6

Continuous-Time Smoothing

6.1 Introduction

The previously-described minimum-mean-square-error and minimum-variance filtering solutions operate on measurements up to the current time. If some processing delay can be tolerated then improved estimation performance can be realised through the use of smoothers. There are three state-space smoothing technique categories, namely, fixed-point, fixed-lag and fixed-interval smoothing. Fixed-point smoothing refers to estimating some linear combination of states at a previous instant in time. In the case of fixed-lag smoothing, a fixed time delay is assumed between the measurement and on-line estimation processes. Fixed-interval smoothing is for retrospective data analysis, where measurements recorded over an interval are used to obtain the improved estimates. Compared to filtering, smoothing has a higher implementation cost, as it has increased memory and calculation requirements.

A large number of smoothing solutions have been reported since Wiener's and Kalman's development of the optimal filtering results – see the early surveys [1] – [2]. The minimum-variance fixed-point and fixed-lag smoother solutions are well known. Two fixed-interval smoother solutions, namely the maximum-likelihood smoother developed by Rauch, Tung and Striebel [3], and the two-filter Fraser-Potter formula [4], have been in widespread use since the 1960s. However, the minimum-variance fixed-interval smoother is not well known. This smoother is simply a time-varying state-space generalisation of the optimal Wiener solution.

The main approaches for continuous-time fixed-point, fixed-lag and fixed-interval smoothing are canvassed here. It is assumed throughout that the underlying noise processes are zero mean and uncorrelated. Nonzero means and correlated processes can be handled using the approaches of Chapters 3 and 4. It is also assumed here that the noise statistics and state-space model parameters are known precisely. Note that techniques for estimating parameters and accommodating uncertainty are addressed subsequently.

Some prerequisite concepts, namely time-varying adjoint systems, backwards differential equations, Riccati equation comparison and the continuous-time maximum-likelihood method are covered in Section 6.2. Section 6.3 outlines a derivation of the fixed-point smoother by Meditch [5]. The fixed-lag smoother reported by Sage *et al* [6] and Moore [7], is the subject of Section 6.4. Section 6.5 deals with the Rauch-Tung-Striebel [3], Fraser-Potter [4] and minimum-variance fixed-interval smoother solutions [8] - [10]. As before, the approach

---

"Life has got a habit of not standing hitched. You got to ride it like you find it. You got to change with it. If a day goes by that don't change some of your old notions for new ones, that is just about like trying to milk a dead cow." *Woodrow Wilson Guthrie* here is to accompany the developments, where appropriate, with proofs about performance being attained. Smoothing is not a panacea for all ills. If the measurement noise is negligible then smoothing (and filtering) may be superfluous. Conversely, if measurement noise obliterates the signals then data recovery may not be possible. Therefore, estimator performance is often discussed in terms of the prevailing signal-to-noise ratio.

6.2 Prerequisites

6.2.1 Time-varying Adjoint Systems

Since fixed-interval smoothers employ backward processes, it is pertinent to introduce the adjoint of a time-varying continuous-time system. Let $\mathcal{G}: \mathbb{R}^p \to \mathbb{R}^q$ denote a linear time-varying system

$$\dot{x}(t) = A(t)x(t) + B(t)w(t)\,, \tag{1}$$

$$y(t) = C(t)x(t) + D(t)w(t)\,, \tag{2}$$

operating on the interval [0, T]. Let w denote the set of $w(t)$ over all time t, that is, $w = \{w(t), t \in [0, T]\}$. Similarly, let $y = \mathcal{G} w$ denote $\{y(t), t \in [0, T]\}$. The adjoint of $\mathcal{G}$, denoted by $\mathcal{G}^H : \mathbb{R}^q \to \mathbb{R}^p$, is the unique linear system satisfying

$$\langle y, \mathcal{G} w\rangle = \langle \mathcal{G}^{H} y, w\rangle \tag{3}$$

for all $y \in \mathbb{R}^q$ and $w \in \mathbb{R}^p$.

Lemma 1: *The adjoint $\mathcal{G}^H$ of the system $\mathcal{G}$ described by (1) – (2), with $x(t_0) = 0$, having the realisation*

$$\dot{\zeta}(t) = -A^T(t)\zeta(t) - C^T(t)u(t)\,, \tag{4}$$

$$z(t) = B^T(t)\zeta(t) + D^T(t)u(t)\,, \tag{5}$$

with $\zeta(T) = 0$, satisfies (3).

The proof follows *mutatis mutandis* from that of Lemma 1 of Chapter 3 and is set out in [11]. The original system (1) – (2) needs to be integrated forwards in time, whereas the adjoint system (4) – (5) needs to be integrated backwards in time. Some important properties of backward systems are discussed in the next section. The simplification $D(t) = 0$ is assumed below unless stated otherwise.

6.2.2 Backwards Differential Equations

The adjoint state evolution (4) is rewritten as

$$-\dot{\zeta}(t) = A^T(t)\zeta(t) + C^T(t)u(t)\,. \tag{6}$$

"The simple faith in progress is not a conviction belonging to strength, but one belonging to acquiescence and hence to weakness." *Norbert Wiener*

The negative sign of the derivative within (6) indicates that this differential equation proceeds backwards in time. The corresponding state transition matrix is defined below.

Lemma 2: *The differential equation (6) has the solution*

$$\xi(t) = \Phi^H(t,t_0)\zeta(t_0) - \int_{t_0}^{t} \Phi^H(s,t)C^T(s)u(s)ds\,, \tag{7}$$

where the adjoint state transition matrix, $\Phi^H(t,t_0)$ *, satisfies*

$$\dot{\Phi}^H(t,t_0) = \frac{d\Phi^H(t,t_0)}{dt} = -A^T(t)\Phi^H(t,t_0)\,, \tag{8}$$

with boundary condition

$$\Phi^H(t,t) = I. \tag{9}$$

Proof: *Following the proof of Lemma 1 of Chapter 3, by differentiating (7) and substituting (4) – (5), it is easily verified that (7) is a solution of (6).* □

The Lyapunov equation corresponding to (6) is described next because it is required in the development of backwards Riccati equations.

Lemma 3: *In respect of the backwards differential equation (6), assume that u(t) is a zero-mean white process with* $E\{u(t)u^T(\tau)\} = U(t)\delta(t-\tau)$ *that is uncorrelated with* $\zeta(t_0)$*, namely,* $E\{u(t)\xi^T(t_0)\} = 0$. *Then the covariances* $P(t,\tau) = E\{\xi(t)\xi^T(t)\}$ *and* $\dot{P}(t,\tau) = \dfrac{dE\{\xi(t)\xi^T(t)\}}{dt}$ *satisfy the Lyapunov differential equation*

$$-\dot{P}(t,\tau) = A(t)^T P(t,\tau) + P(t,\tau)A(t) - C^T(t)U(t)C(t)\,. \tag{10}$$

Proof: *The backwards Lyapunov differential equation (10) can be obtained by using (6) and (7) within* $\dfrac{dE\{\xi(t)\xi^T(t)\}}{dt} = E\{\dot{\xi}(t)\xi^T(t) + \xi(\tau)\dot{\xi}^T(k)\}$ *(see the proof of Lemma 2 in Chapter 3).* □

6.2.3 Comparison of Riccati Equations

The following Riccati Equation comparison theorem is required subsequently to compare the performance of filters and smoothers.

Theorem 1 (Riccati Equation Comparison Theorem) [12], [8]: *Let* $P_1(t) \geq 0$ *and* $P_2(t) \geq 0$ *denote solutions of the Riccati differential equations*

$$\dot{P}_1(t) = A_1(t)P_1(t) + P_1(t)A_1^T(t) - P_1(t)S_1(t)P_1(t) + B_1(t)Q_1(t)B_1^T(t) + B(t)Q(t)B^T(t) \tag{11}$$

and

"Progress always involves risk; you can't steal second base and keep your foot on first base." *Frederick James Wilcox*

$$\dot{P}_2(t) = A_2(t)P_2(t) + P_2(t)A_2^T(t) - P_2(t)S_2(t)P_2(t) + B_2(t)Q_2(t)B_2^T(t) + B(t)Q(t)B^T(t) \tag{12}$$

with $S_1(t) = C_1^T(t)R_1^{-1}(t)C_1(t)$, $S_2(t) = C_2^T(t)R_2^{-1}(t)C_2(t)$, *where* $A_1(t)$, $B_1(t)$, $C_1(t)$, $Q_1(t) \geq 0$, $R_1(t) \geq 0$, $A_2(t)$, $B_2(t)$, $C_2(t)$, $Q_2(t) \geq 0$ and $R_2(t) \geq 0$ *are of appropriate dimensions. If*

(i) $P_1(t_0) \geq P_2(t_0)$ *for a* $t_0 \geq 0$ *and*

(ii) $\begin{bmatrix} Q_1(t) & A_1(t) \\ A_1^T(t) & -S_1(t) \end{bmatrix} \geq \begin{bmatrix} Q_2(t) & A_2(t) \\ A_2^T(t) & -S_2(t) \end{bmatrix}$ *for all* $t \geq t_0$.

Then

$$P_1(t) \geq P_2(t) \tag{13}$$

for all $t \geq t_0$.

Proof: *Condition (i) of the theorem is the initial step of an induction argument. For the induction step, denote* $\dot{P}_3(t) = \dot{P}_1(t) - \dot{P}_2(t)$, $P_3(t) = P_1(t) - P_2(t)$ *and* $\bar{A}(t) = -A_1^T(t) + S_1(t)P_2(t) - 0.5S_1(t)P_3(t)$. *Then*

$$\dot{P}_3(t) = \bar{A}(t)P_3(t) + P_3(t)\bar{A}^T(t) + \begin{bmatrix} I & P_2(t) \end{bmatrix}\left(\begin{bmatrix} Q_1(t) & A_1(t) \\ A_1^t(t) & -S_1(t) \end{bmatrix} - \begin{bmatrix} Q_2(t) & A_2(t) \\ A_2^t(t) & -S_2(t) \end{bmatrix}\right)\begin{bmatrix} I \\ P_2(t) \end{bmatrix}$$

which together with condition (ii) yields

$$\dot{P}_3(t) \geq \bar{A}(t)P_3(t) + P_3(t)\bar{A}^T(t)\,. \tag{14}$$

Lemma 5 of Chapter 3 and (14) imply $\dot{P}_3(t) \geq 0$ *and the claim (13) follows.* □

6.2.4 The Maximum-Likelihood Method

Rauch, Tung and Streibel famously derived their fixed-interval smoother [3] using a maximum-likelihood technique which is outlined as follows. Let $x(t) \sim \mathcal{N}(\mu, R_{xx})$ denote a continuous random variable having a Gaussian (or normal) distribution within mean $E\{x(t)\} = \mu$ and covariance $E\{(x(t) - \mu)(x(t) - \mu)^T\} = R_{xx}$. The continuous-time Gaussian probability density function of $x(t) \in \mathbb{R}^n$ is defined by

$$p(x(t)) = \frac{1}{(2\pi)^{n/2}\left|R_{xx}\right|^{1/2}} \exp\left\{-0.5(x(t)-\mu)^T R_{xx}^{-1}(x(t)-\mu)\right\}, \tag{15}$$

in which $|R_{xx}|$ denotes the determinant of R_{xx}. The probability that the continuous random variable $x(t)$ with a given probability density function $p(x(t))$ lies within an interval [a, b] is given by the likelihood function (which is also known as the cumulative distribution function)

"The price of doing the same old thing is far higher than the price of change." *William Jefferson (Bill) Clinton*

$$P(a \le x(t) \le b) = \int_a^b p(x(t))dx \,. \tag{16}$$

The Gaussian likelihood function for $x(t)$ is calculated from (15) and (16) as

$$f(x(t)) = \frac{1}{(2\pi)^{n/2}\left|R_{xx}\right|^{1/2}} \int_{-\infty}^{\infty} \exp\left\{-0.5(x(t)-\mu)^T R_{xx}^{-1}(x(t)-\mu)\right\} dx \,. \tag{17}$$

It is often more convenient to work with the log-probability density function

$$\log\ p(x(t)) = -\log\ (2\pi)^{n/2}\left|R_{xx}\right|^{1/2} - 0.5(x(t)-\mu)^T R_{xx}^{-1}(x-\mu)dx \tag{18}$$

and the log-likelihood function

$$\log\ f(x(t)) = -\log\ (2\pi)^{n/2}\left|R_{xx}\right|^{1/2} - 0.5\int_{-\infty}^{\infty}(x(t)-\mu)^T R_{xx}^{-1}(x-\mu)dx. \tag{19}$$

Suppose that a given record of $x(t)$ is assumed to be belong to a Gaussian distribution that is a function of an unknown quantity θ. A statistical approach for estimating the unknown θ is the method of maximum likelihood. This typically involves finding an estimate $\hat{\theta}$ that either maximises the log-probability density function

$$\hat{\theta} = \arg \max_{\theta}\ \log\ p(\theta \mid x(t)) \tag{20}$$

or maximises the log-likelihood function

$$\hat{\theta} = \arg \max_{\theta}\ \ \log\ f(\theta \mid x(t)) \,. \tag{21}$$

So-called maximum likelihood estimates can be found by setting either $\frac{\partial \log\ p(\theta \mid x(t))}{\partial \theta}$ or $\frac{\partial \log\ f(\theta \mid x(t))}{\partial \theta}$ to zero and solving for the unknown θ. Continuous-time maximum likelihood estimation is illustrated by the two examples that follow.

