *IEEE Trans. autom. Control* **AC-26**, 1144–1149 (1981).
2. Bellman R., Kagiwada H., Kalaba R. & Sridhar R., Invariant imbedding and nonlinear filtering theory, *J. astronaut. Sci.* **13**, 110–115 (1966).
3. Detchmendy D. & Sridhar R., Sequential estimation of states and parameters in noisy nonlinear dynamical systems, *Trans. A.S.M.E.* 362–368 (1966).
4. Kagiwada H., Kalaba R., Schumitsky A. & Sridhar R., Invariant imbedding and sequential interpolating filters for nonlinear processes, *J. Basic Engng* **91**, 195–200 (1969).
5. Sugisaka M. & Sagara S., A nonlinear filter with higher-order weight functions via invariant imbedding, *Int. J. Control* **21**, 801–823 (1975).
6. Kalaba R., Spingarn K. & Tesfatsion L., A sequential method for nonlinear filtering: numerical implementation and comparisons, *J. Optim. Theor. Applic.* **34**, 541–559 (1981).
7. Kalaba R. & Tesfatsion L., A least-squares model specification test for a class of dynamic nonlinear economic models with systematically varying parameters, *J. Optim. Theor. Applic.* **32**, 538–567 (1980).
8. Tesfatsion L., A new approach to filtering and adaptive control, *J. Optim. Theor. Applic.* **25**, 247–261 (1978).
9. Tesfatsion L., Direct updating of intertemporal criterion functions for a class of adaptive control problems, *IEEE Trans. Syst. Man. Cybernet.* **SMC-9**, 143–151 (1979).
10. Tesfatsion L., A dual approach to Bayesian inference and adaptive control, *Theory Dec.* **14**, 177–194 (1982).
11. Bartle E., *Elements of Real Analysis*. John Wiley, New York (1976).
12. Takayama A., *Mathematical Economics*, Dryden Press, Hinsdale, Ill. (1974).
13. Saridis G., *Self-Organizing Control of Stochastic Systems*, Dekker, New York (1977).

APPENDIX

A.1. Technical notes for Section 2

In analogy to (12), the minimized time-$(T+1)$ least-squares cost function can be expressed as

$$\min_{x_{T+1}} \{|y_{T+1} - H(x_{T+1})|^2 + \phi_T(x_{T+1})\} = |y_{T+1} - H(\hat{x}(T+1|T+1))|^2 + \phi_T(\hat{x}(T+1|T+1)). \tag{A.1}$$

By definition, the optimal smoothed estimate $\hat{x}(T|T+1)$ for the state x_T at time T conditioned on the observation vectors $y_0, \ldots, y_{T+1}$ is the value of x_T obtained when the time-$(T+1)$ least-squares cost function is minimized with respect to $(x_0, \ldots, x_{T+1})$. Substituting (10) into (A.1), it is clear that $\hat{x}(T|T+1)$ coincides with the x_T which satisfies

(10) for $x_{T+1} = \hat{x}(T+1|T+1)$. By definition, the latter value is $S_T(\hat{x}(T+1|T+1))$; hence, $\hat{x}(T|T+1) = S_T(\hat{x}(T+1|T+1))$ as asserted in (15). The extension to (16) is obtained by a simple backwards induction argument.

A.2. Technical notes for Section 3

It will first be shown that the cost recurrence relation (22) is valid. By construction, the function $\psi_T(\beta_T)$ defined by (20) is bounded below by 0 for all β_T in R^n and $T \geq 0$. Thus, for any given β_{T+1} in R^n, it follows by the principle of iterated infima [11, p. 42] that

$$\psi_{T+1}(\beta_{T+1}) = \min_{x_0,\ldots,x_{T+1}} \left(\sum_{t=0}^{T} |y_t - H(x_t)|^2 + |\beta_{T+1} - H(x_{T+1})|^2 + k \sum_{t=0}^{T} |x_{t+1} - F(x_t)|^2 \right)$$

$$= \min_{x_T} \left[\min_{x_0,\ldots,x_{T-1}} \left(\sum_{t=0}^{T} |y_t - H(x_t)|^2 + k \sum_{t=0}^{T-1} |x_{t+1} - F(x_t)|^2 \right) + \min_{x_{T+1}} \left(|\beta_{T+1} - H(x_{T+1})|^2 + k|x_{T+1} - F(x_T)|^2 \right) \right]. \tag{A.2}$$

