3

Continuous-Time Minimum-Variance Filtering

3.1 Introduction

Rudolf E. Kalman studied discrete-time linear dynamic systems for his master's thesis at MIT in 1954. He commenced work at the Research Institute for Advanced Studies (RIAS) in Baltimore during 1957 and nominated Richard S. Bucy to join him in 1958 [1]. Bucy recognised that the nonlinear ordinary differential equation studied by an Italian mathematician, Count Jacopo F. Riccati, in around 1720, now called the Riccati equation, is equivalent to the Wiener-Hopf equation for the case of finite dimensional systems [1], [2]. In November 1958, Kalman recasted the frequency domain methods developed by Norbert Wiener and Andrei N. Kolmogorov in the 1940s to state-space form [2]. Kalman noted in his 1960 paper [3] that generalising the Wiener solution to nonstationary problems was difficult, which motivated his development of the optimal discrete-time filter in a state-space framework. He described the continuous-time version with Bucy in 1961 [4] and published a generalisation in 1963 [5]. Bucy later investigated the monotonicity and stability of the underlying Riccati equation [6]. The continuous-time minimum-variance filter is now commonly attributed to both Kalman and Bucy.

Compared to the Wiener Filter, Kalman's state-space approach has the following advantages.

- It is applicable to time-varying problems.
- As noted in [7], [8], the state-space parameters can be linearisations of nonlinear models.
- The burdens of spectral factorisation and pole-zero cancelation are replaced by the easier task of solving a Riccati equation.
- It is a more intuitive model-based approach in which the estimated states correspond to those within the signal generation process.

Kalman's research at the RIAS was concerned with estimation and control for aerospace systems which was funded by the Air Force Office of Scientific Research. His explanation of why the dynamics-based Kalman filter is more important than the purely stochastic Wiener filter is that "Newton is more important than Gauss" [1]. The continuous-time Kalman filter produces state estimates $\hat{x}(t)$ from the solution of a simple differential equation

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + K(t)\big(z(t) - C(t)\hat{x}(t)\big),$$

"What a weak, credulous, incredulous, unbelieving, superstitious, bold, frightened, what a ridiculous world ours is, as far as concerns the mind of man. How full of inconsistencies, contradictions and absurdities it is. I declare that taking the average of many minds that have recently come before me ... I should prefer the obedience, affections and instinct of a dog before it." *Michael Faraday* in which it is tacitly assumed that the model is correct, the noises are zero-mean, white and uncorrelated. It is straightforward to include nonzero means, coloured and correlated noises. In practice, the true model can be elusive but a simple (low-order) solution may return a cost benefit.

The Kalman filter can be derived in many different ways. In an early account [3], a quadratic cost function was minimised using orthogonal projections. Other derivation methods include deriving a maximum *a posteriori* estimate, using Itô's calculus, calculus-of-variations, dynamic programming, invariant imbedding and from the Wiener-Hopf equation [6] - [17]. This chapter provides a brief derivation of the optimal filter using a conditional mean (or equivalently, a least mean square error) approach.

The developments begin by introducing a time-varying state-space model. Next, the state transition matrix is defined, which is used to derive a Lyapunov differential equation. The Kalman filter follows immediately from a conditional mean formula. Its filter gain is obtained by solving a Riccati differential equation corresponding to the estimation error system. Generalisations for problems possessing deterministic inputs, correlated process and measurement noises, and direct feedthrough terms are described subsequently. Finally, it is shown that the Kalman filter reverts to the Wiener filter when the problems are time-invariant.

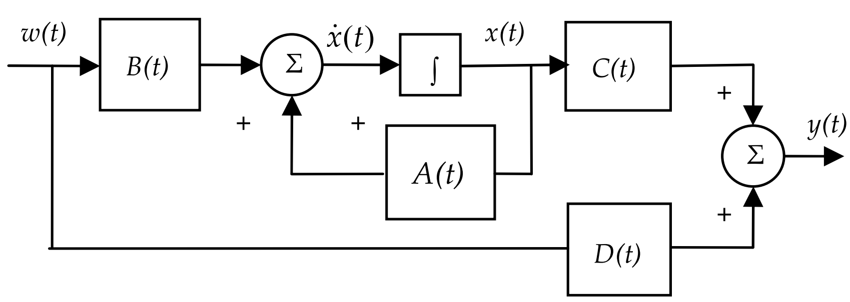


Figure 1. The continuous-time system $\mathcal{G}$ operates on the input signal $w(t) \in \mathbb{R}^m$ and produces the output signal $y(t) \in \mathbb{R}^p$.