Example 1. Consider the first-order autoregressive system

$$\dot{x}(t) = -a_0 x(t) + w(t) \,, \tag{22}$$

where $\dot{x}(t)$ = $\frac{dx(t)}{dt}$, $w(t)$ is a zero-mean Gaussian process and a_0 is unknown. It follows from (22) that $\dot{x}(t)$ ~ $\mathcal{N}(-a_0 x(t),\ \sigma_w^2)$, namely,

$$f(\dot{x}(t)) = \frac{1}{(2\pi)^{n/2}\sigma_w} \int_0^T \exp\left\{-0.5(\dot{x}(t) + a_0 x(t))^2 \sigma_w^{-2}\right\} dt \,. \tag{23}$$

"Faced with the choice between changing one's mind and proving that there is no need to do so, almost everyone gets busy on the proof." *John Kenneth Galbraith*

Taking the logarithm of both sides gives

$$\log\ f(\dot{x}(t)) = -\log\ (2\pi)^{n/2}\sigma_w - 0.5\sigma_w^{-2}\int_0^T (\dot{x}(t) + a_0 x(t))^2\,dt\,. \tag{24}$$

Setting $\dfrac{\partial \log\ f(\dot{x}(t))}{\partial a_0}$ = 0 results in $\int_0^T (\dot{x}(t) + a_0 x(t))x(t)dt$ = 0 and hence

$$\hat{a}_0 = -\left(\int_0^T (x^2(t)dt)\right)^{-1}\int_0^T \dot{x}(t)x(t)dt\,. \tag{25}$$

Example 2. Consider the third-order autoregressive system

$$\dddot{x}(t) + a_2\ddot{x}(t) + a_1\dot{x}(t) + a_0 x(t) = w(t) \tag{26}$$

where $\dddot{x}(t) = \dfrac{d^3x(t)}{dt^3}$ and $\ddot{x}(t) = \dfrac{d^2x(t)}{dt^2}$. The above system can be written in a controllable canonical form as

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\end{bmatrix} = \begin{bmatrix}-a_2 & -a_1 & -a_0\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\\ x_3(t)\end{bmatrix} + \begin{bmatrix}w(t)\\ 0\\ 0\end{bmatrix}. \tag{27}$$

Assuming $\dot{x}_1(t) \sim \mathcal{N}(-a_2x_1(t) - a_1x_2(t) - a_0x_3(t),\ \ \sigma_w^2)$, taking logarithms, setting to zero the partial derivatives with respect to the unknown coefficients, and rearranging yields

$$\begin{bmatrix}\hat{a}_0\\ \hat{a}_1\\ \hat{a}_2\end{bmatrix} = -\begin{bmatrix}\int_0^T x_3^2dt & \int_0^T x_2x_3dt & \int_0^T x_1x_3dt\\ \int_0^T x_2x_3dt & \int_0^T x_2^2dt & \int_0^T x_2x_1dt\\ \int_0^T x_1x_3dt & \int_0^T x_2x_1dt & \int_0^T x_1^2dt\end{bmatrix}^{-1}\begin{bmatrix}\int_0^T \dot{x}_1x_3dt\\ \int_0^T \dot{x}_1x_2dt\\ \int_0^T \dot{x}_1x_1dt\end{bmatrix}, \tag{28}$$

in which state time dependence is omitted for brevity.

6.3 Fixed-Point Smoothing

6.3.1 Problem Definition

In continuous-time fixed-point smoothing, it is desired to calculate state estimates at one particular time of interest, τ, $0 \le \tau \le t$, from measurements $z(t)$ over the interval $t \in [0, T]$. For example, suppose that a continuous measurement stream of a tennis ball's trajectory is available and it is desired to determine whether it bounced within the court boundary. In this case, a fixed-point smoother could be employed to estimate the ball position at the time of the bounce from the past and future measurements.

"When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is probably wrong." *Arthur Charles Clarke*

A solution for the continuous-time fixed-point smoothing problem can be developed from first principles, for example, see [5] - [6]. However, it is recognised in [13] that a simpler solution derivation follows by transforming the smoothing problem into a filtering problem that possesses an augmented state. Following the nomenclature of [14], consider an augmented state vector having two components, namely, $x^{(a)}(t) = \begin{bmatrix} x(t) \\ \xi(t) \end{bmatrix}$. The first component, $x(t) \in \mathbb{R}^n$, is the state of the system $\dot{x}(t) = A(t)x(t) + B(t)w(t)$ and $y(t) = C(t)x(t)$. The second component, $\xi(t) \in \mathbb{R}^n$, equals $x(t)$ at time $t = \tau$, that is, $\xi(t) = x(\tau)$. The corresponding signal model may be written as

$$\dot{x}^{(a)}(t) = A^{(a)}(t)x^{(a)}(t) + B^{(a)}(t)w(t) \tag{29}$$

$$z(t) = C^{(a)}(t)x^{(a)}(t) + v(t), \tag{30}$$

where $A^{(a)} = \begin{bmatrix} A(t) & 0 \\ 0 & \delta_{t\tau}A(t) \end{bmatrix}$, $B^{(a)}(t) = \begin{bmatrix} B(t) \\ \delta_{t\tau}B(t) \end{bmatrix}$ and $C^{(a)}(t) = [C(t) \quad 0]$, in which $\delta_{t\tau} = \begin{cases} 1 & \text{if } t = \tau \\ 0 & \text{if } t \neq \tau \end{cases}$ is the Kronecker delta function. Note that the simplifications $A^{(a)} = \begin{bmatrix} A(t) & 0 \\ 0 & 0 \end{bmatrix}$ and $B^{(a)}(t) = \begin{bmatrix} B(t) \\ 0 \end{bmatrix}$ arise for $t > \tau$. The smoothing objective is to produce an estimate $\hat{\xi}(t)$ of $\xi(t)$ from the measurements $z(t)$ over $t \in [0, T]$.

6.3.2 Solution Derivation

Employing the Kalman-Bucy filter recursions for the system (29) – (30) results in

$$\begin{aligned} \dot{\hat{x}}^{(a)}(t) &= A^{(a)}(t)\hat{x}^{(a)}(t) + K^{(a)}(t)\left(z(t) - C^{(a)}(t)\hat{x}(t\,|\,t)\right) \\ &= \left(A^{(a)}(t) - K^{(a)}(t)C^{(a)}(t)\right)x^{(a)}(t) + K^{(a)}(t)z(t), \end{aligned} \tag{31}$$

where

$$K^{(a)}(t) = P^{(a)}(t)(C^{(a)})^T(t)R^{-1}(t), \tag{32}$$

in which $P^{(a)}(t) \in \mathbb{R}^{2n \times 2n}$ is to be found. Consider the partitioning $K^{(a)}(t) = \begin{bmatrix} K(t) \\ \underline{K}(t) \end{bmatrix}$, then for $t > \tau$, (31) may be written as

$$\begin{bmatrix} \hat{x}(t\,|\,t) \\ \hat{\xi}(t) \end{bmatrix} = \begin{bmatrix} A(t) - K(t)C(t) & 0 \\ -\underline{K}(t)C(t) & 0 \end{bmatrix} \begin{bmatrix} \hat{x}(t\,|\,t) \\ \xi(t) \end{bmatrix} + \begin{bmatrix} K(t) \\ \underline{K}(t) \end{bmatrix} z(t). \tag{33}$$

"Don't be afraid to take a big step if one is indicated. You can't cross a chasm in two small jumps." *David Lloyd George*

Define the augmented error state as $\tilde{x}^{(a)}(t) = x^{(a)}(t) - \hat{x}^{(a)}(t)$, that is,

$$\begin{bmatrix} \tilde{x}(t|t) \\ \tilde{\xi}(t) \end{bmatrix} = \begin{bmatrix} x(t) \\ \xi(\tau) \end{bmatrix} - \begin{bmatrix} \hat{x}(t|t) \\ \hat{\xi}(t) \end{bmatrix}. \tag{34}$$

Differentiating (34) and using $z(t) = C(t)\tilde{x}(t|t) + v(t)$ gives

$$\begin{aligned} \begin{bmatrix} \dot{\tilde{x}}(t|t) \\ \dot{\tilde{\xi}}(t) \end{bmatrix} &= \begin{bmatrix} \dot{x}(t|t) \\ 0 \end{bmatrix} - \begin{bmatrix} \dot{\hat{x}}(t|t) \\ \dot{\hat{\xi}}(t) \end{bmatrix} \\ &= \begin{bmatrix} A(t) - K(t)C(t) & 0 \\ -\underline{K}(t)C(t) & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}(t|t) \\ \tilde{\xi}(t) \end{bmatrix} + \begin{bmatrix} B(t) & -K(t) \\ 0 & -\underline{K}(t) \end{bmatrix} \begin{bmatrix} w(t) \\ v(t) \end{bmatrix}. \end{aligned} \tag{35}$$

Denote $P^{(a)}(t) = \begin{bmatrix} P(t) & \Sigma^T(t) \\ \Sigma(t) & \Omega(t) \end{bmatrix}$, where $P(t) = E\{[x(t) - \hat{x}(t|t)][(x(t) - \hat{x}(t|t)]^T\}$, $\Omega(t) = E\{[\xi(t) - \hat{\xi}(t)][\xi(t) - \hat{\xi}(t)]^T\}$ and $\Sigma(t) = E\{[\xi(t) - \hat{\xi}(t)][x(t) - \hat{x}(t|t)]^T\}$. Applying Lemma 2 of Chapter 3 to (35) yields the Lyapunov differential equation

$$\begin{aligned} \begin{bmatrix} \dot{P}(t) & \dot{\Sigma}^T(t) \\ \dot{\Sigma}(t) & \dot{\Omega}(t) \end{bmatrix} &= \begin{bmatrix} A(t) - K(t)C(t) & 0 \\ -\underline{K}(t)C(t) & 0 \end{bmatrix} \begin{bmatrix} P(t) & \Sigma^T(t) \\ \Sigma(t) & \Omega(t) \end{bmatrix} + \begin{bmatrix} P(t) & \Sigma^T(t) \\ \Sigma(t) & \Omega(t) \end{bmatrix} \begin{bmatrix} A^T(t) - C^T(t)K^T(t) & -C^T(t)\underline{K}^T(t) \\ 0 & 0 \end{bmatrix} \\ &+ \begin{bmatrix} B(t) & -K(t) \\ 0 & -\underline{K}(t) \end{bmatrix} \begin{bmatrix} Q(t) & 0 \\ 0 & R(t) \end{bmatrix} \begin{bmatrix} B^T(t) & 0 \\ -K^T(t) & -\underline{K}^T(t) \end{bmatrix}. \end{aligned}$$

Simplifying the above differential equation yields

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)C^T(t)R^{-1}(t)C(t)P(t) + B(t)Q(t)B^T(t), \tag{36}$$

$$\dot{\Sigma}(t) = \Sigma(t)\left(A^T(t) - C^T(t)K^T(t)\right), \tag{37}$$

$$\dot{\Omega}(t) = -\Sigma(t)C^T(t)R^{-1}(t)C(t)\Sigma^T(t). \tag{38}$$

Equations(37) – (38) can be initialised with

$$\Sigma(\tau) = P(\tau). \tag{39}$$

Thus, the fixed-point smoother estimate is given by

$$\hat{\xi}(t) = \Sigma(t)C^T(t)R^{-1}(t)\left(z(t) - C(t)\hat{x}(t|t)\right), \tag{40}$$

"If you don't like change, you're going to like irrelevance even less." *General Eric Shinseki* which is initialised with $\hat{\xi}(\tau) = \hat{x}(\tau)$. Alternative derivations of (40) are presented in [5], [8], [15]. The smoother (40) and its associated error covariances (36) – (38) are also discussed in [16], [17].

