Fixed Interval Smoothing: Revisited

Stephen Ralph McReynolds*
General Electric Company, Philadelphia, Pennsylvania 19101

Gaps in the derivation of early fixed-interval smoothers are filled in. Particular attention is paid to the use of the variational approach introduced by Bryson and Frazier. The notion of complementary models introduced by Weinert and Desai is used to provide an immediate derivation of the continuous Rauch-Tung-Striebel (R-T-S) smoother from the continuous sweep solution of the Bryson-Frazier two-point boundary-value problem. The discrete version of the sweep is also derived, which leads to a fundamental simplification of Bryson's discrete algorithm discovered by Bierman. Relatively new transformations of the R-T-S smoother that may offer some computational advantages are also discussed. Computational comparisons of different algorithms are given. A simple derivation of Bierman's smoother for the mixed continuous discrete problem is given.

Nomenclature

A = $2n \times 2n$ matrix defined by Eq. (13), used in the continuous two-point boundary-value problem (TPBVP) [Eq. (14)]
B = $n \times n$ symmetric matrix defined by Eq. (98), appears in the propagation of Λ_i^s [Eq. (97)]; alternative expressions are given by Eqs. (99) and (107)
C = smoother gain in Rauch-Tung-Striebel smoother [Eq. (108)]
D = $n \times n$ matrix defined in Eq. (30), which appears in the differential equation for λ^s [Eq. (29)]
E = expectation operation
F = $n \times n$ dynamic matrix appearing in continuous system dynamical equations (1)
G = $n \times r$ process noise input matrix appearing in continuous system dynamical equations (1)
H = Jacobian of the measurement with respect to the state [Eq. (5)]
I = identity matrix
J = negative log likelihood function [Eq. (9)]
K = Kalman Gain [Eqs. (101) and (103)]
L = $P^F\Phi$ [Eq. (124)]
M = $2n \times 2n$ system matrix, defined by Eq. (18) that appears in differential equation for $P(t)$ defined by Eqs. (16) and (17)
N = covariance of the innovation, given by Eq. (100)
n = dimension of state vector
P = an $n \times n$ covariance of state vector estimate; also $2n \times 2n$ covariance of $[x(t), \lambda(t)]$, defined by Eq. (16)
p = dimension of measurement noise
Q = process noise covariance ($r \times r$)
q = n vector defined by Eq. (121)
R = measurement noise covariance ($p \times p$)
r = dimension of process noise
S = $n \times n$ symmetric matrix, defined by Eq. (11); represents information matrix for a set of measurements
t = time
V = $n \times n$ matrix defined by Eq. (84) that appears in Eq. (83) for λ_i^s
v = measurement noise, dimension = p
w = process noise, dimension = r
x = state vector, dimension = n
Y = $n \times n$ matrix, defined by Eq. (12)
y = r vector defined by Eq. (123)
z = measurement vector, dimension = p
Γ = $n \times r$ matrix that appears in discrete dynamical equation (3) as coefficient of process noise
Φ = transition matrix [Eq. (12)]
λ = adjoint vector
Λ = covariance of adjoint vector
ν = innovation of predicted residual [Eq. (31)]
δ = dirac delta function
Δ = $r \times r$ matrix defined by Eq. (119)
Ψ = $n \times n$ matrix defined by Eq. (92) that appears in backward equation for λ_i^s [Eq. (96)]; an equivalent expression is given by Eq. (93)

Superscripts

D = dual of
F = filtered
P = predicted
S = smoothed
T = transposed
$\tilde{}$ = error in

Subscripts

k = discrete or sampled time
n = final
0 = initial

Introduction

THE modern state-space smoothing began with the famous paper by Bryson and Frazier[1] that addressed linear and nonlinear continuous problems. The Bryson-Frazier smoother computes the solution to the state-space estimation problem addressed by Swerling,[2] Kalman,[3] and Kalman and Bucy.[4] The original Bryson-Frazier smoother, which shall be reviewed here, differs from the smoother algorithms that later appeared in Bryson and Ho.[5,6] The original solution, as pointed out by Saaty,[7] is numerically unstable. The value of the original formulation was to introduce modern control techniques to the solution of the smoother problem. The use of an adjoint vector enabled the smoothing problem to be converted to a two-point boundary-value problem (TPBVP). The solution technique formulated by Bryson and Frazier required the generation of a fundamental solution of a ($2n \times 2n$) (n = dimension of state) set of unstable linear differential equations.

Later, as noted by Fraser,[8] Bryson, in a set of class notes, applied the sweep method[9,10,11] to reformulate the solution that required only $(n^2)/2$ set of stable linear differential equa-

Received June 5, 1989; revision received Oct. 2, 1989.

*Senior Member of Technical Staff, Military and Data Systems. Member AIAA.

tions. In the new formulation, the discrete smoothing problem was also addressed. Curiously, the derivation of the discrete adjoint smoother is absent from the literature.

Prior to the book by Bryson and Ho,[5] the famous paper of Rauch et al.[12] manipulated Bayesian probabilities to derive another smoothing approach (R-T-S) that did not explicitly involve the use of adjoint variables. The algorithm was derived for the discrete case, and invoking the use of a limiting process, the authors stated the smoother algorithm for the continuous problem without proof.

Other approaches to the smoothing problem were addressed. In particular it should be noted that Mayne,[13] Fraser,[8] and Potter and Fraser[14] developed the two-filter approach to the fixed-interval smoother. Fixed-point and fixed-lag smoothers were developed. Many approaches were developed for the nonlinear problem. In this paper we do not wish to address these other contributions, which are deftly summarized in the paper by Meditch.[15]

We also wish to avoid the discussion of factorized methods, i.e., square root, U-D, that are used to enhance numerical stability of both filters and smoothers. The reader interested in these techniques is directed to the work of Bierman[18,19] and Watanabe and Tzafestas.[20]

However, there were several other contributions that are of interest in this paper. First of all there is the paper by Bierman,[16] who addressed the mixed continuous-discrete problem, i.e., continuous dynamics with discrete measurements. The solution has particular interest in that it permits the solution of a practical problem in a way not addressed by other algorithms. However, easily overlooked by the casual reader, Bierman's algorithm provides another solution to the discrete smoothing problem that is significantly more elegant than the discrete "Bryson-Frazier" algorithm in Bryson and Ho.[5]

Another contribution of interest is in the notion of "complementary models" introduced by Weinert and Desai.[17] As we shall see, the notion of complementary models will be used to establish a duality relationship between the adjoint smoothers (developed by Bryson, Frazier, and Bierman) and the R-T-S smoother. This duality relationship will permit one to write down the continuous R-T-S smoother from the continuous Bryson smoother and vice versa.