3.2 Prerequisites

3.2.1 The Time-varying Signal Model

The focus initially is on time-varying problems over a finite time interval $t \in [0, T]$. A system $\mathcal{G} : \mathbb{R}^m \to \mathbb{R}^p$ is assumed to have the state-space representation

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

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

where $A(t) \in \mathbb{R}^{n\times n}$, $B(t) \in \mathbb{R}^{n\times m}$, $C(t) \in \mathbb{R}^{p\times n}$, $D(t) \in \mathbb{R}^{p\times p}$ and $w(t)$ is a zero-mean white process noise with $E\{w(t)w^T(\tau)\} = Q(t)\delta(t - \tau)$, in which $\delta(t)$ is the Dirac delta function. This

"A great deal of my work is just playing with equations and seeing what they give." *Paul Arien Maurice Dirac* system in depicted in Fig. 1. In many problems of interest, signals are band-limited, that is, the direct feedthrough matrix, $D(t)$, is zero. Therefore, the simpler case of $D(t) = 0$ is addressed first and the inclusion of a nonzero $D(t)$ is considered afterwards.

3.2.2 The State Transition Matrix

The state transition matrix is introduced below which concerns the linear differential equation (1).

Lemma 1: *The equation (1) has the solution*

$$x(t) = \Phi(t,t_0)x(t_0) + \int_{t_0}^{t} \Phi(t,s)B(s)w(s)ds \,, \tag{3}$$

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

$$\dot{\Phi}(t,t_0) = \frac{d\Phi(t,t_0)}{dt} = A(t)\Phi(t,t_0) \,, \tag{4}$$

with boundary condition

$$\Phi(t,t) = I. \tag{5}$$

Proof: *Differentiating both sides of (3) and using Leibnitz's rule, that is,* $\frac{\partial}{\partial t}\int_{\alpha(t)}^{\beta(t)} f(t,\tau)d\tau$ = $\int_{\alpha(t)}^{\beta(t)} \frac{\partial f(t,\tau)}{\partial t} d\tau - f(t,\alpha)\frac{d\alpha(t)}{dt} + f(t,\beta)\frac{d\beta(t)}{dt}$*, gives*

$$\dot{x}(t) = \dot{\Phi}(t,t_0)x(t_0) + \int_{t_0}^{t} \dot{\Phi}(t,\tau)B(\tau)w(\tau)d\tau + \Phi(t,t)B(t)w(t) \,. \tag{6}$$

Substituting (4) and (5) into the right-hand-side of (6) results in

$$\dot{x}(t) = A(t)\left(\Phi(t,t_0)x(t_0) + \int_{t_0}^{t} \Phi(t,\tau)B(\tau)w(\tau)d\tau\right) + B(t)w(t) \,. \tag{7}$$

□

3.2.3 The Lyapunov Differential Equation

The mathematical expectation, $E\{x(t)x^T(\tau)\}$ of $x(t)x^T(\tau)$, is required below, which is defined as

$$E\{x(t)x^T(\tau)\} = \int_{-\infty}^{\infty} x(t)x^T(\tau) f_{xx} x(t)x^T(\tau))dx(t) \,, \tag{8}$$

where $f_{xx}(x(t)x^T(\tau))$ is the probability density function of $x(t)x^T(\tau)$. A useful property of expectations is demonstrated in the following example.