6.3.3 Performance

It can be seen that the right-hand-side of the smoother error covariance (38) is non-positive and therefore $\Omega(t)$ must be monotonically decreasing. That is, the smoothed estimates improve with time. However, since the right-hand-side of (36) varies inversely with $R(t)$, the improvement reduces with decreasing signal-to-noise ratio. It is shown below the fixed-point smoother improves on the performance of the minimum-variance filter.

Lemma 4: *In respect of the fixed-point smoother (40),*

$$P(t) \geq \Omega(t) . \tag{41}$$

Proof: *The initialisation (39) accords with condition (i) of Theorem 1. Condition (ii) of the theorem is satisfied since*

$$\begin{bmatrix} Q(t) & A(t) \\ A^T(t) & -C^T(t)R^{-1}(t)C(t) \end{bmatrix} \geq \begin{bmatrix} -C^T(t)R^{-1}(t)C(t)0 & 0 \\ 0 & 0 \end{bmatrix}$$

and hence the claim (41) follows. □

6.4 Fixed-Lag Smoothing

6.4.1 Problem Definition

For continuous-time estimation problems, as usual, it assumed that the observations are modelled by $\dot{x}(t) = A(t)x(t) + B(t)w(t)$, $z(t) = C(t)x(t) + v(t)$, with $E\{w(t)w^T(\tau)\} = Q(t)\delta(t-\tau)$ and $E\{v(t)v^T(\tau)\} = R(t)\delta(t-\tau)$. In fixed-lag smoothing, it is desired to calculate state estimates at a fixed time lag behind the current measurements. That is, smoothed state estimates, $\hat{x}(t \mid t+\tau)$, are desired at time t, given data at time $t + \tau$, where τ is a prescribed lag. In particular, fixed-lag smoother estimates are sought which minimise $E\{[x(t) - \hat{x}(t \mid t+\tau)][x(t) - \hat{x}(t \mid t+\tau)]^T\}$. It is found in [18] that the smoother yields practically all the improvement over the minimum-variance filter when the smoothing lag equals several time constants associated with the minimum-variance filter for the problem.

6.4.2 Solution Derivation

Previously, augmented signal models together with the application of the standard Kalman filter recursions were used to obtain the smoother results. However, as noted in [19], it is difficult to derive the optimal continuous-time fixed-lag smoother in this way because an ideal delay operator cannot easily be included within an asymptotically stable state-space system. Consequently, an alternate derivation based on that in [6] is outlined in the

"Change is like putting lipstick on a bulldog. The bulldog's appearance hasn't improved, but now it's really angry." *Rosbeth Moss Kanter* following. Recall that the gain of the minimum-variance filter is calculated as $K(t)$ = $P(t)C^T(t)R^{-1}(t)$, where $P(t)$ is the solution of the Riccati equation (3.36). Let $\Phi(\tau,t)$ denote the transition matrix of the filter error system $\dot{\tilde{x}}(t|t)$ = $(A(t) - K(t)C(t))\tilde{x}(t|t)$ + $B(t)w(t)$ − $K(t)v(t)$, that is,

$$\dot{\Phi}(t,s) = \big(A(\tau) - K(\tau)C(\tau)\big)\Phi(t,s) \tag{42}$$

and $\Phi(s,s)$ = I. It is assumed in [6], [17], [18], [20] that a smoothed estimate $\hat{x}(t|t+\tau)$ of $x(t)$ is obtained as

$$\hat{x}(t|t+\tau) = \hat{x}(t) + P(t)\xi(t,\tau)\,, \tag{43}$$

where

$$\xi(t,t+\tau) = \int_t^{t+\tau} \Phi^T(\tau,t)C^T(\tau)R^{-1}(\tau)\big(z(\tau) - C(\tau)\hat{x}(\tau|\tau)\big)d\tau\,. \tag{44}$$

The formula (43) appears in the development of fixed interval smoothers [21] - [22], in which case $\xi(t)$ is often called an adjoint variable. From the use of Leibniz' rule, that is,

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(t,s)ds = f(t,b(t))\frac{db(t)}{dt} - f(t,a(t))\frac{da(t)}{dt} + \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} f(t,s)ds\,,$$

it can be found that

$$\begin{aligned}\dot{\zeta}(t,t+\tau) &= \Phi^T(t+\tau)C^T(t+\tau)R^{-1}(t+\tau)\big(z(t+\tau) - C(t+\tau)\hat{x}(t+\tau)|t+\tau)\big) \\ &\quad - C^T(t)R^{-1}(t)\big(z(t) - C(t)\hat{x}(t|t)\big) - \big(A(t) - K(t)C(t)\big)^T \xi(t,t+\tau)\,.\end{aligned} \tag{45}$$

Differentiating (43) with respect to t gives

$$\dot{\hat{x}}(t|t+\tau) = \dot{\hat{x}}(t|t) + \dot{P}(t)\xi(t,\tau) + P(t)\dot{\xi}(t,\tau)\,. \tag{46}$$

Substituting $\xi(t,\tau)$ = $P^{-1}(t)\big(\hat{x}(t|t+\tau) - \hat{x}(t|t)\big)$ and expressions for $\dot{\hat{x}}(t)$, $\dot{P}(t)$, $\dot{\xi}(t,t+\tau)$ into (43) yields the fixed–lag smoother differential equation

$$\begin{aligned}\dot{\hat{x}}(t|t+\tau) &= A(t)\dot{\hat{x}}(t|t+\tau) + B(t)Q(t)B^T(t)P^{-1}(t)\big(\hat{x}(t|t+\tau) - \hat{x}(t|t)\big) \\ &\quad + P(t)\Phi^T(t+\tau,t)C^T(t+\tau)R^{-1}(t+\tau)\big(z(t+\tau) - C(t+\tau)\hat{x}(t+\tau|t+\tau)\big)\,.\end{aligned} \tag{47}$$

"An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: What does happen is that the opponents gradually die out." *Max Karl Ernst Ludwig Planck*

6.4.3 Performance

Lemma 5 [18]:

$$P(t) - E\{[x(t) - \dot{\hat{x}}(t \mid t+\tau)][x(t) - \dot{\hat{x}}(t \mid t+\tau)]^T\} > 0. \tag{48}$$

Proof. *It is argued from the references of [18] for the fixed-lag smoothed estimate that*

$$E\{[x(t) - \dot{\hat{x}}(t \mid t+\tau)][x(t) - \dot{\hat{x}}(t \mid t+\tau)]^T\} = P(t) - P(t)\int_t^{\tau} \Phi^T(s,t) C^T(s) R^{-1}(s) C(s) \Phi(s,t) ds \; P(t) \tag{49}$$

Thus, (48) follows by inspection of (49). □

That is to say, the minimum-variance filter error covariance is greater than fixed-lag smoother error covariance. It is also argued in [18] that (48) implies the error covariance decreases monotonically with the smoother lag τ.

6.5 Fixed-Interval Smoothing

6.5.1 Problem Definition

Many data analyses occur off-line. In medical diagnosis for example, reviews of ultra-sound or CAT scan images are delayed after the time of measurement. In principle, smoothing could be employed instead of filtering for improving the quality of an image sequence.

Fixed-lag smoothers are elegant – they can provide a small performance improvement over filters at moderate increase in implementation cost. The best performance arises when the lag is sufficiently large, at the expense of increased complexity. Thus, the designer needs to trade off performance, calculation cost and delay.

Fixed-interval smoothers are a brute-force solution for estimation problems. They provide improved performance without having to fine tune a smoothing lag, at the cost of approximately twice the filter calculation complexity. Fixed interval smoothers involve two passes. Typically, a forward process operates on the measurements. Then a backward system operates on the results of the forward process.

The plants are again assumed to have state-space realisations of the form $\dot{x}(t) = A(t)x(t) + B(t)w(t)$ and $y(t) = C(t)x(t) + D(t)w(t)$. Smoothers are considered which operate on measurements $z(t) = y(t) + v(t)$ over a fixed interval $t \in [0, T]$. The performance criteria depend on the quantity being estimated, *viz.*,

- in input estimation, the objective is to calculate a $\hat{w}(t \mid T)$ that minimises $E\{[w(t) - \hat{w}(t \mid T)][w(t) - \hat{w}(t \mid T)]^T\}$;
- in state estimation, $\hat{x}(t \mid T)$ is calculated which achieves the minimum $E\{[x(t) - \hat{x}(t \mid T)][x(t) - \hat{x}(t \mid T)]^T\}$; and
- in output estimation, $\hat{y}(t \mid T)$ is produced such that $E\{[y(t) - \hat{y}(t \mid T)][y(t) - \hat{y}(t \mid T)]^T\}$ is minimised.

"If you want to truly understand something, try to change it." *Kurt Lewin*

This section focuses on three continuous-time fixed-interval smoother formulations; the maximum-likelihood smoother derived by Rauch, Tung and Streibel [3], the Fraser-Potter smoother [4] and a generalisation of Wiener's optimal unrealisable solution [8] – [10]. Some additional historical background to [3] – [4] is described within [1], [2], [17].

6.5.2 The Maximum Likelihood Smoother

6.5.2.1 Solution Derivation

Rauch, Tung and Streibel [3] employed the maximum-likelihood method to develop a discrete-time smoother for state estimation and then used a limiting argument to obtain a continuous-time version. A brief outline of this derivation is set out here. Suppose that a record of filtered estimates, $\hat{x}(\tau | \tau)$, is available over a fixed interval $\tau \in [0, T]$. Let $\hat{x}(\tau | T)$ denote smoothed state estimates at time $0 \le \tau \le T$ to be evolved backwards in time from filtered states $\hat{x}(\tau | \tau)$. The smoother development is based on two assumptions. First, it is assumed that $-\dot{\hat{x}}(\tau | T)$ is normally distributed with mean $A(\tau)\hat{x}(\tau | T)$ and covariance $B(\tau)Q(\tau)B^T(\tau)$, that is, $-\dot{\hat{x}}(\tau | T) \sim \mathcal{N}(A(\tau)\hat{x}(\tau | T),\ B(\tau)Q(\tau)B^T(\tau))$. The probability density function of $-\dot{\hat{x}}(\tau | T)$ is

$$p(-\dot{\hat{x}}(\tau | T) | \hat{x}(\tau | T)) = \frac{1}{(2\pi)^{n/2} \left| B(\tau)Q(\tau)B^T(\tau) \right|^{1/2}}$$

$$\times \exp\left\{ -0.5(-\dot{\hat{x}}(\tau | T) - A(\tau)\hat{x}(\tau | T))^T (B(\tau)Q(\tau)B^T(\tau))^{-1} (-\dot{\hat{x}}(\tau | T) - A(\tau)\hat{x}(\tau | T)) \right\}$$

Second, it is assumed that $\hat{x}(\tau | T)$ is normally distributed with mean $\hat{x}(\tau | \tau)$ and covariance $P(\tau)$, namely, $\hat{x}(\tau | T) \sim N(\hat{x}(\tau | \tau), P(\tau))$. The corresponding probability density function is

$$p(\hat{x}(\tau | T) | \hat{x}(\tau | \tau)) = \frac{1}{(2\pi)^{n/2} \left| P(\tau) \right|^{1/2}} \times \exp\left\{ -0.5(\hat{x}(\tau | T) - \hat{x}(\tau | \tau))^T P^{-1}(t)(\hat{x}(\tau | T) - \hat{x}(\tau | \tau)) \right\}.$$

From the approach of [3] and the further details in [6],

$$0 = \frac{\partial \log\ p(-\dot{\hat{x}}(\tau | T) | \hat{x}(\tau | T)) p(\hat{x}(\tau | T) | \hat{x}(\tau | \tau))}{\partial \hat{x}(\tau | T)}$$

$$= \frac{\partial \log\ p(-\dot{\hat{x}}(\tau | T) | \hat{x}(\tau | T))}{\partial \hat{x}(\tau | T)} + \frac{\partial \log\ p(\hat{x}(t | T) | \hat{x}(t | t))}{\partial \hat{x}(t | T)}$$

results in

$$0 = \frac{\partial(-\dot{\hat{x}}(\tau | T) - A(\tau)\hat{x}(\tau | T))^T}{\partial \hat{x}(\tau | T)} (B(\tau)Q(\tau)B^T(\tau))^{-1} (-\dot{\hat{x}}(\tau | T) - A(\tau)\hat{x}(\tau | T)) + P^{-1}(\tau)(\hat{x}(t | T) - \hat{x}(\tau | \tau)).$$