Other algorithms of interest are based on transformations of the R-T-S smoother, which transform the covariance propagation into sum of squares. Bierman[19] proposed such an algorithm for problems with one process noise component. Although one can handle more components by repeating the rank 1 noise algorithm, this becomes computationally inefficient. A better approach is to adopt the equations developed by Watanabe and Tzafestas.[20]

The basic purposes of this paper are historical, theoretical, and practical. Historically, we wish to review the original Bryson-Frazier algorithm, which was never published in the open literature. Theoretically, we wish to present derivations of the various discrete smoothers, which are absent from the literature. We also wish to show how the use of complementary models can provide a trivial derivation of the continuous R-T-S smoother (also omitted from the literature) based on the continuous Bryson smoother. On the practical side, we will compare the computations involved with various discrete smoothers.

The Problem

The smoothing problem consists of the construction of the best estimate of the state of a system over a time period using all the measurements in that time interval. The areas of application consist usually of nonreal-time applications, as opposed to the real-time applications that are concerned with filters. A typical application of smoothers is in the postflight construction of a trajectory or altitude profile of a spacecraft in order to support the integration of data from a space-borne sensor.

The problem to be considered is the classic smoothing problem for a linear system driven by white noise, given linear measurements corrupted by white noise. We shall address both the continuous and discrete (or sampled) problems.

Assuming that all variables are Gaussian, the optimal solution can be formulated as the problem of calculating the conditional probability of the state vector, given a set of measurements. In the absence of the Gaussian assumption, the algorithms considered will provide an unbiased estimate with minimum variance.

Several theoretical approaches can be used to derive smoothing algorithms. The method employed by Rauch et al.[12] was to employ the manipulation of conditional probabilities. Other early approaches favored a control oriented approach. Among these we can distinguish the "variational approach" employed by Bryson and Frazier and the dynamic programming approach employed by Mayne.[13] In this paper, we shall view all algorithms as derived from the variational approach. The variational approach is distinguished by the presence of Lagrange multipliers, which transforms the problem into a TPBVP. It is hoped that such an approach will provide additional insight into the nature of the solution provided by other theoretical approaches.

Let

$$x(t) = n \text{ dimensional state at time } t \text{ (continuous time)}$$

Let the continuous dynamics be designated by

$$\dot{x}(t) = F(t)x(t) + G(t)w(t)$$

$$F(x) = n \times n \text{ dynamic matrix}$$

$$G(t) = n \times r \text{ noise input matrix}$$

$$w(t) = r \text{ dimensional vector of dynamic white noise}$$

$$E[w(t)] = 0 \tag{1}$$

$$E[w(t)w^T(\tau)] = Q(t)\delta(t-\tau) \tag{2}$$

$$Q(t) = r \times r \text{ positive symmetric matrix}$$

$$\delta = \text{dirac delta function}$$

The continuous measurements are given by

$$z(t) = H(t)x(t) + v(t) \tag{3}$$

$$H(t) = p \times n \text{ measurement matrix}$$

$$v(t) = p \text{ vector of white noise}$$

$$E[v(t)] = 0 \tag{4a}$$

$$E[v(t)v^T(s)] = R(t)\delta(t-s) \tag{4b}$$

$R(t)$ is $p \times p$ position definite symmetric matrix.

Discrete dynamics involving the sampled state vector x_n are given by

$$x_{k+1} = \Phi_k x_k + \Gamma_k w_k$$

$$\Phi_k = n \times n, \qquad \Gamma_k = n \times r \tag{5}$$

w_k is r vector of discrete white noise.

$$E[w_k] = 0$$

$$E[w_j w_k^T] = Q_k \delta_{jk} \tag{6}$$

Q_k is an $r \times r$ symmetric matrix (positive semidefinite).

Discrete measurements are given by

$$z_k = H_k x_k + v_k \tag{7}$$

$H_k = p \times n$ matrix

$v_k = p$ vector of discrete white noise

$$E[v_k] = 0 \tag{8a}$$

$$E[v_k v_j^T] = R_{kj}\delta_{kj} \tag{8b}$$

R_k is a $p \times p$ positive-definite symmetric matrix.

In the discrete filter/smoother problem, a discrete system is given with a set of discrete measurements $\{z_k \mid k = 1, \ldots, N\}$. Let an estimate of the initial state at t_0 be given and denote it by $\hat{x}_0$. The error in this initial estimate is assumed to be Gaussian with zero mean and covariance P_0. The filtered estimate at time t_j is designated by x_j^F, and is defined to be the "best estimate" of x_j given all the measurements $\{z_k \mid k = 1, \ldots, j\}$.

In contrast, the smoothed estimate at t_j, $j \in [0, N]$ is to be defined as the best estimate at t_j using all the measurements $\{z_k \mid k = 1, \ldots, N\}$. The "best estimate," denoted by $\hat{x}_j^s$, represents the maximum likelihood estimate or conditional mean, which are the same for this problem. One wishes also to calculate the respective covariances of these estimates, denoted by P_j^s.

Most of the implemented filters process measurements sequentially forward in time so that they produce the sequence $\{\hat{x}_k^F, P_k^F, k = 1, \ldots, N\}$. In contrast, most smoothing algorithms proceed backward in time producing the sequence $\{\hat{x}_k^s, P_k^s, k = N, \ldots, 1\}$, with the initial smoothed estimate at N being provided by the filtered estimate at N. The reason for the filtering sequence follows from the fact that the a priori in x_0 specifies the initial conditions of the filters. On the other hand, the final conditions of the filter specify the initial conditions of the smoother.

The formulation of the continuous filter/smoother problem is similar to the discrete problem except that continuous measurements and dynamic models replace the discrete models. In the mixed continuous-discrete problem the dynamics are continuous and the measurements are discrete.

Bryson-Frazier Smoother

In the Bryson-Frazier paper, the continuous problem of smoothing was formulated as a maximum likelihood estimation problem, converted to a linear-quadratic control problem. The object of smoothing was to find $x(t)(t_0 \le t \le t_n)$ to minimize the performance function

$$J = \left[x(t_0) - \hat{x}_0\right]^T P_0^{-1}\left[x(t_0) - \hat{x}_0\right] + \int_{t_0}^{t_N}\left[z - H(t)x(t)\right]^T R^{-1}\left[z - H(t)x(t)\right] + w^T Q^{-1} w \, \mathrm{d}t \tag{9}$$

Subject to constraints

$$\dot{x} = F(x)x(t) + G(t)w(t) \tag{10}$$

The similarity between the estimation and control problem has been noted by Kalman[3] and Bryson and Ho[5] (p. 370) and is referred to as a "duality" relationship.

Formally, if the measurement noise and process noise were white noise, the integral J [Eq. (9)] would become infinite. However, as white noise can be formulated as a limiting case of nonwhite noise, we involve the implicit use of such a limiting process in the theory, thus deriving deftly the correct results while avoiding stochastic calculus.

To solve this problem, standard calculus of variation methods used in optimal control were introduced. A set of "adjoint variables" $\lambda(t)$ corresponding to Lagrange multipliers to the constraint given by Eq. (10) were introduced and were used to convert the problem to a TPBVP.