By assumption, the function $x(T, \beta_T, T)$ defined by (21) is a one-to-one and onto function of β_T. Thus, introducing a change of variable from x_T to β_T, (A.2) can equivalently be expressed as follows:

$$\psi_{T+1}(\beta_{T+1}) = \min_{\beta_T} \left[\min_{x_0,\ldots,x_{T-1}} \left(\sum_{t=0}^{T-1} |y_t - H(x_t)|^2 + |y_T - H(x(T,\beta_T,T))|^2 + k \sum_{t=0}^{T-2} |x_{t+1} - F(x_t)|^2 + k|x(T,\beta_T,T) - F(x_{T-1})|^2 \right) + \min_{x_{T+1}} \left(|\beta_{T+1} - H(x_{T+1})|^2 + k|x_{T+1} - F(x(T,\beta_T,T))|^2 \right) \right]. \tag{A.3}$$

Also, the solution $(x_0, \ldots, x_{T+1})$ to (A.2) is assumed to be unique; hence, the minimizing x_T in (A.2) must also be unique. It follows that the minimizing β_T in (A.3) is unique.

Now let the term $|y_T - H(x(T, \beta_T, T))|^2$ in (A.3) be equivalently expressed as

$$\begin{aligned} &|y_T - H(x(T,\beta_T,T))|^2 \\ &= |y_T - H(x(T,\beta_T,T)) + \beta_T - \beta_T|^2 \\ &= |\beta_T - H(x(T,\beta_T,T))|^2 \\ &\quad + 2[\beta_T - H(x(T,\beta_T,T))]'[y_T - \beta_T] \\ &\quad + |y_T - \beta_T|^2. \end{aligned} \tag{A.4}$$

Substituting (A.4) into (A.3), and recalling definition (19a), (A.3) reduces to

$$\psi_{T+1}(\beta_{T+1}) = \min_{\beta_T} \left[\min_{x_0,\ldots,x_{T-1}} J_T(x_0, \ldots, x_{T-1}, x(T,\beta_T,T), \beta_T) + 2[\beta_T - H(x(T,\beta_T,T))]'[y_T - \beta_T] + |y_T - \beta_T|^2 + \min_{x_{T+1}} \left(|\beta_{T+1} - H(x_{T+1})|^2 + k|x_{T+1} - F(x(T,\beta_T,T))|^2 \right) \right]. \tag{A.5}$$

Finally, by definition of $x(T, \beta_T, T)$,

$$\begin{aligned} &\min_{x_0,\ldots,x_{T-1}} J_T(x_0, \ldots, x_{T-1}, x(T,\beta_T,T), \beta_T) \\ &= J_T(x(0,\beta_T,T), \ldots, x(T-1,\beta_T,T), x(T,\beta_T,T), \beta_T) \\ &= \min_{x_0,\ldots,x_T} J_T(x_0, \ldots, x_T, \beta_T) \\ &\equiv \psi_T(\beta_T). \end{aligned} \tag{A.6}$$

Substituting (A.6) into (A.5), one obtains

$$\psi_{T+1}(\beta_{T+1}) = \min_{\beta_T}\Big[\psi_T(\beta_T) + 2[\beta_T - H(x(T,\beta_T,T))]'[y_T - \beta_T] + |y_T - \beta_T|^2 + \min_{x_{T+1}}\big(|\beta_{T+1} - H(x_{T+1})|^2 + k|x_{T+1} - F(x(T,\beta_T,T))|^2\big)\Big], \tag{A.7}$$

which is the desired recurrence relation (22).

By construction, the smoothed estimate $x(T, \beta_{T+1}, T+1)$ for the time-T state x_T conditioned on the observation vectors $(y_0, \ldots, y_T, \beta_{T+1})$ is the minimizing value for x_T in (A.2). As in Section 3 (24), let $R_T(\beta_{T+1})$ denote the unique value of β_T which satisfies the minimization in (A.7) for the given β_{T+1}. By the above arguments, it follows that the minimizing x_T in (A.2) is given by $x(T, R_T(\beta_{T+1}), T)$; hence,

$$x(T,\beta_{T+1},T+1) = x(T,R_T(\beta_{T+1}),T). \tag{A.8}$$

In particular, letting $\hat{\beta}_T \equiv R_T(y_{T+1})$,

$$\hat{x}(T|T+1) \equiv x(T,y_{T+1},T+1) = x(T,\hat{\beta}_T,T), \tag{A.9}$$

as claimed in (26).

The extension to (28) then follows by a simple backwards induction argument.