"The soft-minded man always fears change. He feels security in the status quo, and he has an almost morbid fear of the new. For him, the greatest pain is the pain of a new idea." *Martin Luther King Jr.*

Hence, the solution is given by

$$-\dot{\hat{x}}(\tau \mid T) = A(\tau)\hat{x}(\tau \mid T) + G(\tau)\left(\hat{x}(\tau \mid T) - \hat{x}(\tau \mid \tau)\right), \tag{50}$$

where

$$G(\tau) = -B(\tau)Q(\tau)B^T(\tau)\frac{\partial(-\dot{\hat{x}}(\tau \mid T) - A(\tau)\hat{x}(\tau \mid T))^T}{\partial\hat{x}(\tau \mid T)}P^{-1}(\tau) \tag{51}$$

is the smoother gain. Suppose that $\hat{x}(\tau \mid T)$, $A(\tau)$, $B(\tau)$, $Q(\tau)$, $P^{-1}(\tau)$ are sampled at integer k multiples of T_s and are constant during the sampling interval. Using the Euler approximation $-\dot{\hat{x}}(kT_s \mid T)$ = $\dfrac{\hat{x}((k-1)T_s \mid T) - \hat{x}(kT_s \mid T)}{T_s}$, the sampled gain may be written as

$$G(kT_s) = B(kT_s)T_s^{-1}Q(kT_s)B^T(kT_s)(I + AT_s)P^{-1}(kT_s)\,. \tag{52}$$

Recognising that $T_s^{-1}Q(kT_s)$ = $Q(\tau)$, see [23], and taking the limit as $T_s \rightarrow 0$ and yields

$$G(\tau) = B(\tau)Q(\tau)B^T(\tau)P^{-1}(\tau)\,. \tag{53}$$

To summarise, the above fixed-interval smoother is realised by the following two-pass procedure.

(i) In the first pass, the (forward) Kalman-Bucy filter operates on measurements $z(\tau)$ to obtain state estimates $\hat{x}(\tau \mid \tau)$.

(ii) In the second pass, the differential equation (50) operates on the filtered state estimates $\hat{x}(\tau \mid \tau)$ to obtain smoothed state estimates $\hat{x}(\tau \mid T)$. Equation (50) is integrated backwards in time from the initial condition $\hat{x}(\tau \mid T)$ = $\hat{x}(\tau \mid \tau)$ at $\tau = T$.

Alternative derivations of this smoother appear in [6], [20], [23], [24].

6.5.2.2 Alternative Form

For the purpose of developing an alternate form of the above smoother found in the literature, consider a fictitious forward version of (50), namely,

$$\begin{aligned}\dot{\hat{x}}(t \mid T) &= A(t)\hat{x}(t \mid T) + B(t)Q(t)B^T(t)P^{-1}(t)\left(\hat{x}(t \mid T) - \hat{x}(t \mid t)\right) \\ &= A(t)\hat{x}(t \mid T) + B(t)Q(t)B^T(t)\xi(t \mid T)\,,\end{aligned} \tag{54}$$

where

$$\xi(t \mid T) = P^{-1}(t)(\hat{x}(t \mid T) - \hat{x}(t \mid t)) \tag{55}$$

"There is a certain relief in change, even though it be from bad to worse. As I have often found in travelling in a stagecoach that it is often a comfort to shift one's position, and be bruised in a new place." *Washington Irving* is an auxiliary variable. An expression for the evolution of $\xi(t \mid T)$ is now developed. Writing (55) as

$$\hat{x}(\tau \mid T) = \hat{x}(t \mid t) + P(t)\xi(t \mid T) \tag{56}$$

and taking the time differential results in

$$\dot{\hat{x}}(t \mid T) = \dot{\hat{x}}(t \mid t) + \dot{P}(t)\xi(t \mid T) + P(t)\dot{\xi}(t \mid T) . \tag{57}$$

Substituting $\dot{\hat{x}}(t \mid t) = A(t)\hat{x}(t \mid t) + P(t)C^T(t)R^{-1}(z(t) - C(t)\hat{x}(t \mid t))$ into (57) yields

$$P(t)\dot{\xi}(t \mid T) = P(t)C^T(t)R^{-1}C(t) - P(t)C^T(t)R^{-1}z(t) + A(t)P(t)\xi(t) + B(t)Q(t)B^T(t) - \dot{P}(t)\xi(t). \tag{58}$$

Using $\hat{x}(t \mid t) = \hat{x}(t \mid T) - P(t)\xi(t \mid T)$, $-P(t)A^T(t) = A(t)P(t) - P(t)C^T(t)R^{-1}(t)C(t)P(t)\xi(t \mid T)$ + $B(t)Q(t)B^T(t)$ – $\dot{P}(t)$ within (58) and rearranging gives

$$-\dot{\xi}(t \mid T) = -C^T(t)R^{-1}(t)C(t)\hat{x}(t \mid T) + A^T(t)\xi(t) - C^T(t)R^{-1}(t)z(t) . \tag{59}$$

The filter (54) and smoother (57) may be collected together as

$$\begin{bmatrix} \dot{\hat{x}}(t \mid T) \\ -\dot{\xi}(t \mid T) \end{bmatrix} = \begin{bmatrix} A(t) & B(t)Q(t)B^T(t) \\ -C^T(t)R^{-1}(t)C(t) & A^T(t) \end{bmatrix} \begin{bmatrix} \hat{x}(t \mid T) \\ \xi(t \mid T) \end{bmatrix} + \begin{bmatrix} 0 \\ C^T(t)R^{-1}(t)z(t) \end{bmatrix} . \tag{60}$$

Equation (60) is known as the Hamiltonian form of the Rauch-Tung-Striebel smoother [17].

6.5.2.3 Performance

In order to develop an expression for the smoothed error state, consider the backwards signal model

$$-\dot{x}(\tau) = A(\tau)x(\tau) + B(\tau)w(\tau) . \tag{61}$$

Subtracting (50) from (61) results in

$$-\dot{x}(\tau) + \dot{\hat{x}}(\tau \mid T) = (A(\tau) + G(\tau))(x(\tau) - \hat{x}(\tau \mid T)) - G(\tau)(x(\tau) - \hat{x}(\tau \mid \tau)) + B(\tau)w(\tau) . \tag{62}$$

Let $\tilde{x}(\tau \mid T) = x(\tau) - \hat{x}(\tau \mid T)$ denote the smoothed error state and $\tilde{x}(\tau \mid \tau) = x(\tau) - \hat{x}(\tau \mid \tau)$ denote the filtered error state. Then the differential equation (62) can simply be written as

$$-\dot{\tilde{x}}(\tau \mid T) = (A(\tau) + G(\tau))(\tilde{x}(\tau \mid T) - G(\tau)\tilde{x}(\tau \mid \tau) + B(\tau)w(\tau) , \tag{63}$$

where $-\dot{\tilde{x}}(\tau \mid T) = -(\dot{\hat{x}}(\tau \mid T) - \dot{x}(\tau))$. Applying Lemma 3 to (63) and using $E\{\hat{x}(\tau \mid \tau)$, $w^T(\tau)\}$ = 0 gives

$$-\dot{\Sigma}(\tau \mid T) = (A(\tau) + G(\tau))\Sigma(\tau \mid T) + \Sigma(\tau \mid T)(A(\tau) + G(\tau))^T - B(\tau)Q(\tau)B^T(\tau) , \tag{64}$$

"That which comes into the world to disturb nothing deserves neither respect nor patience." *Rene Char* where $\Sigma(\tau | T) = E\{\hat{x}(\tau | T),\ \hat{x}^T(\tau | T)\}$ is the smoother error covariance and $\dot{\Sigma}(\tau | T) = \frac{d\Sigma(\tau | T)}{d\tau}$. The smoother error covariance differential equation (64) is solved backwards in time from the initial condition

$$\Sigma(\tau | T) = P(t | t) \tag{65}$$

at $t = T$, where $P(t | t)$ is the solution of the Riccati differential equation

$$\begin{aligned} \dot{P}(t) &= (A(t) - K(t)C(t))P(t) + P(t)(A^T(t) - C^T(t)K^T(t)) + K(t)R(t)K^T(t) + B(t)Q(t)B^T(t) \\ &= A(t)P(t) + P(t)A^T(t) - K(t)R(t)K^T(t) + B(t)Q(t)B^T(t) \end{aligned} \tag{66}$$

It is shown below that this smoother outperforms the minimum-variance filter. For the purpose of comparing the solutions of forward Riccati equations, consider a fictitious forward version of (64), namely,

$$\dot{\Sigma}(t | T) = (A(t) + G(t))\Sigma(t | T) + \Sigma(t | T)(A(t) + G(t))^T - B(t)Q(t)B^T(t) \tag{67}$$

initialised with

$$\Sigma(t_0 | T) = P(t_0 | t_0) > 0\ . \tag{68}$$

Lemma 6: *In respect of the fixed-interval smoother (50),*

$$P(t | t) \geq \Sigma(t | T)\ . \tag{69}$$

Proof: *The initialisation (68) satisfies condition (i) of Theorem 1. Condition (ii) of the theorem is met since*

$$\begin{bmatrix} B(t)Q(t)B^T(t) + K(t)R(t)K^T(t) & A(t) - K(t)C(t) \\ A^T(t) - C^T(t)K^T(t) & 0 \end{bmatrix} \geq \begin{bmatrix} -B(t)Q(t)B^T(t) & A(t) + G(t) \\ A^T(t) + G^T(t) & 0 \end{bmatrix}$$

for all $t \geq t_0$ and hence the claim (69) follows. □

6.5.3 The Fraser-Potter Smoother

The Central Limit Theorem states that the mean of a sufficiently large sample of independent identically distributed random variables will be approximately normally distributed [25]. The same is true of partial sums of random variables. The Central Limit Theorem is illustrated by the first part of the following lemma. A useful generalisation appears in the second part of the lemma.

Lemma 7: *Suppose that y_1, y_2, …, y_n are independent random variables and W_1, W_2, … W_n are independent positive definite weighting matrices. Let $\mu = E\{y\}$, $u = y_1 + y_2 + … + y_n$ and*

$$v = (W_1 y_1 + W_2 y_2 + … + W_n y_n)\ (W_1 + W_2 + … W_n)^{-1}. \tag{70}$$

"Today every invention is received with a cry of triumph which soon turns into a cry of fear." *Bertolt Brecht*

(i) *If* $y_i \sim \mathcal{N}(\mu, R)$, *i = 1 to n, then* $u \sim \mathcal{N}(n\mu, nR)$;

(ii) *If* $y_i \sim \mathcal{N}(0, I)$, *i = 1 to n, then* $v \sim \mathcal{N}(0, I)$.

Proof:

(i) $E\{u\} = E\{y_1\} + E(y_2) + \ldots + E\{y_n\} = n\mu$. $E\{(u-\mu)(u-\mu)^T\} = E\{(y_1-\mu)(y_1-\mu)^T\} + E\{(y_2-\mu)(y_2-\mu)^T\} + \ldots + E\{(y_n-\mu)(y_n-\mu)^T\} = nR$.