"Life is good for only two things, discovering mathematics and teaching mathematics." *Siméon Denis Poisson*

Example 1. Suppose that $x(t)$ is a stochastic random variable and $h(t)$ is a continuous function, then

$$E\left\{\int_a^b h(t)x(t)x^T(\tau)dt\right\} = \int_a^b h(t)E\{x(t)x^T(\tau)\}dt \ . \tag{9}$$

To verify this, expand the left-hand-side of (9) to give

$$\begin{aligned} E\left\{\int_a^b h(t)x(t)x^T(\tau)dt\right\} &= \int_{-\infty}^{\infty}\int_a^b h(t)x(t)x^T(\tau)dt f_{xx}(x(t)x^T(\tau))dx(t) \\ &= \int_{-\infty}^{\infty}\int_a^b h(t)x(t)x^T(\tau)f_{xx}(x(t)x^T(\tau)dtdx(t) \ . \end{aligned} \tag{10}$$

Using Fubini's theorem, that is, $\int_c^d\int_a^b g(x,y)dxdy = \int_a^b\int_c^d g(x,y)dydx$, within (10) results in

$$\begin{aligned} E\left\{\int_a^b h(t)x(t)x^T(\tau)dt\right\} &= \int_a^b\int_{-\infty}^{\infty} h(t)x(t)x^T(\tau)f_{xx}(x(t)x^T(\tau))dx(t)dt \\ &= \int_a^b h(t)\int_{-\infty}^{\infty} x(t)x^T(\tau)f_{xx}(x(t)x^T(\tau))dx(t)dt \ . \end{aligned} \tag{11}$$

The result (9) follows from the definition (8) within (11).

The Dirac delta function, $\delta(t) = \begin{cases} \infty & t = 0 \\ 0 & t \neq 0 \end{cases}$, satisfies the identity $\int_{-\infty}^{\infty}\delta(t)dt = 1$. In the foregoing development, use is made of the partitioning

$$\int_{-\infty}^{0}\delta(t)dt = \int_0^{\infty}\delta(t)dt = 0.5 \ . \tag{12}$$

Lemma 2: *In respect of equation (1), assume that $w(t)$ is a zero-mean white process with $E\{w(t)w^T(\tau)\} = Q(t)\delta(t - \tau)$ that is uncorrelated with $x(t_0)$, namely, $E\{w(t)x^T(t_0)\} = 0$. Then the covariances $P(t,\tau) = E\{x(t)x^T(\tau)\}$ and $\dot{P}(t,\tau) = \frac{d}{dt}E\{x(t)x^T(\tau)\}$ satisfy the Lyapunov differential equation*

$$\dot{P}(t,\tau) = A(t)P(t,\tau) + P(t,\tau)A^T(\tau) + B(t)Q(t)B^T(t) \ . \tag{13}$$

Proof: *Using (1) within $\frac{d}{dt}E\{x(t)x^T(\tau)\} = E\{\dot{x}(t)x^T(\tau) + x(t)\dot{x}^T(\tau)\}$ yields*

$$\begin{aligned} \dot{P}(t,\tau) &= E\{A(t)x(t)x^T(\tau) + B(t)w(t)x^T(\tau)\} + E\{x(t)x^T(\tau)A^T(\tau) + x(t)w^T(\tau)B^T(\tau)\} \\ &= A(t)P(t,\tau) + P(t,\tau)A^T(\tau) + E\{B(t)w(t)x^T(\tau)\} + E\{x(t)w^T(\tau)B^T(\tau)\} \ . \end{aligned} \tag{14}$$

"It is a mathematical fact that the casting of this pebble from my hand alters the centre of gravity of the universe." *Thomas Carlyle*

It follows from (1) and (3) that

$$\begin{aligned} E\{B(t)w(t)x^T(\tau)\} &= B(t)E\{w(t)x^T(0)\Phi(t,0)\} + B(t)E\left\{\int_{t_0}^{t} w(t)w^T(\tau)B^T(\tau)\Phi(t,\tau)d\tau\right\} \\ &= B(t)E\{w(t)x^T(0)\Phi(t,0)\} + B(t)\int_{t_0}^{t} E\{w(t)w^T(\tau)\}B^T(\tau)\Phi(t,\tau)d\tau \ . \end{aligned} \tag{15}$$

The assumptions $E\{w(t)x^T(t_0)\} = 0$ *and* $E\{w(t)w^T(\tau)\} = Q(t)\delta(t-\tau)$ *together with (15) lead to*

$$\begin{aligned} E\{(B(t)w(t)x^T(\tau)\} &= B(t)Q(t)\int_{t_0}^{t} \delta(t-\tau)B^T(\tau)\Phi(t,\tau)d\tau \\ &= 0.5B(t)Q(t)B^T(t) \ . \end{aligned} \tag{16}$$