Let us introduce the notation

$$S \triangleq H^T R^{-1} H \tag{11}$$

$$Y \triangleq GQG^T \tag{12}$$

$$A \triangleq \begin{bmatrix} F & Y \\ S & -F^T \end{bmatrix} \tag{13}$$

Then, the TPBVP derived by Bryson and Frazier[5] is

$$\begin{bmatrix} \dot{\hat{x}}^s \\ \dot{\lambda}^s \end{bmatrix} = A \begin{bmatrix} \hat{x}^s \\ \lambda^s \end{bmatrix} + \begin{bmatrix} 0 \\ -H^T R^{-1} z \end{bmatrix} \tag{14}$$

with boundary conditions

$$\lambda^s(t_N) = 0$$

$$\hat{x}^s(t_0) - P_0\lambda(t_0) = \hat{x}(t_0) \tag{15}$$

From these equations, a TPBVP for the expanded covariance

$$\begin{bmatrix} P_{xx}(t) & P_{x\lambda(t)} \\ P_{\lambda x}(t) & P_{\lambda\lambda}(t) \end{bmatrix} \triangleq P(t) \triangleq E\left[\begin{bmatrix} x^s(t) \\ \lambda^s(t) \end{bmatrix} \left[x^{sT}(t) \lambda^{sT}(t)\right]\right] \tag{16}$$

was derived

$$\dot{P} = AP + PA^T + M \tag{17}$$

where

$$M \triangleq \begin{bmatrix} Y & | & 0 \\ 0 & | & S \end{bmatrix} \tag{18}$$

$P(t)$ is a $2n \times 2n$ (symmetric) matrix that must satisfy boundary conditions

$$P_{\lambda x}(t_N) = 0 \tag{19a}$$

$$P_{x\lambda}(t_N) = 0 \tag{19b}$$

$$P_{xx}(t_0) - P_0 P_{x\lambda}^T(t_0) - P_{x\lambda}(t_0)P_0 + \bar{P}_0 P_{\lambda\lambda}(t_0)\bar{P}_0 = P_0 \tag{19c}$$

To solve this problem, a particular solution was introduced, which was shown to be equivalent to the Kalman filter. This provided the optimal smoothed solution for $x(t_N)$ and $P_{xx}(t_N)$, which converted the problem to an initial value problem. Thus the smoothed solution was generated by integrating the $n + 4n^2$ differential equation given by Eqs. (15) and (18) with boundary conditions

$$\lambda(t_N) = 0 \tag{20a}$$

$$\hat{x}^s(t_N) = \hat{x}^F(x_N) \tag{20b}$$

$$P_{xx}(t_N) = P^F(t_N) \tag{20c}$$

$$P_{x\lambda}(t_N) = 0 \tag{20d}$$

$$P_{\lambda\lambda}(t_N) = 0 \tag{20e}$$

This author is not aware of any significant applications of this algorithm, for if one tried to implement it, one would find the combined equations unstable.[7] The significance of this algorithm was the theoretical approach, which led to a two-point boundary problem.

Sweep Solution to the Bryson-Frazier TPBVP

To solve this problem in a more stable manner required the "sweep method" developed by Bryson and McReynolds.[5,9-11] With the sweep method, Bryson was able to attack the discrete smoothing problem as well as the continuous problem. These equations were placed in class notes of courses Bryson taught.

A version appeared in Fraser's 1967 doctoral thesis[8] for both the discrete and continuous smoothing, along with Fraser's derivation of the continuous smoother. Different versions of these smoothers appeared in Bryson and Ho[5] with some printing errors, some of which were corrected in Bryson and Ho.[6]

In the sweep approach, a vector $\hat{x}(t)$ and matrix $P(t)$ are chosen so that 1) the relation $\lambda(t)=P^{-1}(t)[x(t)-\hat{x}(t)]$ or $x(t)-\hat{x}(t)=P(t)\lambda(t)$ is an invariant of the optimization equations (14), and 2) the $P(t_0)=P_0, \hat{x}(t_0)=x_0$. This implies that any solution of the TPBVP must satisfy requirement 1 and conversely, if 1 is satisfied and the proper boundary conditions at the final time are satisfied, the solution will be the optimal smoothed solutions.

To make the equation invariant, we can choose $P(t)=P^F(t), \hat{x}(t)=\hat{x}^F(t), \lambda^F(t)=0$. So one can see that requirement 1 will be satisfied by the particular smoothed solution that ends at t. The important point is that 1 must hold for all smoothed solutions for $t_N \geq t$. Thus we write

$$\lambda^s(t)=P^{F^{-1}}(t)\left[\hat{x}^s(t)-\hat{x}^F(t)\right] \tag{21}$$

Thus generating $\lambda^s(t)$ is key to obtaining $\hat{x}^s(t)$ through the preceding relationship, which can be inverted to become

$$\hat{x}^s(t)=\hat{x}^F(t)+P^F(t)\lambda^s(t) \tag{22}$$

This relation can also be used to generate the smoothed covariance. One needs the property

$$E\left[\tilde{x}^s(t)\tilde{x}^F(t)^T\right]=P^s(t) \tag{23}$$

This implies

$$\text{cov}\left[\hat{x}^F(t)-\hat{x}^{(s)}\right]=P^F(t)-P^s(t) \tag{24}$$

Thus defining

$$\Lambda^s(t)=E\left[\lambda^{(s)}(t)\lambda^{(s)}(t)^T\right] \tag{25}$$

One has, from Eqs. (21) and (24), the relation

$$\Lambda^s(t)=P^{F^{-1}}(t)\left[P^F(t)-P^s(t)\right]P^{F^{-1}} \tag{26}$$

Solving this equation for $P^s(t)$ gives

$$P^s(t)=P^F(t)-P^F(t)\Lambda^s(t)P^F(t) \tag{27}$$

Thus, one can derive $P^s(t)$ once $\Lambda^s(t)$ is obtained [having saved $P^F(t)$].

One should also note an interesting property of $\lambda^s(t)$, noted by Fraser,[8] namely that it is orthogonal to the error in $x^s(t)$.

$$E\left[\lambda^s(t)\tilde{x}^s(t)^T\right]=P^{F^{-1}}(t)\left[E\left(\tilde{x}^s(t)\tilde{x}^s(t)\right)-E\left[\tilde{x}^F(t)\tilde{x}^s(t)\right]\right]=0 \tag{28}$$