(ii) $E\{v\} = W_1(W_1 + W_2 + \ldots + W_n)^{-1}E\{y_1\} + W_2(W_1 + W_2 + \ldots + W_n)^{-1}E(y_2) + \ldots + W_n(W_1 + W_2 + \ldots W_n)^{-1}E\{y_n\}) = 0$. $E\{vv^T\} = E\{(W_1^T + W_2^T + \ldots + W_n^T)^{-1}W_1^T y_1^T y_1 W_1(W_1 + W_2 + \ldots W_N)^{-1}\} + E\{(W_1^T + W_2^T + \ldots + W_n^T)^{-1}W_2^T y_2^T y_2 W_2(W_1 + W_2 + \ldots W_N)^{-1}\} + \ldots + E\{(W_1^T + W_2^T + \ldots + W_n^T)^{-1}W_n^T y_n^T y_n W_n(W_1 + W_2 + \ldots W_N)^{-1}\} = (W_1^T + W_2^T + \ldots + W_n^T)^{-1}(W_1^T W_1 + W_2^T W_2 + \ldots W_n^T W_n)(W_1 + W_2 + \ldots W_n)^{-1} = I$. □

Fraser and Potter reported a smoother in 1969 [4] that combined state estimates from forward and backward filters using a formula similar to (70) truncated at $n = 2$. The inverses of the forward and backward error covariances, which are indicative of the quality of the respective estimates, were used as weighting matrices. The combined filter and Fraser-Potter smoother equations are

$$\dot{\hat{x}}(t\,|\,t) = A(t)\hat{x}(t\,|\,t) + P(t\,|\,t)C^T(t)R^{-1}(t)(z(t) - C(t)\hat{x}(t\,|\,t)), \tag{71}$$

$$-\dot{\xi}(t\,|\,t) = A(t)\xi(t\,|\,t) + \Sigma(t\,|\,t)C^T(t)R^{-1}(t)(z(t) - C(t)\xi(t\,|\,t)), \tag{72}$$

$$\hat{x}(t\,|\,T) = (P^{-1}(t\,|\,t) + \Sigma^{-1}(t\,|\,t))^{-1}(P^{-1}(t\,|\,t)\hat{x}(t\,|\,t) + \Sigma^{-1}(t\,|\,t)\xi(t\,|\,t)), \tag{73}$$

where $P(t\,|\,t)$ is the solution of the forward Riccati equation $\dot{P}(t\,|\,t) = A(t)P(t\,|\,t) + P(t\,|\,t)A^T(t) - P(t\,|\,t)C^T(t)R^{-1}(t)C(t)P(t\,|\,t) + B(t)Q(t)B^T(t)$ and $\Sigma(t\,|\,t)$ is the solution of the backward Riccati equation $-\dot{\Sigma}(t\,|\,t) = A(t)\Sigma(t\,|\,t) + \Sigma(t\,|\,t)A^T(t) - \Sigma(t\,|\,t)C^T(t)R^{-1}(t)C(t)\Sigma(t\,|\,t) + B(t)Q(t)B^T(t)$.

It can be seen from (72) that the backward state estimates, $\zeta(t)$, are obtained by simply running a Kalman filter over the time-reversed measurements. Fraser and Potter's approach is pragmatic: when the data is noisy, a linear combination of two filtered estimates is likely to be better than one filter alone. However, this two-filter approach to smoothing is *ad hoc* and is not a minimum-mean-square-error design.

"If there is dissatisfaction with the status quo, good. If there is ferment, so much the better. If there is restlessness, I am pleased. Then let there be ideas, and hard thought, and hard work." *Hubert Horatio Humphrey.*

6.5.4 The Minimum-Variance Smoother

6.5.4.1 Problem Definition

The previously described smoothers are focussed on state estimation. A different signal estimation problem shown in Fig. 1 is considered here. Suppose that observations $z = y + v$ are available, where $y_2 = \mathcal{G}_2 w$ is the output of a linear time-varying system and v is measurement noise. A solution $\mathcal{H}$ is desired which produces estimates $\hat{y}_1$ of a second reference system $y_1 = \mathcal{G}_1 w$ in such a way to meet a performance objective. Let $e = y_1 - \hat{y}_1$ denote the output estimation error. The optimum minimum-variance filter can be obtained by finding the solution that minimises $\left\| ee^T \right\|_2$. Here, in the case of smoothing, the performance objective is to minimise $\left\| ee^H \right\|_2$.

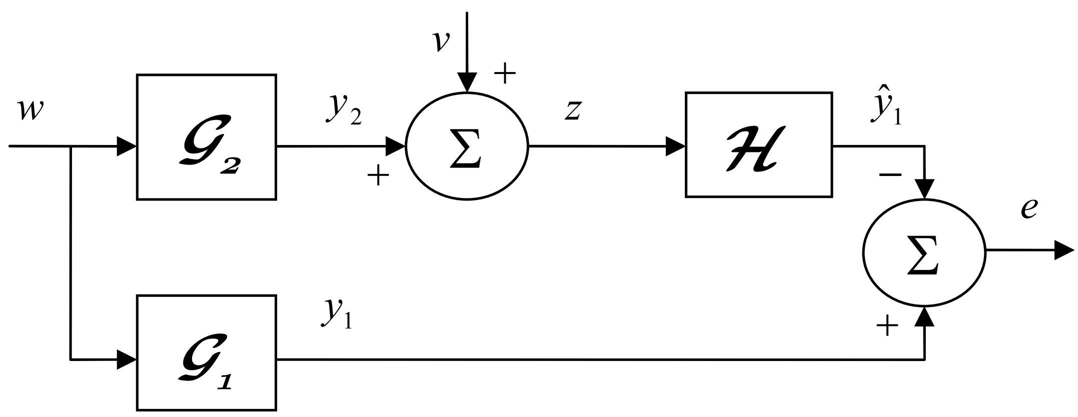


Figure 1. The general estimation problem. The objective is to produce estimates $\hat{y}_1$ of y_1 from measurements z.

6.5.4.2 Optimal Unrealisable Solutions

The minimum-variance smoother is a more recent innovation [8] - [10] and arises by generalising Wiener's optimal noncausal solution for the above time-varying problem. The solution is obtained using the same completing-the-squares technique that was previously employed in the frequency domain (see Chapters 1 and 2). It can be seen from Fig. 1 that the output estimation error is generated by $e = \mathcal{R}_{ei} i$, where

$$\mathcal{R}_{ei} = -\begin{bmatrix} \mathcal{H} & \mathcal{H}\mathcal{G}_2 - \mathcal{G}_1 \end{bmatrix} \tag{74}$$

is a linear system that operates on the inputs $i = \begin{bmatrix} v \\ w \end{bmatrix}$.

Consider the factorisation

$$\Delta\Delta^H = \mathcal{G}_2 Q \mathcal{G}_2^H + R , \tag{75}$$

"Restlessness and discontent are the first necessities of progress. Show me a thoroughly satisfied man and I will show you a failure." *Thomas Alva Edison* in which the time-dependence of $Q(t)$ and $R(t)$ is omitted for notational brevity. Suppose that $\Delta: \mathbb{R}^p \to \mathbb{R}^p$ is causal, namely Δ and its inverse, Δ^{-1}, are bounded systems that proceed forward in time. The system Δ is known as a Wiener-Hopf factor.

Lemma 8: *Assume that the Wiener-Hopf factor inverse, Δ^{-1}, exists over $t \in [0, T]$. Then the smoother solution*

$$\begin{aligned} \mathcal{H} &= \mathcal{G}_1 Q \mathcal{G}_2^H \Delta^{-H} \Delta^{-1} \\ &= \mathcal{G}_1 Q \mathcal{G}_2^H (\Delta\Delta^H)^{-1} \\ &= \mathcal{G}_1 Q \mathcal{G}_2^H (\mathcal{G}_2 Q \mathcal{G}_2^H + R)^{-1} . \end{aligned} \tag{76}$$

minimises $\left\| ee^H \right\|_2 = \left\| \mathcal{R}_{ei} \mathcal{R}_{ei}^H \right\|_2$.

Proof: *It follows from (74) that* $\mathcal{R}_{ei}\mathcal{R}_{ei}^H = \mathcal{G}_1 Q \mathcal{G}_2^H - \mathcal{G}_1 Q \mathcal{G}_2^H \mathcal{H}^H - \mathcal{H}\mathcal{G}_2 Q \mathcal{G}_1^H + \mathcal{H}\Delta\Delta^H\mathcal{H}^H$. *Completing the square leads to* $\mathcal{R}_{ei}\mathcal{R}_{ei}^H = \mathcal{R}_{ei1}\mathcal{R}_{ei1}^H + \mathcal{R}_{ei2}\mathcal{R}_{ei2}^H$, *where*

$$\mathcal{R}_{ei2}\mathcal{R}_{ei2}^H = (\mathcal{H}\Delta - \mathcal{G}_1 Q \mathcal{G}_2^H \Delta^{-H})(\mathcal{H}\Delta - \mathcal{G}_1 Q \mathcal{G}_2^H \Delta^{-H})^H \tag{77}$$

and

$$\mathcal{R}_{ei1}\mathcal{R}_{ei1}^H = \mathcal{G}_1 Q \mathcal{G}_1^H - \mathcal{G}_1 Q \mathcal{G}_2^H (\Delta\Delta^H)^{-1} \mathcal{G}_2 Q \mathcal{G}_1^H . \tag{78}$$

By inspection of (77), the solution (76) achieves

$$\left\| \mathcal{R}_{ei2}\mathcal{R}_{ei2}^H \right\|_2 = 0. \tag{79}$$

Since $\left\| \mathcal{R}_{ei1}\mathcal{R}_{ei1}^H \right\|_2$ *excludes the estimator solution* $\mathcal{H}$, *this quantity defines the lower bound for* $\left\| \mathcal{R}_{ei}\mathcal{R}_{ei}^H \right\|_2$. □

Example 3. Consider the output estimation case where $\mathcal{G}_1 = \mathcal{G}_2$ and

$$\mathcal{H}_{OE} = \mathcal{G}_2 Q \mathcal{G}_2^H (\mathcal{G}_2 Q \mathcal{G}_2^H + R)^{-1} , \tag{80}$$

which is of order n^4 complexity. Using $\mathcal{G}_2 Q \mathcal{G}_2^H = \Delta\Delta^H - R$ leads to the n^2-order solution

$$\mathcal{H}_{OE} = I - R(\Delta\Delta^H)^{-1} . \tag{81}$$

"Whatever has been done before will be done again. There is nothing new under the sun." *Ecclesiastes 1:9*

It is interesting to note from (81) and

$$\mathcal{R}_{ei1}\mathcal{R}_{ei1}^{H} = \mathcal{G}_2 Q \mathcal{G}_2^{H} - \mathcal{G}_2 Q \mathcal{G}_2^{H} (\mathcal{G}_2 Q \mathcal{G}_2^{H} - R)^{-1} \mathcal{G}_2 Q \mathcal{G}_2^{H} \tag{82}$$

that $\lim_{R \to 0} \mathcal{H} = I$ and $\lim_{R \to 0} \mathcal{R}_{ei}\mathcal{R}_{ei}^{H} = 0$. That is, output estimation is superfluous when measurement noise is absent. Let $\{\mathcal{R}_{ei}\mathcal{R}_{ei}^{H}\}_+$ = $\{\mathcal{R}_{ei1}\mathcal{R}_{ei1}^{H}\}_+$ + $\{\mathcal{R}_{ei2}\mathcal{R}_{ei2}^{H}\}_+$ denote the causal part of $\mathcal{R}_{ei}\mathcal{R}_{ei}^{H}$. It is shown below that minimum-variance filter solution can be found using the above completing-the squares technique and taking causal parts.

Lemma 9: *The filter solution*

$$\begin{aligned} \{\mathcal{H}\}_+ &= \{\mathcal{G}_1 Q \mathcal{G}_2^{H} \Delta^{-H} \Delta^{-1}\}_+ \\ &= \{\mathcal{G}_1 Q \mathcal{G}_2^{H} \Delta^{-H}\}_+ \Delta^{-1} \end{aligned} \tag{83}$$

minimises $\left\| \{ee^{H}\}_+ \right\|_2$ = $\left\| \{\mathcal{R}_{ei}\mathcal{R}_{ei}^{H}\}_+ \right\|_2$*, provided that the inverses exist.*

Proof: *It follows from (77) that*

$$\{\mathcal{R}_{ei2}\mathcal{R}_{ei2}^{H}\}_+ = \{(\mathcal{H}\Delta - \mathcal{G}_1 Q \mathcal{G}_2^{H} \Delta^{-H})(\mathcal{H}\Delta - \mathcal{G}_1 Q \mathcal{G}_2^{H} \Delta^{-H})^{H}\}_+ . \tag{84}$$

By inspection of (84), the solution (83) achieves

$$\left\| \{\mathcal{R}_{ei2}\mathcal{R}_{ei2}^{H}\}_+ \right\|_2 = 0. \tag{85}$$ □

It is worth pausing at this juncture to comment on the significance of the above results.