The above Lyapunov differential equation follows by substituting (16) into (14). □

In the case τ = t, denote $P(t,t) = E\{x(t)x^T(t)\}$ and $\dot{P}(t,t) = \frac{d}{dt}E\{x(t)x^T(t)\}$. Then the corresponding Lyapunov differential equation is written as

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) + B(t)Q(t)B^T(t) \ . \tag{17}$$

3.2.4 Conditional Expectations

The minimum-variance filter derivation that follows employs a conditional expectation formula, which is set out as follows. Consider a stochastic vector $[x^T(t) \ \ y^T(t)]^T$ having means and covariances

$$E\left\{\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}\right\} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \tag{18}$$

and

$$E\left\{\begin{bmatrix} x(t)-\bar{x} \\ y(t)-\bar{y} \end{bmatrix}\begin{bmatrix} x^T(t)-\bar{x}^T & y^T(t)-\bar{y}^T \end{bmatrix}\right\} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} . \tag{19}$$

respectively, where $\Sigma_{yx} = \Sigma_{xy}^T$. Suppose that it is desired to obtain an estimate of $x(t)$ given $y(t)$, denoted by $E\{x(t) \mid y(t)\}$, which minimises $E\{(x(t) - E\{x(t) \mid y(t)\})(x(t) - E\{x(t) \mid y(t)\})^T\}$. A standard approach (*e.g.*, see [18]) is to assume that the solution for $E\{x(t) \mid y(t)\}$ is affine to $y(t)$, namely,

$$E\{x(t) \mid y(t)\} = Ay(t) + b, \tag{20}$$

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality." *Albert Einstein* where A and b are unknowns to be found. It follows from (20) that

$$E\{(x(t)-E\{x(t)\mid y(t)\})(x(t)-E\{x(t)\mid y(t)\})^T\}$$
$$=E\{x(t)x^T(t)-x(t)y^T(t)A^T-x(t)b^T-Ay(t)x^T(t)$$
$$+Ay(t)y^T(t)A^T-Ay(t)b^T-bx^T(t)+by^T(t)A^T+bb^T\}\,. \tag{21}$$

Substituting $E\{x(t)x^T(t)\} = \overline{x}\overline{x}^T + \Sigma_{xx}$, $E\{x(t)y^T(t)\} = \overline{x}\overline{y}^T + \Sigma_{xy}$, $E\{y(t)x^T(t)\} = \overline{y}\overline{x}^T + \Sigma_{yx}$, $E\{y(t)y^T(t)\} = \overline{y}\overline{y}^T + \Sigma_{yy}$ into (21) and completing the squares yields

$$E\left\{(x(t)-E\{x(t)\mid y(t)\})(x(t)-E\{x(t)\mid y(t)\})^T\right\}$$
$$=(\overline{x}-A\overline{y}-b)(\overline{x}-A\overline{y}-b)^T+\begin{bmatrix} I & -A\end{bmatrix}\begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}\begin{bmatrix} I \\ -A^T\end{bmatrix}. \tag{22}$$

The second term on the right-hand-side of (22) can be rearranged as

$$\begin{bmatrix} I & -A\end{bmatrix}\begin{bmatrix}\Sigma_{xx} & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}\begin{bmatrix} I \\ -A^T\end{bmatrix}=(A-\Sigma_{xy}\Sigma_{yy}^{-1})\Sigma_{yy}(A-\Sigma_{xy}\Sigma_{yy}^{-1})^T+\Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\,.$$

Thus, the choice $A = \Sigma_{xy}\Sigma_{yy}^{-1}$ and $b = \overline{x}-A\overline{y}$ minimises (22), which gives

$$E\{x(t)\mid y(t)\}=\overline{x}+\Sigma_{xy}\Sigma_{yy}^{-1}\left(y(t)-\overline{y}\right) \tag{23}$$

and

$$E\left\{(x(t)-E\{x(t)\mid y(t)\})(x(t)-E\{x(t)\mid y(t)\})^T\right\}=\Sigma_{xx}+\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\,. \tag{24}$$

The conditional mean estimate (23) is also known as the linear least mean square estimate [18]. An important property of the conditional mean estimate is established below.