To obtain an expression for $\Lambda^s(t)$, one uses Eq. (22) to substitute for x in the differential equation for λ^s obtaining

$$\dot{\lambda}^s=-D(t)\lambda^s(t)-H^TR^{-1}\nu(t) \tag{29}$$

where

$$D(t)\triangleq F(t)-S(t)P^{F^{-1}}(t) \tag{30}$$

$$\nu(t)\triangleq z-H\hat{x}^F(t) \tag{31}$$

$\nu(t)$ is the innovation process that feeds the filter. The $\nu(t)$ is white noise with the statistics

$$E\left[\nu(t)\right]=0 \tag{32}$$

$$E\left[\nu(t)\nu^T(\tau)\right]=R(t)\delta(t-\tau)\Big] \tag{33}$$

The preceding equation for $\lambda^s(t)$ is the linear stochastic differential equation for x where the following duality relationships exist:

$$\lambda^s(t)\leftrightarrow x(t) \tag{34}$$

$$R^{-1}\nu(t)\leftrightarrow\Gamma w(t) \tag{35}$$

To obtain full duality, one must also reverse time, as the differential equation must be integrated backward in time. Thus, introducing

$$\tau=t_N-t \tag{36}$$

The differential equation for λ^s is given by

$$\frac{\mathrm{d}\lambda^s}{\mathrm{d}\tau}=D^T(t)\lambda^s(t)-H^T(t)R^{-1}(t)\nu \tag{37}$$

This leads to the following other dual relationships

$$D^T(t)\leftrightarrow F(t)$$

$$-H^T(t)\leftrightarrow G(t)$$

$$R^{-1}(t)\leftrightarrow Q(t) \tag{38}$$

Using these relationships, one can transform the time prediction formula for the covariance of x, namely,

$$\dot{P}=FP+PF^T+Y \tag{39}$$

to one for $\Lambda^s(t)$

$$\frac{\mathrm{d}\Lambda^s}{\mathrm{d}\tau}(t)=D^T(t)\Lambda^s(t)+\Lambda^s(t)D(t)+S(t) \tag{40}$$

This, and the boundary conditions $\Lambda^s(t_N)=0$ from Eq. (15), generates $\Lambda^s(t)$.

In terms of time t, Eq. (40) becomes

$$\dot{\Lambda}^s=-D^T\Lambda^s-\Lambda^sD-S \tag{41}$$

This equation was carefully derived by Fraser.[8]

The minus sign of the last term seems unusual as there is a plus sign in front of Q in Eq. (39). However, it is clear that a minus sign is necessary to ensure $\Lambda^s(t)\geq 0$. Note that $\Lambda^s(t)=-\Lambda(t)$ [Eq. (13.3.10) in Ref. 6].

Covariance Propagation Fixed-Interval Smoothers

In the covariance propagation approach to fixed-interval smoothing by Rauch et al.[12] the smoothed covariance is propagated directly, without use of adjoint variables. The result was first developed for discrete problems by manipulating conditional probabilities. The formulas for continuous problem, stated but not proved,[12] can be derived by applying a limiting process to the discrete problem. Another approach to proving the continuous result, suggested by Bryson and Ho, is to differentiate Eqs. (22) and (27). A more direct proof is to use the technique used to derive the continuous adjoint smoother.

Using Eq. (21), $\lambda^s(t)$ can be eliminated from the equation for $\hat{x}^s$ to obtain

$$\dot{\hat{x}}^s=F\hat{x}^s+Y(t)P^{F^{-1}}(t)\left[\hat{x}^s(t)-\hat{x}^F(t)\right] \tag{42}$$

To obtain the equation for $\dot{P}^s(t)$, we exploit the derivation of the continuous adjoint smoother. Let us subtract the equation $\dot{x}=Fx+Gw$ from the equation for $\hat{x}^s$ to obtain the equation for $\tilde{x}^s=\hat{x}^s-x$. Thus the joint equations for $\tilde{x}^s$ and $\tilde{\lambda}^s$ become

$$\begin{bmatrix}\dot{\tilde{x}}^s\\ \dot{\tilde{\lambda}}^s\end{bmatrix}=\begin{bmatrix}F & Y\\ S & -F^T\end{bmatrix}\begin{bmatrix}\tilde{x}^s\\ \tilde{\lambda}^s\end{bmatrix}+\begin{bmatrix}-Gw\\ -H^TR^{-1}\nu\end{bmatrix} \tag{43}$$

We note that there is a similarity in the roles of $\tilde{x}^s$ and $\tilde{\lambda}^s$. In fact, let us perform the change $\tilde{x}^s \leftrightarrow \tilde{\lambda}^s$, writing the preceding equations as

$$\begin{bmatrix} \dot{\tilde{\lambda}}^s \\ \dot{\tilde{x}}^s \end{bmatrix} = \begin{bmatrix} -F^T & S \\ Y & F \end{bmatrix} \begin{bmatrix} \tilde{\lambda}^s \\ \tilde{x}^s \end{bmatrix} + \begin{bmatrix} -H^T R^{-1} \nu \\ -Gw \end{bmatrix} \tag{44}$$

Now, if $\tilde{\lambda}^s$ is treated as a state variable, $\tilde{x}^s$ has the form of its adjoint variable. Thus one can construct an estimation problem for which $\tilde{\lambda}^s$ represents the state vectors and $\tilde{x}^s$ represents the adjoint vector. This problem can be constructed by exploiting the "mutually" dual relations

$$\begin{aligned} &\lambda^s \leftrightarrow \tilde{x}^s \\ &F \leftrightarrow -F^T \\ &Y \leftrightarrow S \\ &R^{-1}\nu \leftrightarrow w \\ &R^{-1} \leftrightarrow Q \\ &H^T \leftrightarrow G \end{aligned} \tag{45}$$

This observation was first made by Weinert and Desai,[17] who use these dual relations to derive smoothing results. They referred to this dual estimation problem as "complementary models."

The duality relationship indicated here is distinct from that mentioned earlier with regard to estimation and control. In that duality, there is a unique 1:1 correspondence between quantities in both problems. In this duality relationship, we established a 1:1 correspondence within a problem. This correspondency further more has the "mutually dual" property that if one quantity in the original problem, say Y, is assigned a quantity in the other problem, namely S, then S in the original problem is assigned Y in the complementary problem. Thus, the complementary problem of the complementary problem is the original problem.

Thus, using these dual relationships, the equation for covariance $P^s(t)$ is given by the dual of the equation for $\Lambda^s(t)$. The dual expressions for D are given by

$$\begin{aligned} D^D &= F^D - S^D(P^F)^{-1} \\ &= -F^T(t) - Y(t)P^F(t)^{-1} \end{aligned} \tag{46}$$

Thus,

$$\dot{P}^s = (-D^D)^T P^s - P^s(D^D) - Y \tag{47}$$

This expression, together with Eq. (42) for $\dot{x}^s$ and the boundary conditions, completes the definition of the continuous version of the R-T-S smoother.[12]

Another way of expressing this is as follows: The covariance propagation smoother of the complementary model is equivalent to the adjoint smoother with respect to the original problem, and conversely, the adjoint smoother of the complementary problem is equivalent to the covariance propagation smoother of the original problem.

Derivation of the Discrete Smoothers Using Adjoint Theory

Let the performance index to be minimized be given by

$$\begin{aligned} J_N = \sum_{i=1}^{N} (z_i - H_i x_i)^T R_i^{-1}(z_i - H_i x_i) + w_i^T Q_i^{-1} w_i \\ + (x_0 - \hat{x}_0)^T P_0^{-1}(x_0 - \hat{x}_0) \end{aligned} \tag{48}$$

subject to the dynamic constraints

$$x_{i+1} = \Phi_i x_i + \Gamma_i w_i \tag{49}$$

J_N is the negative of the log likelihood function.