- The formulation (76) is an optimal solution for the time-varying smoother problem since it can be seen from (79) that it achieves the best-possible performance.
- Similarly, (83) is termed an optimal solution because it achieves the best-possible filter performance (85).
- By inspection of (79) and (85) it follows that the minimum-variance smoother outperforms the minimum-variance filter.
- In general, these optimal solutions are not very practical because of the difficulty in realising an exact Wiener-Hopf factor.

Practical smoother (and filter) solutions that make use of an approximate Wiener-Hopf factor are described below.

6.5.4.3 Optimal Realisable Solutions

Output Estimation

The Wiener-Hopf factor is modelled on the structure of the spectral factor which is described Section 3.4.4. Suppose that $R(t) > 0$ for all $t \in [0, T]$ and there exist $R^{1/2}(t) > 0$ such

---

"There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things." *Niccolo Di Bernado dei Machiavelli* that $R(t) = R^{1/2}(t)\, R^{1/2}(t)$. An approximate Wiener-Hopf factor $\hat{\Delta} : \mathbb{R}^p \to \mathbb{R}^p$ is defined by the system

$$\begin{bmatrix} \dot{x}(t) \\ \delta(t) \end{bmatrix} = \begin{bmatrix} A(t) & K(t)R^{1/2}(t) \\ C(t) & R^{1/2}(t) \end{bmatrix} \begin{bmatrix} x(t) \\ z(t) \end{bmatrix}, \tag{86}$$

where $K(t) = P(t)C^T(t)R^{-1}(t)$ is the Kalman gain in which $P(t)$ is the solution of the Riccati differential equation

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)C^T(t)R^{-1}(t)C(t)P(t) + B(t)Q(t)B^T(t)\,. \tag{87}$$

The output estimation smoother (81) can be approximated as

$$\begin{aligned} \mathcal{H}_{OE} &= I - R(\hat{\Delta}\hat{\Delta}^H)^{-1} \\ &= I - R\hat{\Delta}^{-H}\hat{\Delta}^{-1}. \end{aligned} \tag{88}$$

An approximate Wiener-Hopf factor inverse, $\hat{\Delta}^{-1}$, within (88) is obtained from (86) and the Matrix Inversion Lemma, namely,

$$\begin{bmatrix} \dot{\hat{x}}(t) \\ \alpha(t) \end{bmatrix} = \begin{bmatrix} A(t) - K(t)C(t) & K(t) \\ -R^{-1/2}(t)C(t) & R^{-1/2}(t) \end{bmatrix} \begin{bmatrix} x(t) \\ z(t) \end{bmatrix}, \tag{89}$$

where $\hat{x}(t) \in \mathbb{R}^n$ is an estimate of the state within $\hat{\Delta}^{-1}$. From Lemma 1, the adjoint of $\hat{\Delta}^{-1}$, which is denoted by $\hat{\Delta}^{-H}$, has the realisation

$$\begin{bmatrix} -\dot{\xi}(t) \\ \beta(t) \end{bmatrix} = \begin{bmatrix} A^T(t) - C^T(t)K^T(t) & -C^T(t)R^{-1/2}(t) \\ K^T(t) & R^{-1/2}(t) \end{bmatrix} \begin{bmatrix} \xi(t) \\ \alpha(t) \end{bmatrix}. \tag{90}$$

where $\xi(t) \in \mathbb{R}^p$ is an estimate of the state within $\hat{\Delta}^{-H}$. Thus, the smoother (88) is realised by (89), (90) and

$$\hat{y}(t \mid T) = z(t) - R(t)\beta(t)\,. \tag{91}$$

Procedure 1. The above output estimator can be implemented via the following three steps.

Step 1. Operate $\hat{\Delta}^{-1}$ on the measurements $z(t)$ using (89) to obtain $\alpha(t)$.

Step 2. In lieu of the adjoint system (90), operate (89) on the time-reversed transpose of $\alpha(t)$. Then take the time-reversed transpose of the result to obtain $\beta(t)$.

Step 3. Calculate the smoothed output estimate from (91).

"If I have a thousand ideas and only one turns out to be good, I am satisfied." *Alfred Bernhard Nobel*

Example 4. Consider an estimation problem parameterised by a = – 1, b = $\sqrt{2}$, c = 1, d = 0, σ_w^2 = σ_v^2 = 1, which leads to p = k = $\sqrt{3}$ – 1 [26]. Smoothed output estimates may be obtained by evolving

$$\dot{\hat{x}}(t) = \sqrt{3}\hat{x}(t) + \sqrt{3}z(t) \,,\ \alpha(t) = -\hat{x}(t) + z(t) \,,$$

time-reversing the $\alpha(t)$ and evolving

$$\dot{\xi}(t) = \sqrt{3}\xi(t) + \sqrt{3}\alpha(t) \,,\ \beta(t) = -\xi(t) + \alpha(t) \,,$$

then time-reversing $\beta(t)$ and calculating

$$\hat{y}(t \mid T) = z(t) - \beta(t) \,.$$

Filtering

The causal part $\{\mathcal{H}_{OE}\}_+$ of the minimum-variance smoother (88) is given by

$$\begin{aligned} \{\mathcal{H}_{OE}\}_+ &= I - R\{\hat{\Delta}^{-H}\}_+\hat{\Delta}^{-1} \\ &= I - RR^{-1/2}\hat{\Delta}^{-1} \\ &= I - R^{1/2}\hat{\Delta}^{-1} \,. \end{aligned} \tag{92}$$

Employing (89) within (92) leads to the standard minimum-variance filter, namely,

$$\dot{\hat{x}}(t \mid t) = (A(t) - K(t)C(t))\hat{x}(t \mid t) + K(t)z(t) \tag{93}$$

$$\hat{y}(t \mid t) = C(t)\hat{x}(t \mid t) \,. \tag{94}$$

Input Estimation

As discussed in Chapters 1 and 2, input estimates can be found using $\mathcal{G}_1$ = I, and substituting $\hat{\Delta}$ for Δ within (76) yields the solution

$$\mathcal{H}_{IE} = Q\mathcal{G}_2^{-1}(\hat{\Delta}\hat{\Delta}^H)^{-1} = Q\mathcal{G}_2^{-1}\hat{\Delta}^{-H}\hat{\Delta}^{-1} \,. \tag{95}$$

As expected, the low-measurement-noise-asymptote of this equaliser is given by $\lim_{R \to 0} \mathcal{H}_{IE} = \mathcal{G}_2^{-1}$. That is, at high signal-to-noise-ratios the equaliser approaches $\mathcal{G}_2^{-1}$, provided the inverse exists.

The development of a differential equation for the smoothed input estimate, $\hat{w}(t \mid T)$, makes use of the following formula [27] for the cascade of two systems. Suppose that two linear

---

"Ten geographers who think the world is flat will tend to reinforce each others errors….Only a sailor can set them straight." *John Ralston Saul* systems $\mathcal{G}_1$ and $\mathcal{G}_2$ have state-space parameters $\begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix}$ and $\begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix}$, respectively. Then $\mathcal{G}_2\mathcal{G}_1$ is parameterised by $\begin{bmatrix} A_1 & 0 & B_1 \\ B_2C_1 & A_2 & B_2D_1 \\ D_2C_1 & C_2 & D_2D_1 \end{bmatrix}$. It follows that $\hat{w}(t\,|\,T) = Q\mathcal{G}^H\hat{\Delta}^{-H}\alpha(t)$ is realised by

$$\begin{bmatrix} -\dot{\xi}(t) \\ -\dot{\gamma}(t) \\ -\hat{w}(t\,|\,T) \end{bmatrix} = \begin{bmatrix} A^T(t) - C^T(t)K^T(t) & 0 & -C^T(t)R^{-1/2}(t) \\ -C^T(t)K^T(t) & A^T(t) & C^T(t)R^{-1/2}(t) \\ Q(t)D^T(t)K^T(t) & Q(t)B^T(t) & Q(t)D^T(t)R^{-1/2}(t) \end{bmatrix} \begin{bmatrix} \xi(t) \\ \gamma(t) \\ \alpha(t) \end{bmatrix}. \tag{96}$$

in which $\gamma(t) \in \mathbb{R}^n$ is an auxiliary state.

Procedure 2. Input estimates can be calculated via the following two steps.

Step 1. Operate $\hat{\Delta}^{-1}$ on the measurements $z(t)$ using (89) to obtain $\alpha(t)$.

Step 2. In lieu of (96), operate the adjoint of (96) on the time-reversed transpose of $\alpha(t)$. Then take the time-reversed transpose of the result.

State Estimation

Smoothed state estimates can be obtained by defining the reference system $\mathcal{G}_1$ within (76) as

$$\dot{\hat{x}}(t\,|\,T) = A(t)\hat{x}(t\,|\,T) + B(t)\hat{w}(t\,|\,T). \tag{97}$$

That is, a smoother for state estimation is given by (89), (96) and (97). In frequency-domain estimation problems, minimum-order solutions are found by exploiting pole-zero cancellations, see Example 1.13 of Chapter 1. Here in the time-domain, (89), (96), (97) is not a minimum-order solution and some numerical model order reduction may be required.

Suppose that $C(t)$ is of rank n and $D(t) = 0$. In this special case, an n^2-order solution for state estimation can be obtained from (91) and

$$\hat{x}(t\,|\,T) = C^{\#}(t)\hat{y}(t\,|\,T), \tag{98}$$

where

$$C^{\#}(t) = \left(C^T(t)C(t)\right)^{-1} C^T(t) \tag{99}$$

denotes the Moore-Penrose pseudoinverse.

"In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual." *Galileo Galilei*

6.5.4.4 Performance

An analysis of minimum-variance smoother performance requires an identity which is described after introducing some additional notation. Let $\alpha = \mathcal{G}_0 w$ denote the output of linear time-varying system having the realisation

$$\dot{x}(t) = A(t)x(t) + w(t) \tag{100}$$

$$\alpha(t) = x(t), \tag{101}$$

where $w(t) \in \mathbb{R}^n$ and $A(t) \in \mathbb{R}^{n\times n}$. By inspection of (100) – (101), the output of the inverse system $w = \mathcal{G}_0^{-1} y$ is given by

$$w(t) = \dot{\alpha}(t) - A(t)\alpha(t). \tag{102}$$

Similarly, let $\beta = \mathcal{G}_0^H u$ denote the output of the adjoint system $\mathcal{G}_0^H$, which from Lemma 1 has the realisation

$$-\dot{\zeta}(t) = A^T(t)\zeta(t) + u(t) \tag{103}$$

$$\beta(t) = \zeta(t). \tag{104}$$

It follows that the output of the inverse system $u = \mathcal{G}_0^{-H}\beta$ is given by

$$u(t) = -\dot{\beta}(t) - A^T(t)\beta(t). \tag{105}$$

The following identity is required in the characterisation of smoother performance

$$-P(t)A^T(t) - A(t)P(t) = P(t)\mathcal{G}_0^{-H} + \mathcal{G}_0^{-1}P(t), \tag{106}$$

where $P(t)$ is an arbitrary matrix of compatible dimensions. The above equation can be verified by using (102) and (105) within (106). Using the above notation, the exact Wiener-Hopf factor satisfies

$$\Delta\Delta^H = C(t)\mathcal{G}_0 B(t)Q(t)B^T(t)\mathcal{G}_0^H C^T(t) + R(t). \tag{107}$$

It is observed below that the approximate Wiener-Hopf factor (86) approaches the exact Wiener Hopf-factor whenever the problem is locally stationary, that is, whenever $A(t)$, $B(t)$, $C(t)$, $Q(t)$ and $R(t)$ change sufficiently slowly, so that $\dot{P}(t)$ of (87) approaches the zero matrix.