Lemma 3 (Orthogonal projections): *In respect of the conditional mean estimate (23), in which the mean and covariances are respectively defined in (18) and (19), the error vector*

$$\tilde{x}(t)=x(t)-E\{x(t)\mid y(t)\}\,. \tag{25}$$

is orthogonal to y(t), that is, $E\{\tilde{x}(t)y^T(t)\} = 0$.

"Statistics: The only science that enables different experts using the same figures to draw different conclusions." *Evan Esar*

Proof [8],[18]: *From (23) and (25), it can be seen that*

$$E\left\{(\tilde{x}(t)-E\{\tilde{x}(t)\})(y(t)-E\{y(t)\})^T\right\} = E\left\{\left(x(t)-\overline{x}-\Sigma_{xy}\Sigma_{yy}^{-1}(y(t)-\overline{y})\right)\left(y(t)-\overline{y}\right)^T\right\}$$
$$= \Sigma_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy}$$
$$= 0 .$$

□

Sufficient background material has now been introduced for the finite-horizon filter (for time-varying systems) to be derived.

3.3 The Continuous-time Minimum-Variance Filter

3.3.1 Derivation of the Optimal Filter

Consider again the linear time-varying system $\mathcal{G}: \mathbb{R}^m \to \mathbb{R}^p$ having the state-space realisation

$$\dot{x}(t) = A(t)x(t) + B(t)w(t) , \tag{26}$$
$$y(t) = C(t)x(t) , \tag{27}$$

where $A(t)$, $B(t)$, $C(t)$ are of appropriate dimensions and $w(t)$ is a white process with

$$E\{w(t)\} = 0, \;\; E\{w(t)w^T(\tau)\} = Q(t)\delta(t-\tau). \tag{28}$$

Suppose that observations

$$z(t) = y(t) + v(t) \tag{29}$$

are available, where $v(t) \in \mathbb{R}^p$ is a white measurement noise process with

$$E\{v(t)\} = 0, E\{v(t)v^T(\tau)\} = R(t)\delta(t-\tau) \tag{30}$$

and

$$E\{w(t)v^T(\tau)\} = 0. \tag{31}$$

The objective is to design a linear system $\mathcal{H}$ that operates on the measurements $z(t)$ and produces an estimate $\hat{y}(t|t) = C(t)\hat{x}(t|t)$ of $y(t) = C(t)x(t)$ given measurements at time t, so that the covariance $E\{e(t|t)e^T(t|t)\}$ is minimised, where $e(t|t) = x(t) - \hat{x}(t|t)$. This output estimation problem is depicted in Fig. 2.

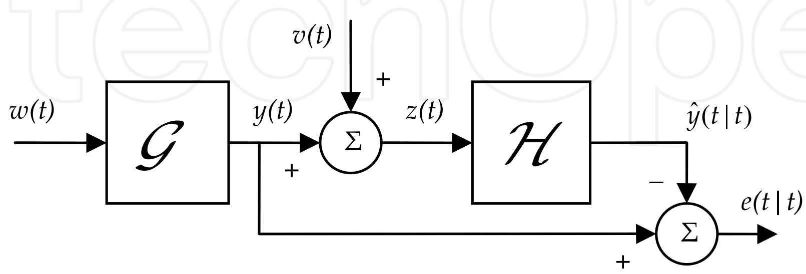


Figure 2. The continuous-time output estimation problem. The objective is to find an estimate $\hat{y}(t|t)$ of $y(t)$ which minimises $E\{(y(t)-\hat{y}(t|t))(y(t)-\hat{y}(t|t))^T\}$.

"Art has a double face, of expression and illusion, just like science has a double face: the reality of error and the phantom of truth." *René Daumal*

It is desired that $\hat{x}(t|t)$ and the estimate $\dot{\hat{x}}(t|t)$ of $\dot{x}(t)$ are unbiased, namely

$$E\{x(t) - \hat{x}(t|t)\} = 0 , \tag{32}$$

$$E\{\dot{x}(t|t) - \dot{\hat{x}}(t|t)\} = 0 . \tag{33}$$

If $\hat{x}(t|t)$ is a conditional mean estimate, from Lemma 3, criterion (32) will be met. Criterion (33) can be satisfied if it is additionally assumed that $E\{\dot{\hat{x}}(t|t)\} = A(t)\hat{x}(t|t)$, since this yields $E\{\dot{x}(t|t) - \dot{\hat{x}}(t|t)\} = A(t)(E\{x(t) - \hat{x}(t|t)\} = 0$. Thus, substituting $E\left\{\begin{bmatrix}\dot{\hat{x}}(t|t) \\ z(t)\end{bmatrix}\right\} = \begin{bmatrix}A(t)\hat{x}(t|t) \\ C(t)\hat{x}(t|t)\end{bmatrix}$ into (23), yields the conditional mean estimate