Introducing λ^s_{i+1} as the adjoint variable (Lagrange multiplier) for this latter constraint, the following discrete TPBVP can be derived using variation techniques.[5]

$$\hat{w}^s_i = Q_i \Gamma_i^T \lambda^s_{i+1} \tag{50}$$

$$\lambda^s_i = \Phi_i^T \lambda^s_{i+1} - H_i^T R_i^{-1}(z_i - H_i \hat{x}^s_i) \tag{51}$$

$$\hat{x}^s_{i+1} = \Phi_i \hat{x}^s_i + Y_i \lambda^s_{i+1} \tag{52}$$

$$Y_i = \Gamma_i Q_i \Gamma_i^T \tag{53}$$

$$\lambda^s_0 = P_0^{-1}(\hat{x}^s_0 - \hat{x}_0) \tag{54}$$

$$\lambda^s_{N+1} = 0 \tag{55}$$

The $\hat{x}_0$ is referred to as the a priori estimate of x_0; P_0 is the a priori covariance, i.e., covariance of $\hat{x}_0$. The superscript s denotes the smoothed solution. The smoothed solution at N will be the same as the filtered solution, denoted by superscript F; i.e., $\hat{x}^F_N$ denotes the filtered solution at time N. The solution at $N+1$ will be the one-step prediction, denoted by $\hat{x}^P_{N+1}$.

Written as a function of x_{N+1}, the optimal value of J_N can be written as

$$J^{\text{opt}}_N(x_{N+1}) = (x_{N+1} - \hat{x}^P_{N+1}) P^{P-1}_{N+1}(x_{N+1} - \hat{x}^P_{N+1})$$
$$+ \text{ terms not depending on } x_{N+1}$$

The P^P_{N+1} is the covariance of $\hat{x}^P_{N+1}$. The preceding statement follows from basic rules derived concerning Gaussian distribution and the fact that $\hat{x}^P_{N+1}$ is the maximum likelihood estimate of x_{N+1}.

From variation principles one has

$$\frac{\delta J^{\text{opt}}_N}{\delta x_{N+1}} = 2\lambda^T_{N+1} \tag{56}$$

On the other hand, one has

$$\frac{\delta J^{\text{opt}}_N}{\delta x_{N+1}} = 2(x_{N+1} - \hat{x}^P_{N+1})^T (P^P_{N+1})^{-1} \tag{57}$$

Combine Eqs. (56) and (57):

$$\lambda^s_{N+1} = P^{P-1}_{N+1}(\hat{x}^s_{N+1} - \hat{x}^P_{N+1}) \tag{58}$$

One can indeed see that the optimal condition $\lambda^s_{N+1} = 0$ [Eq. (55)] when $\hat{x}^s_{N+1} = \hat{x}^P_{N+1}$. Now we shall prove that the preceding relationship can be applied to any point on the time grid, i.e.,

$$\lambda^s_k = [P^P_k]^{-1}[\hat{x}^s_k - \hat{x}^P_k] \tag{59}$$

holds for all k, $k \in [0, N+1]$.

Another proof of Eq. (59) will be developed based on the variational equations.

Defining

$$P^P_0 = P_0, \qquad \hat{x}^P_0 = x_0 \tag{60}$$

we see that Eq. (59) is equivalent to the boundary condition given by Eq. (54). Now we shall prove Eq. (59) must hold for all k by induction. Assume

$$\lambda^s_i = [P^P_i]^{-1}(\hat{x}^s_i - \hat{x}^P_i) \tag{61}$$

To simplify the derivation Eq. (51) is modified.

Define

$$\nu_i \triangleq z_i - H_i\hat{x}_i^P \tag{62}$$

ν_i is the predicted residual (or innovation) at i. Using ν_i Eq. (51) can be written as

$$\lambda_i^s = \Phi_i^T\lambda_{i+1}^s - H_i^TR_i^{-1}\nu_i + S_i(\hat{x}_i^s - \hat{x}_i^P) \tag{63}$$

where

$$S_i = H_i^TR_i^{-1}H_i \tag{64}$$

Now, introducing P_i^F (the filtered covariance)

$$P_i^{F^{-1}} \triangleq P_i^{P^{-1}} + S_i \tag{65}$$

and $\hat{x}_i^F$ (the filtered estimate)

$$\hat{x}_i^F \triangleq \hat{x}_i^P + P_i^FH_i^TR_i^{-1}\nu_i \tag{66}$$

Then using Eqs. (61), (65), and (66), Eq. (63) can be written

$$[P_i^F]^{-1}[x_i^s - \hat{x}_i^P] = \Phi_i^T\lambda_{i+1} + [P_i^F]^{-1}[\hat{x}_i^F - \hat{x}_i^P] \tag{67}$$

or

$$\hat{x}_i^s - \hat{x}_i^F = P_i^F\Phi_i^T\lambda_{i+1}^s \tag{68}$$

Multiply Eq. (68) by Φ_i giving

$$\Phi_i x_i^s - \Phi_i x_i^F = \Phi_i P_i^F\Phi_i^T\lambda_{i+1}^s \tag{69}$$

Now define

$$\hat{x}_{i+1}^P \triangleq \Phi_i\hat{x}_i^F \tag{70}$$

Using Eqs. (70) and (52) in Eq. (69)

$$\hat{x}_{i+1}^s - \hat{x}_{i+1}^P = [\Phi_iP_i^F\Phi_i^T + Y_i]\lambda_{i+1}^s \tag{71}$$

Thus defining

$$P_{i+1}^P \triangleq \Phi_iP_i^F\Phi_i^T + Y_i \tag{72}$$

Eq. (71) is Eq. (59) for $k = i+1$. This completes the induction.

It should be recognized that Eqs. (62), (65), (66), (70), and (72) are equivalent to the discrete Kalman-Bucy[4] filter. Eq. (65) is the Gaussian least-squares update of the covariance, originally suggested for sequential processing by Swerling.[2]

The preceding equations can be used to derive various smoothers. Eq. (59) is the key equation. To derive the discrete version of the adjoint smoother, it can be used to derive $\hat{x}_k^s$ from knowledge of λ_k^s

$$\hat{x}_k^s = \hat{x}_k^P + P_k^P\lambda_k^s \tag{73}$$

To develop a formula for the smoothed covariance, we introduce the covariance of λ_k^s:

$$\Lambda_k^s \triangleq E\left[\lambda_k^s(\lambda_k^s)^T\right] \tag{74}$$

Now, transposing Eq. (59) gives

$$\lambda_k^{sT} = (\hat{x}_k^s - \hat{x}_k^P)^T[P_k^P]^{-1} \tag{75}$$

Multiplying Eq. (59) by Eq. (75) gives

$$\lambda_k^s\lambda_k^{s^T} = (P_k^P)^{-1}(\hat{x}_n^s - \hat{x}_k^P)(\hat{x}_k^s - \hat{x}_k^P)^T(P_k^P)^{-1} \tag{76}$$