Lemma 10 [8]: *In respect of the signal model (1) – (2) with $D(t) = 0$, $E\{w(t)\} = E\{v(t)\} = 0$, $E\{w(t)w^T(t)\} = Q(t)$, $E\{v(t)v^T(t)\} = R(t)$, $E\{w(t)v^T(t)\} = 0$ and the quantities defined above,*

$$\hat{\Delta}\hat{\Delta}^H = \Delta\Delta^H - C(t)\mathcal{G}_0 \dot{P}(t)\mathcal{G}_0^H C^T(t). \tag{108}$$

"Every great advance in natural knowledge has involved the absolute rejection of authority." *Thomas Henry Huxley*

Proof: *The approximate Wiener-Hopf factor may be written as* $\hat{\Delta} = C(t)\mathcal{G}_0 K(t)R^{1/2}(t) + R^{1/2}(t)$. *It is easily shown that* $\hat{\Delta}\hat{\Delta}^H = C(t)\mathcal{G}_0(P\mathcal{G}_0^{-H} + \mathcal{G}_0^{-1}P + K(t)R(t)K^T(t))\mathcal{G}_0^H C^T(t)$ *and using (106) gives* $\hat{\Delta}\hat{\Delta}^H = C(t)\mathcal{G}_0(B(t)Q(t)B^T(t) - \dot{P}(t)\mathcal{G}_0^H C^T(t) + R(t)$. *The result follows by comparing* $\hat{\Delta}\hat{\Delta}^H$ *and (107).* □

Consequently, the minimum-variance smoother (88) achieves the best-possible estimator performance, namely $\left\|\mathcal{R}_{ei2}\mathcal{R}_{ei2}^H\right\|_2 = 0$, whenever the problem is locally stationary.

Lemma 11 [8]: *The output estimation smoother (88) satisfies*

$$\mathcal{R}_{ei2} = R(t)[(\Delta\Delta^H)^{-1} - (\Delta\Delta^H - C(t)\mathcal{G}_0 \dot{P}(t)\mathcal{G}_0^H C^T(t))^{-1}]\Delta . \tag{109}$$

Proof: *Substituting (88) into (77) yields*

$$\mathcal{R}_{ei2} = R(t)[(\Delta\Delta^H)^{-1} - (\hat{\Delta}\hat{\Delta}^H)^{-1}]\Delta . \tag{110}$$

The result is now immediate from (108) and (110). □

Conditions for the convergence of the Riccati difference equation solution (87) and hence the asymptotic optimality of the smoother (88) are set out below.

Lemma 12 [8]: *Let* $S(t) = C^T(t)R^{-1}(t)C(t)$. *If(i) there exist solutions* $P(t) \ge P(t+\delta_t)$ *of (87) for a* $t > \delta_t > 0$; *and*

(ii)

$$\begin{bmatrix} Q(t) & A(t) \\ A^T(t) & -S(t) \end{bmatrix} \ge \begin{bmatrix} Q(t+\delta_t) & A(t+\delta_t) \\ A^T(t+\delta_t) & -S(t+\delta_t) \end{bmatrix} \tag{111}$$

for all $t > \delta_t$ *then*

$$\lim_{t\to\infty}\left\|\mathcal{R}_{ei2}\mathcal{R}_{ei2}^H\right\|_2 = 0 . \tag{112}$$

Proof: *Conditions (i) and (ii) together with Theorem 1 imply* $P(t) \ge P(t+\delta_t)$ *for all* $t > \delta_t$ *and* $\lim_{t\to\infty}\dot{P}(t) = 0$. *The claim (112) is now immediate from Lemma 11.* □

6.5.5 Performance Comparison

The following scalar time-invariant examples compare the performance of the minimum-variance filter (92), maximum-likelihood smoother (50), Fraser-Potter smoother (73) and minimum-variance smoother (88) under Gaussian and nongaussian noise conditions.

Example 5 [9]. Suppose that $A = -1$ and $B = C = Q = 1$. Simulations were conducted using $T = 100$ s, $\delta t = 1$ ms and 1000 realisations of Gaussian noise processes. The mean-square-error (MSE) exhibited by the filter and smoothers as a function of the input signal-to-noise ratio

"The definition of insanity is doing the same thing over and over again and expecting different results." *Albert Einstein*

(SNR) is shown in Fig. 2. As expected, it can be seen that the smoothers outperform the filter. Although the minimum-variance smoother exhibits the lowest mean-square error, the performance benefit diminishes at high signal-to-noise ratios.

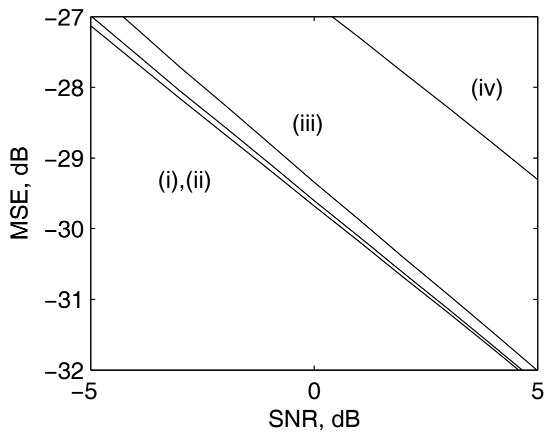


Figure 2. MSE versus SNR for Example 4: (i) minimum-variance smoother, (ii) Fraser-Potter smoother, (iii) maximum-likelihood smoother and (iv) minimum-variance filter.

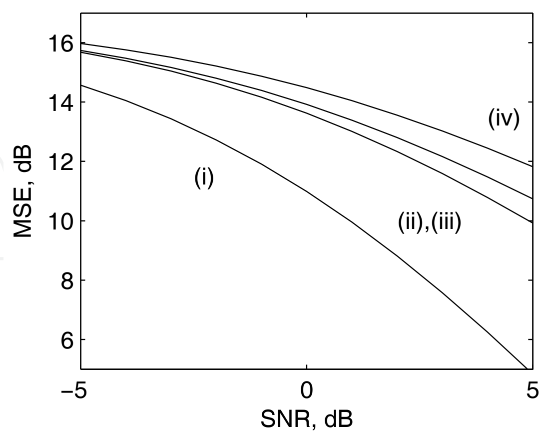


Figure 3. MSE versus SNR for Example 5: (i) minimum-variance smoother, (ii) Fraser-Potter smoother, (iii) maximum-likelihood smoother and (iv) minimum-variance filter.

Example 6 [9]. Suppose instead that the process noise is the unity-variance deterministic signal $w(t)$ = $\sin(t)\sigma_{\sin(t)}^{-1}$, where $\sigma_{\sin(t)}^{2}$ denotes the sample variance of sin(t). The results of simulations employing the sinusoidal process noise and Gaussian measurement noise are shown in Fig. 3. Once again, the smoothers exhibit better performance than the filter. It can be seen that the minimum-variance smoother provides the best mean-square-error performance. The minimum-variance smoother appears to be less perturbed by nongaussian noises because it does not rely on assumptions about the underlying distributions.

6.6 Conclusion

The fixed-point smoother produces state estimates at some previous point in time, that is,

$$\hat{\xi}(t) = \Sigma(t)C^T(t)R^{-1}(t)\big(z(t) - C(t)\hat{x}(t\,|\,t)\big),$$

where $\Sigma(t)$ is the smoother error covariance.

In fixed-lag smoothing, state estimates are calculated at a fixed time delay τ behind the current measurements. This smoother has the form

$$\dot{\hat{x}}(t\,|\,t+\tau) = A(t)\dot{\hat{x}}(t\,|\,t+\tau) + B(t)Q(t)B^T(t)P^{-1}(t)\big(\hat{x}(t\,|\,t+\tau) - \hat{x}(t\,|\,t)\big)$$

$$+P(t)\Phi^T(t+\tau, t)C^T(t+\tau)R^{-1}(t+\tau)\big(z(t+\tau) - C(t+\tau)\hat{x}(t+\tau)\big),$$

where $\Phi(t + \tau, t)$ is the transition matrix of the minimum-variance filter.

"He who rejects change is the architect of decay. The only human institution which rejects progress is the cemetery." *James Harold Wilson*

Three common fixed-interval smoothers are listed in Table 1, which are for retrospective (or off-line) data analysis. The Rauch-Tung-Streibel (RTS) smoother and Fraser-Potter (FP) smoother are minimum-order solutions. The RTS smoother differential equation evolves backward in time, in which $G(\tau) = B(\tau)Q(\tau)B^T(\tau)P^{-1}(\tau)$ is the smoothing gain. The FP smoother employs a linear combination of forward state estimates and backward state estimates obtained by running a filter over the time-reversed measurements. The optimum minimum-variance solution, in which $\bar{A}(t) = A(t) - K(t)C(t)$, where K($t$) is the predictor gain, involves a cascade of forward and adjoint predictions. It can be seen that the optimum minimum-variance smoother is the most complex and so any performance benefits need to be reconciled with the increased calculation cost.

	ASSUMPTIONS	MAIN RESULTS
Signals and system	$E\{w(t)\} = E\{v(t)\} = 0$. $E\{w(t)w^T(t)\} = Q(t) > 0$ and $E\{v(t)v^T(t)\} = R(t) > 0$ are known. $A(t)$, $B(t)$ and $C(t)$ are known.	$\dot{x}(t) = A(t)x(t) + B(t)w(t)$ $y(t) = C(t)x(t)$ $z(t) = y(t) + v(t)$
RTS smoother	Assumes that the filtered and smoothed states are normally distributed. $\hat{x}(t \mid t)$ previously calculated by Kalman filter.	$-\dot{\hat{x}}(\tau \mid T) = A(\tau)\hat{x}(\tau \mid T) + G(\tau)\big(\hat{x}(\tau \mid T) - \hat{x}(\tau \mid \tau)\big)$
FP smoother	$\hat{x}(t \mid t)$ previously calculated by Kalman filter.	$\hat{x}(t \mid T) = (P^{-1}(t \mid t) + \Sigma^{-1}(t \mid t))^{-1}$ $\times (P^{-1}(t \mid t)\hat{x}(t \mid t) + \Sigma^{-1}(t \mid t)\xi(t \mid t))$
Optimal smoother		$\begin{bmatrix} \dot{\hat{x}}(t) \\ \alpha(t) \end{bmatrix} = \begin{bmatrix} \bar{A}(t) & K(t) \\ -R^{-1/2}(t)C(t) & R^{-1/2}(t) \end{bmatrix} \begin{bmatrix} \hat{x}(t) \\ z(t) \end{bmatrix}$ $\begin{bmatrix} -\dot{\xi}(t) \\ \beta(t) \end{bmatrix} = \begin{bmatrix} \bar{A}^T(t) & -C^T(t)R^{-1/2}(t) \\ K^T(t) & R^{-1/2}(t) \end{bmatrix} \begin{bmatrix} \xi(t) \\ \alpha(t) \end{bmatrix}$ $\hat{y}(t \mid T) = z(t) - R(t)\beta(t)$

Table 1. Continuous-time fixed-interval smoothers.

The output estimation error covariance for the general estimation problem can be written as $\mathcal{R}_{ei}\mathcal{R}_{ei}^H = \mathcal{R}_{ei1}\mathcal{R}_{ei1}^H + \mathcal{R}_{ei2}\mathcal{R}_{ei2}^H$, where $\mathcal{R}_{ei1}\mathcal{R}_{ei1}^H$ specifies a lower performance bound and

"Remember a dead fish can float downstream but it takes a live one to swim upstream." *William Claude Fields*

$\mathcal{R}_{ei2}\mathcal{R}_{ei2}^H$ is a function of the estimator solution. The optimal smoother solution achieves $\left\| \{\mathcal{R}_{ei2}\mathcal{R}_{ei2}^H\}_+ \right\|_2$ = 0 and provides the best mean-square-error performance, provided of course that the problem assumptions are correct. The minimum-variance smoother solution also attains best-possible performance whenever the problem is locally stationary, that is, when $A(t)$, $B(t)$, $C(t)$, $Q(t)$ and $R(t)$ change sufficiently slowly.

6.7 Problems

Problem 1. Write down augmented state-space matrices $A^{(a)}(t)$, $B^{(a)}(t)$ and $C^{(a)}(t)$ for the continuous-time fixed-point smoother problem.

(i) Substitute the above matrices into $\dot{P}^{(a)}(t) = A^{(a)}(t)P^{(a)}(t) + P^{(a)}(t)(A^{(a)})^T(t) - P^{(a)}(t)(C^{(a)})^T(t)R^{-1}(t)C^{(a)}(t)P^{(a)}(t) + B^{(a)}(t)Q(t)(B^{(a)}(t))^T$ to obtain the component Riccati differential equations.

(ii) Develop expressions for the continuous-time fixed-point smoother estimate and the smoother gain.