$$\begin{aligned}\dot{\hat{x}}(t|t) &= A(t)\hat{x}(t|t) + K(t)\big(z(t) - C(t)\hat{x}(t|t)\big) \\ &= \big(A(t) - K(t)C(t)\big)\hat{x}(t|t) + K(t)z(t) ,\end{aligned} \tag{34}$$

where $K(t) = E\{x(t)z^T(t)\}E\{z(t)z^T(t)\}^{-1}$. Equation (34) is known as the continuous-time Kalman filter (or the Kalman-Bucy filter) and is depicted in Fig. 3. This filter employs the state matrix $A(t)$ akin to the signal generating model $\mathcal{G}$, which Kalman and Bucy call the message process [4]. The matrix $K(t)$ is known as the filter gain, which operates on the error residual, namely the difference between the measurement $z(t)$ and the estimated output $C(t)\hat{x}(t)$. The calculation of an optimal gain is addressed in the next section.

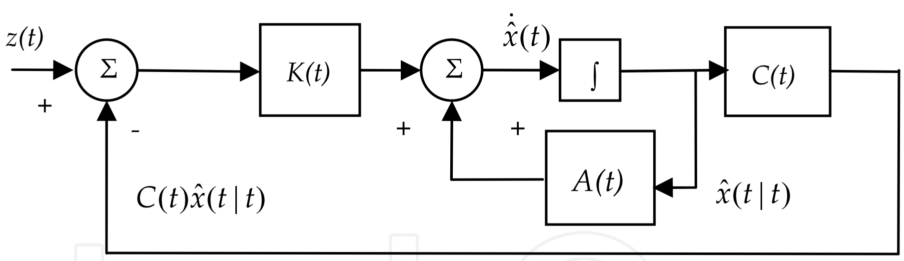


Figure 3. The continuous-time Kalman filter which is also known as the Kalman-Bucy filter. The filter calculates conditional mean estimates $\hat{x}(t|t)$ from the measurements $z(t)$.

3.3.2 The Riccati Differential Equation

Denote the state estimation error by $\tilde{x}(t|t) = x(t) - \hat{x}(t|t)$. It is shown below that the filter minimises the error covariance $E\{\tilde{x}(t|t)\tilde{x}^T(t|t)\}$ if the gain is calculated as

$$K(t) = P(t)C^T(t)R^{-1}(t) , \tag{35}$$

"Somewhere, something incredible is waiting to be known." *Carl Edward Sagan* in which $P(t) = E\{\tilde{x}(t|t)\tilde{x}^T(t|t)\}$ is the solution of the Riccati differential equation

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

Lemma 4: *In respect of the state estimation problem defined by (26) - (31), suppose that there exists a solution*

$$P(t) = P^T(t) \geq 0 \tag{37}$$

for the algebratic Riccati equation (36) satisfying

$$A(t) - P(t)C^T(t) \geq 0 \tag{38}$$

for all t in the interval [0,T]. Then the filter (34) having the gain (35) minimises P(t) = $E\{\tilde{x}(t|t)\tilde{x}^T(t|t)\}$.

Proof: *Subtracting (34) from (26) results in*

$$\dot{\tilde{x}}(t|t) = \big(A(t) - K(t)C(t)\big)\tilde{x}(t|t) + B(t)w(t) - K(t)v(t) . \tag{39}$$

Applying Lemma 2 to the error system (39) gives

$$\dot{P}(t) = \big(A(t) - K(t)C(t)\big)P(t) + P(t)(A(t) - K(t)C(t))^T + K(t)R(t)K^T(t) + B(t)Q(t)B^T(t) \tag{40}$$

which can be rearranged as

$$\begin{aligned}\dot{P}(t) &= A(t)P(t) + P(t)A^T(t) + B(t)Q(t)B^T(t)) \\ &+ \big(K(t) - P(t)C^T(t)R^{-1}(t)\big)R(t)\big(K^T(t) - R^{-1}(t)C^T(t)P(t)\big) + P(t)C^T(t)R^{-1}(t)C(t)P(t) .\end{aligned} \tag{41}$$