We wish to take expectations of Eq. (76) using the relationship [see Eq. (24)]

$$E\left[(\hat{x}_k^s - x_k^P)(\hat{x}_k^s - \hat{x}_k^P)^T\right] = P_k^P - P_k^s \tag{77}$$

Thus, taking expectations of Eq. (76), using Eqs. (74) and (77), gives

$$\Lambda_k^s = (P_k^P)^{-1}(P_k^P - P_k^s)(P_k^P)^{-1} \tag{78}$$

Solving this equation for P_k^s gives

$$P_k^s = P_k^P - P_k^P\Lambda_k^sP_k^P \tag{79}$$

The discrete adjoint smoother consists of developing propagation equations for λ_k and Λ_k and using Eqs. (73) and (79). To obtain $\hat{x}_k^s$ and P_k^s initially

$$\lambda_{N+1} = 0$$

$$\Lambda_{N+1} = 0 \tag{80}$$

Using Eq. (73) in Eq. (63)

$$\lambda_i^s = \Phi_i^T\lambda_{i+1}^s - S_iP_i^P\lambda_i^s + H_i^TR_i^{-1}\nu_i \tag{81}$$

$$V_i\lambda_i^s = \Phi_i^T\lambda_{i+1}^s + H_i^TR_i^{-1}\nu_i \tag{82}$$

$$\lambda_i^s = V_i^{-1}(\Phi_i^T\lambda_{i+1}^s + H_i^TR_i^{-1}\nu_i) \tag{83}$$

where

$$V_i \triangleq I + S_iP_i^P \tag{84}$$

Then from Eq. (83)

$$\lambda_i^s = V_i^{-1}\Phi_i^T\lambda_{i+1}^s + V_i^{-1}H_i^TR^{-1}\nu_i \tag{85}$$

To compute V_i^{-1}, we use the identity

$$[P_i^F]^{-1} = S_i + [P_i^P]^{-1} \tag{86}$$

Multiplying this equation on the right by P_i^P and using Eq. (84) gives

$$[P_i^F]^{-1}P_i^P = S_iP_i^P + I = V_i \tag{87}$$

Thus

$$V_i^{-1} = [P_i^P]^{-1}P_i^F \tag{88}$$

Multiplying Eq. (86) on the right by P_i^F gives

$$I = S_iP_i^F + [P_i^P]^{-1}P_i^F \tag{89}$$

or

$$V_i^{-1} = I - S_iP_i^F = I - H_i^TK_i^T \tag{90}$$

Now substituting Eq. (90) for V_i^{-1} in Eq. (83)

$$\lambda_i = (I - S_iP_i^F)(\Phi_i^T\lambda_{i+1} + H_i^TR_i^{-1}\nu_i) \tag{91}$$

This is equivalent to Eq. (13.2.9) in Bryson and Ho.[5]

To derive an expression for the propagation for Λ_i, it is convenient to introduce the modified transition

$$\Psi_i \triangleq \Phi_i(I - P_i^FS_i) \tag{92}$$

Equivalently,

$$\Psi_i = \Phi_i(I - K_iH_i) \tag{93}$$

Then Eq. (83) can be written

$$\lambda_i^s = \Psi_i^T \lambda_{i+1}^s + V_i^{-1} H_i R_i^{-1} \nu_i \tag{94}$$

One can also show

$$V_i^{-1} H_i^T R_i^{-1} = (I - S_i P_i^F) H_i^T R_i^{-1} = H_i^T R_i^{-1} - S_i K_i$$
$$= (P_i^F)^{-1} K_i - S_i K_i = (P_i^P)^{-1} K_i$$
$$= H_i^T N_i^{-1} \tag{95}$$

Thus, Eq. (94) can be written as

$$\lambda_i^s = \Psi_i^T \lambda_{i+1}^s - H_i^T N_i^{-1} \nu_i \tag{96}$$

Eq. (96) was derived by Bierman.[16]

Eqs. (94) or (96) imply

$$\Lambda_i^s = \Psi_i^T \Lambda_{i+1}^s \Psi_i + B_i \tag{97}$$

where

$$B_i \triangleq V_i^{-1} H_i^T R_i^{-1} N_i R_i^{-1} H_i (V_i^{-1})^T \tag{98}$$

or

$$B_i = H_i^T N_i^{-1} H_i \tag{99}$$

where

$$N_i \triangleq Cov(\nu_i) = H_i P_i^P H_i^T + R_i \tag{100}$$

Eq. (97), using Eq. (99) to define B_i, and Eq. (93) to define Ψ_i, was derived by Bierman.[16]

Eq. (100) can be used to derive the original Bryson smoother. We recognize that

$$P_i^F H_i^T R_i^{-1} = K_i \qquad \text{(Kalman and Bucy)}^4 \tag{101}$$

appears in Eq. (98). Thus, Eq. (98), with the help of Eq. (88) becomes

$$B_i = [P_i^P]^{-1} K_i N_i R_i^{-1} H_i [I - P_i^F S_i] \tag{102}$$

Now one also has

$$K_i = P_i^P H_i^T N_i^{-1} \tag{103}$$

Thus, substituting further

$$B_i \triangleq S_i (I - P_i^F S_i) \tag{104}$$

With this definition of B_i, and Eq. (92) to define Ψ_i, Eq. (97) becomes Eq. (13.2.12) in Bryson and Ho.[6]

To obtain a covariance smoother, one can substitute Eq. (59) into Eq. (52) to obtain

$$\hat{x}_{i+1}^s = \Phi_i \hat{x}_i^s + Y_i P_{i+1}^{P^{-1}} (\hat{x}_{i+1}^s - \hat{x}_{i+1}^P) \tag{105}$$

subtracting $\hat{x}_{i+1}^P = \Phi_i \hat{x}_i^F$ gives the relation

$$\hat{x}_{i+1}^s - \hat{x}_{i+1}^P = \Phi_i (\hat{x}_i^s - x_i^F) + Y_i P_{i+1}^{P^{-1}} (\hat{x}_{i+1}^s - \hat{x}_{i+1}^P) \tag{106}$$

Solving for $\hat{x}_i^s - x_i^F$ gives

$$\hat{x}_i^s - \hat{x}_i^F = C_i (\hat{x}_{i+1}^s - \hat{x}_{i+1}^P) \tag{107}$$

where

$$C_i = \Phi_i^{-1} (I - Y_i P_{i+1}^{P^{-1}}) \tag{108}$$

Now the covariance of the left side is $P_i^F - P_i^s$ and the covariance of $\hat{x}_{i+1}^s - \hat{x}_{i+1}^P$, which appears on the right side, is $P_{i+1}^P - P_{i+1}^s$.