Problem 2. The Hamiltonian equations (60) were derived from the forward version of the maximum likelihood smoother (54). Derive the alternative form

$$\begin{bmatrix} -\dot{\hat{x}}(t\,|\,T) \\ \dot{\xi}(t\,|\,T) \end{bmatrix} = \begin{bmatrix} A(t) & B(t)Q(t)B^T(t) \\ C^T(t)R^{-1}(t)C(t) & A^T(t) \end{bmatrix} \begin{bmatrix} \hat{x}(t\,|\,T) \\ \xi(t\,|\,T) \end{bmatrix} - \begin{bmatrix} 0 \\ C^T(t)R^{-1}(t)z(t) \end{bmatrix}.$$

from the backward smoother (50). Hint: use the backward Kalman-Bucy filter and the backward Riccati equation.

Problem 3. It is shown in [6] and [17] that the intermediate variable within the Hamiltonian equations (60) is given by

$$\xi(t\,|\,T) = \int_t^T \Phi^T(s,t)C^T(s)R^{-1}(s)(z(s) - C(s)\hat{x}(s\,|\,s))ds\,,$$

where $\Phi^T(s,t)$ is the transition matrix of the Kalman-Bucy filter. Use the above equation to derive

$$-\dot{\xi}(t\,|\,T) = -C^T(t)R^{-1}(t)C(t)\hat{x}(t\,|\,T) + A^T(t)\xi(t) - C^T(t)R^{-1}(t)z(t)\,.$$

Problem 4. Show that the adjoint of system having state space parameters $\begin{bmatrix} A(t) & B(t) \\ C(t) & D(t) \end{bmatrix}$ is parameterised by $\begin{bmatrix} -A^T(t) & -C^T(t) \\ B^T(t) & D^T(t) \end{bmatrix}$.

"It is not the strongest of the species that survive, nor the most intelligent, but the most responsive to change." *Charles Robert Darwin*

Problem 5. Suppose $\mathcal{G}_0$ is a system parameterised by $\begin{bmatrix} A(t) & I \\ I & 0 \end{bmatrix}$, show that $-P(t)A^T(t)$ - $A(t)P(t)$ = $P(t)\mathcal{G}_0^{-H}$ + $\mathcal{G}_0^{-1}P(t)$.

Problem 6. The optimum minimum-variance smoother was developed by finding the solution that minimises $\left\| \tilde{y}_2 \tilde{y}_2^H \right\|_2$. Use the same completing-the-square approach to find the optimum minimum-variance filter. (Hint: Find the solution that minimises $\left\| \tilde{y}_2 \tilde{y}_2^T \right\|_2$.)

Problem 7 [9]. Derive the output estimation minimum-variance filter by finding a solution Let $a \in \mathbb{R}$, $b = 1$, $c \in \mathbb{R}$ and $d = 0$ denote the time-invariant state-space parameters of the plant $\mathcal{G}$. Denote the error covariance, gain of the Kalman filter and gain of the maximum-likelihood smoother by p, k and g, respectively. Show that

$$H_1(s) = k(s-a+kc)^{-1},$$

$$H_2(s) = cgk(-s-a+g)^{-1}(s-a+kc)^{-1},$$

$$H_3(s) = kc(-a + kc)(s - a + kc)^{-1}(-s - a + kc)^{-1},$$

$$H_4(s) = ((-a + kc)^2 - (-a + kc - k)^2)(s - a + kc)^{-1}(-s - a + kc)^{-1}$$

are the transfer functions of the Kalman filter, maximum-likelihood smoother, the Fraser-Potter smoother and the minimum variance smoother, respectively.

Problem 8.

(i) Develop a state-space formulation of an approximate Wiener-Hopf factor for the case when the plant includes a nonzero direct feedthrough matrix (that is, $D(t) \neq 0$).

(ii) Use the matrix inversion lemma to obtain the inverse of the approximate Wiener-Hopf factor for the minimum-variance smoother.

6.8 Glossary

$p(x(t))$	Probability density function of a continuous random variable $x(t)$.
$x(t) \sim \mathcal{N}(\mu, R_{xx})$	The random variable $x(t)$ has a normal distribution with mean μ and covariance R_{xx}.
$f(x(t))$	Cumulative distribution function or likelihood function of $x(t)$.
$\hat{x}(t \mid t+\tau)$	Estimate of $x(t)$ at time t given data at fixed time lag τ.
$\hat{x}(t \mid T)$	Estimate of $x(t)$ at time t given data over a fixed interval T.
$\hat{w}(t \mid T)$	Estimate of $w(t)$ at time t given data over a fixed interval T.

"Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road." *Stewart Brand*

$G(t)$	Gain of the minimum-variance smoother developed by Rauch, Tung and Striebel.
$\mathcal{R}_{ei}$	A linear system that operates on the inputs $i = \begin{bmatrix} v^T & w^T \end{bmatrix}^T$ and generates the output estimation error e.
$C^{\#}(t)$	Moore-Penrose pseudoinverse of $C(t)$.
Δ	The Wiener-Hopf factor which satisfies $\Delta\Delta^H = \mathcal{G}Q\mathcal{G}^H + R$.
$\hat{\Delta}$	Approximate Wiener-Hopf factor.
$\hat{\Delta}^H$	The adjoint of $\hat{\Delta}$.
$\hat{\Delta}^{-H}$	The inverse of $\hat{\Delta}^H$.
$\{\hat{\Delta}^{-H}\}_+$	The causal part of $\hat{\Delta}^{-H}$.

6.9 References

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

[2] T. Kailath, "A View of Three Decades of Linear Filtering Theory", *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 146 – 181, Mar., 1974.

[3] H. E. Rauch, F. Tung and C. T. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems", *AIAA Journal*, vol. 3, no. 8, pp. 1445 – 1450, Aug., 1965.

[4] D. C. Fraser and J. E. Potter, "The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters", *IEEE Transactions on Automatic Control*, vol. AC-14, no. 4, pp. 387 – 390, Aug., 1969.

[5] J. S. Meditch, "On Optimal Fixed Point Linear Smoothing", *International Journal of Control*, vol. 6, no. 2, pp. 189 – 199, 1967.

[6] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*, McGraw-Hill Book Company, New York, 1971.

[7] J. B. Moore, "Fixed-Lag Smoothing Results for Linear Dynamical Systems", *A.T.R.*, vol. 7, no. 2, pp. 16 – 21, 1973.

[8] G. A. Einicke, "Asymptotic Optimality of the Minimum-Variance Fixed-Interval Smoother", *IEEE Transactions on Signal Processing,* vol. 55, no. 4, pp. 1543 – 1547, Apr. 2007.

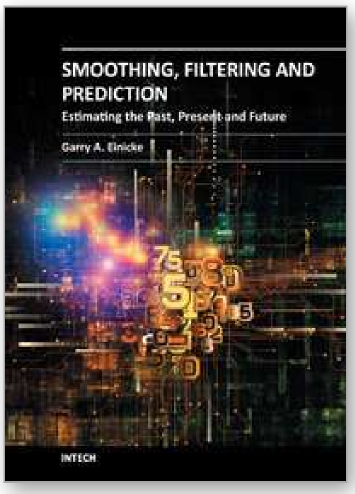
[9] G. A. Einicke, J. C. Ralston, C. O. Hargrave, D. C. Reid and D. W. Hainsworth, "Longwall Mining Automation, An Application of Minimum-Variance Smoothing", *IEEE Control Systems Magazine*, vol. 28, no. 6, pp. 28 – 37, Dec. 2008.

[10] G. A. Einicke, "A Solution to the Continuous-Time H-infinity Fixed-Interval Smoother Problem", *IEEE Transactions on Automatic Control,* vo. 54, no. 12, pp. 2904 – 2908, Dec. 2009.

"I don't want to be left behind. In fact, I want to be here before the action starts." *Kerry Francis Bullmore Packer*

[11] D. J. N. Limebeer, B. D. O. Anderson, P. Khargonekar and M. Green, "A Game Theoretic Approach to H_∞ Control for Time-varying Systems", *SIAM Journal on Control and Optimization*, vol. 30, no. 2, pp. 262 – 283, 1992.
[12] G. Freiling, G. Jank and H. Abou-Kandil, "Generalized Riccati Difference and Differential Equations", *Linear Algebra And Its Applications*, pp. 291 – 303, 199
[13] L. E. Zachrisson, "On Optimal Smoothing of Continuous Time Kalman Processes, *Information Sciences*, vol. 1, pp. 143 – 172, 1969.
[14] U. Shaked and Y. Theodore, "H_∞ – Optimal Estimation: A Tutorial", *Proc. 31st IEEE Conference on Decision Control*, Ucson, Arizona, pp. 2278 – 2286, Dec. 1992.
[15] J. S. Meditch, "On Optimal Linear Smoothing Theory", *Information and Control*, vol. 10, pp. 598 – 615, 1967.
[16] P. S. Maybeck, *Stochastic Models, Estimation and Control*, Vol. 2, Academic press, New York, 1982.
[17] T. Kailath, A. H. Sayed and B. Hassibi, *Linear Estimation*, Pearson Prentice Hall, New Jersey, 2000.
[18] B. D. O. Anderson, "Properties of Optimal Linear Smoothing", *IEEE Transactions on Automatic Control*, vol. 14, no. 1, pp. 114 – 115, Feb. 1969.
[19] S. Chirarattananon and B. D. O. Anderson, "Stable Fixed-Lag Smoothing of Continuous-time Processes", IEEE *Transactions on Information Theory*, vol. 20, no. 1, pp. 25 – 36, Jan. 1974.
[20] A. Gelb, *Applied Optimal Estimation*, The Analytic Sciences Corporation, USA, 1974.
[21] L. Ljung and T. Kailath, "A Unified Approach to Smoothing Formulas", *Automatica*, vol. 12, pp. 147 – 157, 197
[22] T. Kailath and P. Frost, "An Innovations Approach to Least-Squares Estimation Part II: Linear Smoothing in Additive White Noise", *IEEE Transactions on Automat. Control*, vol. 13, no. 6, pp. 655 – 660, 1968.
[23] F. L. Lewis, L. Xie and D. Popa, *Optimal and Robust Estimation With an Introduction to Stochastic Control Theory*, Second Edition, CRC Press, Taylor & Francis Group, 2008.
[24] F. A. Badawi, A. Lindquist and M. Pavon, "A Stochastic Realization Approach to the Smoothing Problem", *IEEE Transactions on Automatic Control*, vol. 24, no. 6, Dec. 1979.
[25] J. A. Rice, *Mathematical Statistics and Data Analysis* (Second Edition), Duxbury Press, Belmont, California, 1995.
[26] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, Second Edition, John Wiley & Sons, Inc., New York, 1992.
[27] M. Green and D. J. N. Limebeer, *Linear Robust Control*, Prentice-Hall Inc, Englewood Cliffs, New Jersey, 1995.

"In times of profound change, the learners inherit the earth, while the learned find themselves beautifully equipped to deal with a world that no longer exists." *Al Rogers*



Smoothing, Filtering and Prediction - Estimating The Past, Present and Future

Edited by

ISBN 978-953-307-752-9
Hard cover, 276 pages
Publisher InTech
Published online 24, February, 2012
Published in print edition February, 2012

This book describes the classical smoothing, filtering and prediction techniques together with some more recently developed embellishments for improving performance within applications. It aims to present the subject in an accessible way, so that it can serve as a practical guide for undergraduates and newcomers to the field. The material is organised as a ten-lecture course. The foundations are laid in Chapters 1 and 2, which explain minimum-mean-square-error solution construction and asymptotic behaviour. Chapters 3 and 4 introduce continuous-time and discrete-time minimum-variance filtering. Generalisations for missing data, deterministic inputs, correlated noises, direct feedthrough terms, output estimation and equalisation are described. Chapter 5 simplifies the minimum-variance filtering results for steady-state problems. Observability, Riccati equation solution convergence, asymptotic stability and Wiener filter equivalence are discussed. Chapters 6 and 7 cover the subject of continuous-time and discrete-time smoothing. The main fixed-lag, fixed-point and fixed-interval smoother results are derived. It is shown that the minimum-variance fixed-interval smoother attains the best performance. Chapter 8 attends to parameter estimation. As the above-mentioned approaches all rely on knowledge of the underlying model parameters, maximum-likelihood techniques within expectation-maximisation algorithms for joint state and parameter estimation are described. Chapter 9 is concerned with robust techniques that accommodate uncertainties within problem specifications. An extra term within Riccati equations enables designers to trade-off average error and peak error performance. Chapter 10 rounds off the course by applying the afore-mentioned linear techniques to nonlinear estimation problems. It is demonstrated that step-wise linearisations can be used within predictors, filters and smoothers, albeit by forsaking optimal performance guarantees.

Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821