Setting $\dot{P}(t)$ *equal to the zero matrix results in a stationary point at (35) which leads to (40). From the differential of (40)*

$$\ddot{P}(t) = \big(A(t) - P(t)C^T(t)R^{-1}(t)C(t)\big)\dot{P}(t) + \dot{P}(t)\big(A^T(t) - P(t)C^T(t)R^{-1}(t)C(t)\big) \tag{42}$$

and it can be seen that $\ddot{P}(t)$ *≥ 0 provided that the assumptions (37) - (38) hold. Therefore, P(t) =* $E\{\tilde{x}(t|t)\tilde{x}^T(t|t)\}$ *is minimised at (35).* □

The above development is somewhat brief and not very rigorous. Further discussions appear in [4] – [17]. It is tendered to show that the Kalman filter minimises the error covariance, provided of course that the problem assumptions are correct. In the case that it is desired to estimate an arbitrary linear combination $C_1(t)$ of states, the optimal filter is given by the system

$$\dot{\hat{x}}(t|t) = A(t)\hat{x}(t|t) + K(t)\big(z(t) - C(t)\hat{x}(t|t)\big) , \tag{43}$$

$$\hat{y}_1(t) = C_1(t)\hat{x}(t) . \tag{44}$$

"The worst wheel of the cart makes the most noise." *Benjamin Franklin*

This filter minimises the error covariance $C_1(t)P(t)C_1^T(t)$. The generalisation of the Kalman filter for problems possessing deterministic inputs, correlated noises, and a direct feed-through term is developed below.

3.3.3 Including Deterministic Inputs

Suppose that the signal model is described by

$$\dot{x}(t) = A(t)x(t) + B(t)w(t) + \mu(t) \tag{45}$$

$$y(t) = C(t)x(t) + \pi(t), \tag{46}$$

where $\mu(t)$ and $\pi(t)$ are deterministic (or known) inputs. In this case, the filtered state estimate can be obtained by including the deterministic inputs as follows

$$\dot{\hat{x}}(t\,|\,t) = A(t)\hat{x}(t\,|\,t) + K(t)\big(z(t) - C(t)\hat{x}(t\,|\,t) - \pi(t)\big) + \mu(t) \tag{47}$$

$$\hat{y}(t) = C(t)\hat{x}(t) + \pi(t). \tag{48}$$

It is easily verified that subtracting (47) from (45) yields the error system (39) and therefore, the Kalman filter's differential Riccati equation remains unchanged.

Example 2. Suppose that an object is falling under the influence of a gravitational field and it is desired to estimate its position over [0, t] from noisy measurements. Denote the object's vertical position, velocity and acceleration by $x(t)$, $\dot{x}(t)$ and $\ddot{x}(t)$, respectively. Let g denote the gravitational constant. Then $\ddot{x}(t) = -g$ implies $\dot{x}(t) = \dot{x}(0) - gt$, so the model may be written as

$$\begin{gathered}\begin{bmatrix}\dot{x}(t)\\ \ddot{x}(t)\end{bmatrix} = A\begin{bmatrix}x(t)\\ \dot{x}(t)\end{bmatrix} + \mu(t),\\ z(t) = C\begin{bmatrix}x(t)\\ \dot{x}(t)\end{bmatrix} + v(t),\end{gathered} \tag{49}$$

where $A = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$ is the state matrix, $\mu(t) = \begin{bmatrix}\dot{x}(0) - gt\\ -g\end{bmatrix}$ is a deterministic input and $C = \begin{bmatrix}1 & 0\end{bmatrix}$ is the output mapping. Thus, the Kalman filter has the form

$$\begin{bmatrix}\dot{\hat{x}}(t\,|\,t)\\ \ddot{\hat{x}}(t\,|\,t)\end{bmatrix} = A\begin{bmatrix}\hat{x}(t\,|\,t)\\ \dot{\hat{x}}(t\,|\,t)\end{bmatrix} + K\left(z(t) - C\begin{bmatrix}\hat{x}(t\,|\,t)\\ \dot{\hat{x}}(t\,|\,t)\end{bmatrix}\right) + \mu(t), \tag{50}$$

$$\hat{y}(t\,|\,t) = C(t)\hat{x}(t\,|\,t), \tag{51}$$

where the gain K is calculated from (35) and (36), in which $BQB^T = 0$.