Thus Eq. (107) implies the relationship

$$P_i^s = P_i^F - C_i (P_{i+1}^P - P_{i+1}^s) C_i^T \tag{109}$$

Another derivation for the covariance smoother is by using Eq. (59) in Eq. (68) giving

$$C_i = P_i^F \Phi_i^T P_{i+1}^{P^{-1}} \tag{110}$$

Eqs. (107), (109), and (110) correspond to the classic discrete R-T-S smoother.

Mixed Discrete-Continuous Smoothing

Many applications are of a mixed type: discrete measurements and continuous dynamics. In this case, a rigorous mixed smoother is useful particularly when it is necessary to obtain a high data rate of smoother output in comparison to the input data rate. Bierman[16] derived the analogue of the adjoint smoother for the mixed problem from the discrete R-T-S smoother. The form of this smoother is a solution of differential equations for λ and Λ, with discrete updates at the measurement times. The formulas for the differential equations can be more directly obtained from the continuous adjoint smoother, i.e., with $S = 0$ giving

$$\dot{\lambda} = -F^T \lambda \tag{111}$$

$$\dot{\Lambda} = -F^T \Lambda - \Lambda F \tag{112}$$

These equations are particularly attractive in that they do not involve inverting P^F, as in the continuous problem.

The discrete updates at the measurements can be obtained from the discrete update formulas setting $\Phi_k = I_n$ (the $n \times n$ identity matrix). Using the Bierman form of the smoother [Eqs. (92), (96), (97), (99)] gives

$$\lambda_k^- = (I - K_k H_k) \lambda_k^+ + H_k^T N_k^{-1} \nu_k \tag{113}$$

$$\Lambda_k^- = (I - K_k H_k) \Lambda_k^+ (I - K_k H_k)^T + H_k^T N_k^{-1} H_k \tag{114}$$

A covariance smoother for the same problem will yield the same R-T-S smoother as for the continuous problem. There are no discontinuities in P^S and $\hat{x}^s$ at measurement points, but the derivatives are discontinuous. It is desirable to allow for these discontinuities in integrating the differential equations.

Alternate Forms of the Discrete Rauch-Tung-Striebel Smoother

In this section alternate forms of the R-T-S smoother that have potential numerical advantages are discussed. Bierman[19] developed an algorithm that is particularly suited for systems with a single component of process noise. In this form the propagated covariance is written as a sum of squares, which, as Bierman's information solution, avoids the problem of differencing. This also reduces computation exploiting the lower dimension of the process noise. The numerical efficiency of this algorithm loses its advantages with higher-order process noise because it must be repeated once for every process noise component. To handle multiple process noise components, it is more efficient to exploit a generalization of the Bierman's basic formula to p dimensions developed by Watanabe.[20] Adopting Watanabe's notation, the smoother backward recursion relation is performed in two steps. (The notation in this section is inconsistent with other sections.)

$$\tilde{x}_{k/N}^s = (L_k) \hat{x}_{n+1}^s + \Gamma_k \Delta_k y_k \tag{115}$$

$$\hat{x}_k^s = \Phi_k^{-1} \tilde{x}_k^s \tag{116}$$

Table 1 Comparison of discrete adjoint smoothers for the case of a single scalar measurement

Bryson's algorithm		Bierman's algorithm	
Operation	Multiplications	Operation	Multiplications
$\Lambda_N = 0,\ \lambda_N = 0$		$\Lambda_N = 0,\ \lambda_N = 0$	
$S_i = H_i^T R_i^{-1} H_i^T$	n^2	—	
$\Psi_i = \Phi_i[I - P_i^F S_i]$	$2n^3$	$\Psi_i = \Phi_i(I - K_i H_i)$	$2n^2$
$B_i = S_i[I - P_i^F S_i]$	n^3	$N_i = H_i P_i^P H_i^T + R_i$	n^2
$N_i = H_i P_i^P H_i^T + R_i$	$\frac{1}{2}n^2$	$B_i = H_i^T N_i^{-1} H_i$	$\frac{1}{2}n^2$
$\Lambda_i^s = \Psi_i^T \Lambda_{i+1}^s \Psi_i + B_i$	$1\frac{1}{2}n^3$	$\Lambda_i^s = \Psi_i^T \Lambda_{i+1}^s \Psi_i + B_i$	$1\frac{1}{2}n^3$
$\lambda_i^s = \Psi_i^T \lambda_{i+1}^s + N_i^{-1} H_i^T R_i^{-1} \nu_i$	n^2	$\lambda_i^s = \Psi_i^T \lambda_{i+1} + H_i N_i^{-1} \nu_i$	n^2
Total computations	$3\frac{1}{2}n^3 + 4n^3$	—	$1\frac{1}{2}n^3 + 4\frac{1}{2}n^2$
$\hat{x}_i^s = \hat{x}_i^F - P_i \Phi_i^T \lambda_i^s$ [a]	$2n^2$	$\hat{x}_i^s = x_i^P + P_i^P \lambda_i^s$	n^2
$P_i^s = P_i^F - P_i^F \Phi_i^T \Lambda_i^s \Phi_i P_i^F$ [b]	$2\frac{1}{2}n^3$	$P_i^s = P_i^P - P_i^P \Lambda_i^s P_i^P$	$1\frac{1}{2}n^3$
Total computations	$2\frac{1}{2}n^3 + 2n^2$	—	$1\frac{1}{2}n^3 + n^2$
Combined total	$6n^3 + 5n^2$	—	$3n^3 + 5\frac{1}{2}n^2$

[a] This is Eq. (68).
[b] This follows from Eq. (68).

Table 2 Computations for discrete R-T-S smoother

	Number of multiplications
$P_N^S = P_N^F$	
$\hat{x}_N^s = \hat{x}_N^F$	—
$L_i = P_i^F \Phi_i^T$	(saved from filter)
$C_i = LP_{i+1}^{-1}$	$1\frac{1}{2}n^3$ (exploiting symmetry of P_{i+1} when computing P_{i+1}^{-1})
$\hat{x}_i^s = \hat{x}_i^F + C_i(\hat{x}_{i+1}^s - \hat{x}_{i+1}^P)$	n^2
$P_i^s = P_i^F - C_i(P_{i+1}^P - P_{i+1}^s)C_i^T$	$\frac{1}{2}n^3$ (exploiting symmetry of r-h-s)
	Total computations: $2n^3 + n^2$

$$\bar{P}_k^s = L_k P_{k+1}^s L_k^T + \Gamma_k \Delta_k \Gamma_k^T \tag{117}$$

$$P_k^s = \Phi_k^{-1} \bar{P}_{k+1}^s \Phi_k^{-1} \tag{118}$$

where

$$\Delta_k \triangleq (\Gamma_k^T S_k \Gamma_k + Q_k^{-1})^{-1} \tag{119}$$

$$S_k \triangleq (\Phi_k^{-1})^T P_k^{F^{-1}} \Phi_k^{-1} = (\Phi_k P_k^F \Phi_k^T)^{-1} \tag{120}$$

$$q_k \triangleq (\Phi_k^{-1})^T P_k^{F^{-1}} \hat{x}_k^F \tag{121}$$

$$K_k \triangleq S_k \Gamma_k \Delta_k \tag{122}$$

$$y_k \triangleq \Gamma_k^T q_k \tag{123}$$

$$L_k = I - \Gamma_k K_k \tag{124}$$

Although these formulas appear more complicated than the R-T-S, the computational savings is obtained through exploiting the following:

1) Φ_k^{-1} is often available from Φ_k without much additional computations.

2) Q_k is often a diagonal matrix.

3) $S_k \Gamma_k$ can be computed directly from $\Phi_k P_k^F \Phi_k^T$ (which is computed in the filter) without computing S_k.

4) Δ_k is low order e.g., a scalar.

5) L_k is a rank 1 matrix; these products such as $L_k P_{k+1}^s$ require only $2n^2$ comultiplications.

Computation Comparison of Smoothers

In Table 1, a summary of the two versions of the adjoint smoothers are given: one as derived originally by Bryson and the other by Bierman. As can be seen, for a scalar measurement, Bierman's is more than twice as efficient for large values of n ($1\frac{1}{2}n^3$ vs $4n^3$). As the number of measurements increases, the number of computations increase, and the computational advantage of Bierman's algorithm decreases.

In the case of n measurements, the CPU leverage is $7\frac{1}{2}n^3$ vs $9\frac{1}{2}n^3$.

In Table 2 the computations for the discrete R-T-S smoother are summarized in comparison. As can be seen, if the smoothed covariance is not desired at every point, Bierman's formulation can be more efficient than the R-T-S, but Bryson's is less efficient. If the covariance is desired at every point, then the R-T-S is more efficient than either formulation. This efficiency increases as the number of measurements increases. Thus, the primary utility of the adjoint smoother is due to superior numerical stability. Using the transformed algorithms for rank 1 process noise, the number of multiplications is approximately the same, i.e., $2n^3 + 3\frac{1}{2}n^2$, as the discrete R-T-S smoother, assuming that the inverse of the transition matrix is available without additional computations. However, it should share the numerical advantages of Bierman's adjoint smoother, but it is more efficient.

Conclusions

Some historical fixed-interval smoothers were revisited theoretically, using variation calculus first introduced by Bryson and Frazier. It was shown to provide a theoretical basis for both the discrete and continuous smoothers. The notion of complementary models, introduced by Weinert and Desai and establishing a mutual dual relationship within the smoothing problem, permitted us to take a shortcut to derive the continuous Rauch-Tung-Striebel (R-T-S) smoother from Bryson's continuous smoother adjoint smoother. Viewed from the "complementary model point of view," the Bryson-Frazier and the R-T-S smoothers are the same algorithm!

In the process of deriving the discrete Bryson-Frazier smoothers, we also derived a discrete smoother, originally derived by Bierman, that is twice as fast as the Bryson-Frazier algorithm. The Bierman algorithm is not as efficient as the R-T-S smoother. Transformations of the R-T-S smoother introduced by Bierman and Watanabe were reviewed. These forms are competitive with the R-T-S smoother under special circumstances and share the advantage of the Bryson-Frazier-Bierman algorithm in that they avoid differencing matrices.

The avoidance of numerical stability problems in filtering and smoothing would not be complete without a discussion of factorized methods. But that is the subject of another paper.

References

[1]Bryson, A. E., Jr., and Frazier, M., "Smoothing for Linear and Nonlinear Dynamic Systems," AeroSystem Division, Wright Patterson AFB, OH, TDR 63-119, 1963, pp. 353-364.

[2]Swerling, P., "A Proposed Stagewise Differential Correction for Satellite Tracking and Prediction," *Journal Astronautic Science*, Vol. 6, 1959, pp. 46-52.

[3]Kalman, R. E., "A New Approach to Problems in Stochastic Estimation and Control," *Journal of Basic Engineering*, Vol. 83, ASME Series D, 1960, pp. 35-45.

[4]Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, Vol. 83, ASME Series D, 1961, pp. 95-108.

[5]Bryson, A. E., Jr., and Ho, Y. C., *Applied Optimal Control*, Blaisdell, Waltham, MA, 1969, Chap. 13.

[6]Bryson, A. E., Jr., and Ho, Y. C., *Applied Optimal Control*, Hemisphere, Washington, DC, 1975, pp. 390-407.

[7]Saaty, T. L., and Bram, J., *Non-linear Mathematics*, McGraw-Hill, New York, 1964, p. 358.

[8]Fraser, D. C., "A New Technique for the Optimal Smoothing of Data," Sc.D. Dissertation, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 1967.

[9]McReynolds, S. R., and Bryson, A. E., "A Successive Sweep Method for Solving Optimal Programming Problems," *Joint Automatic Control Conference*, American Automatic Control Council, New York, 1965, pp. 551-555.

[10]McReynolds, S. R., "A Successive Sweep Method for Solving Optimal Programming Problems," Ph.D. Thesis, Harvard University, Dept. of Engineering and Applied Physics, Cambridge, MA, 1966.

[11]Dyer, P., and McReynolds, S. R., *The Computation and Theory of Optimal Control*, Academic, New York, 1970.

[12]Rauch, H. E., Tung, F., and Striebel, C., "Maximum Likelihood Estimates of Linear Dynamic Systems," *AIAA Journal*, Vol. 3, No. 8, 1965, pp. 1445-1450.

[13]Mayne, D. Q., "A Solution to the Smoothing Problem for Linear-Dynamic Systems," *Automatica*, Vol. 4, 1966, pp. 73-92.

[14]Fraser, D. C., and Potter, J. E., "The Optimal Linear Smoother as a Combination of Two Optimum Linear Filters," *IEEE Transactions on Automatic Control*, Vol. AC-14, Aug. 1969, pp. 387-390.

[15]Meditch, J. S., "A Survey of Data Smoothing for Linear and Nonlinear Dynamic Systems," *Automatica*, Vol. 9, Pergammon, London, 1973, pp. 151-162.

[16]Bierman, G. J., "Fixed Interval Smoothing with Discrete Measurement," *International Journal of Control*, Vol. 18, No. 1, 1973, pp. 65-75.

[17]Weinert, H. L., and Desai, V. B., "On Complementary Models and Fixed Interval Smoothing," *IEEE Transactions on Automatic Control*, Vol. AC-26, No. 4, 1981, pp. 863-867.

[18]Bierman, G. J., *Factorization Methods for Discrete Sequential Estimations*, Academic, New York, 1977.

[19]Bierman, G. J., "A New Computationally Efficient Fixed-Interval Discrete Time Smoother," *Automatica*, Vol. 19, No. 5, pp. 503-561.

[20]Watanabe, K., and Tzafestas, S. G., "New Computationally Efficient Formula for Backward-Pass Fixed Interval Smoother and its UD. Factorization Algorithm," *IEE Proceedings*, Vol. 136, No. 2, Pt.D., 1989, pp. 73-78.