"These, Gentlemen, are the opinions upon which I base my facts." *Winston Leonard Spencer-Churchill*

3.3.4 Including Correlated Process and Measurement Noise

Suppose that the process and measurement noises are correlated, that is,

$$E\left\{\begin{bmatrix} w(t) \\ v(t) \end{bmatrix}\begin{bmatrix} w^T(\tau) & v^T(\tau) \end{bmatrix}\right\} = \begin{bmatrix} Q(t) & S(t) \\ S^T(t) & R(t) \end{bmatrix}\delta(t-\tau) . \tag{52}$$

The equation for calculating the optimal state estimate remains of the form (34), however, the differential Riccati equation and hence the filter gain are different. The generalisation of the optimal filter that takes into account (52) was published by Kalman in 1963 [5]. Kalman's approach was to first work out the corresponding discrete-time Riccati equation and then derive the continuous-time version.

The correlated noises can be accommodated by defining the signal model equivalently as

$$\dot{x}(t) = \bar{A}(t)x(t) + B(t)\bar{w}(t) + \mu(t) , \tag{53}$$

where

$$\bar{A}(t) = A(t) - B(t)S(t)R^{-1}(t)C(t) \tag{54}$$

is a new state matrix,

$$\bar{w}(t) = w(t) - S(t)R^{-1}(t)v(t) \tag{55}$$

is a new stochastic input that is uncorrelated with $v(t)$, and

$$\mu(t) = B(t)S(t)R^{-1}(t)y(t) \tag{56}$$

is a deterministic signal. It can easily be verified that the system (53) with the parameters (54) – (56), has the structure (26) with $E\{w(t)v^T(\tau)\} = 0$. It is convenient to define

$$\begin{aligned} \bar{Q}(t)\delta(t-\tau) &= E\{\bar{w}(t)\bar{w}^T(\tau)\} \\ &= E\{w(t)w^T(\tau)\} - E\{w(t)v^T(\tau)R^{-1}(t)S^T(t) - S(t)R^{-1}(t)E\{v(t)w^T(\tau) \\ &\quad + S(t)R^{-1}(t)E\{v(t)v^T(\tau)\}R^{-1}(t)S^T(t) \\ &= \left(Q(t) - S(t)R^{-1}(t)S^T(t)\right)\delta(t-\tau) . \end{aligned} \tag{57}$$

"I am tired of all this thing called science here. We have spent millions in that sort of thing for the last few years, and it is time it should be stopped." *Simon Cameron*

The corresponding Riccati differential equation is obtained by substituting $\overline{A}(t)$ for $A(t)$ and $\overline{Q}(t)$ for $Q(t)$ within (36), namely,

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

This can be rearranged to give

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - K(t)(t)R^{-1}(t)K^T(t) + B(t)Q(t)B^T(t) , \tag{59}$$

in which the gain is now calculated as

$$K(t) = \left(P(t)C^T(t) + B(t)S(t)\right)R^{-1}(t) . \tag{60}$$

3.3.5 Including a Direct Feedthrough Matrix

The approach of the previous section can be used to address signal models that possess a direct feedthrough matrix, namely,

$$\dot{x}(t) = A(t)x(t) + B(t)w(t) , \tag{61}$$

$$y(t) = C(t)x(t) + D(t)w(t) . \tag{62}$$

As before, the optimal state estimate is given by

$$\dot{\hat{x}}(t \mid t) = A(t)\hat{x}(t \mid t) + K(t)\left(z(t) - C(t)\hat{x}(t \mid t)\right) , \tag{63}$$

where the gain is obtained by substituting $S(t) = Q(t)D^T(t)$ into (60),

$$K(t) = \left(P(t)C^T(t) + B(t)Q(t)D^T(t)\right)R^{-1}(t) , \tag{64}$$

in which $P(t)$ is the solution of the Riccati differential equation

$$\dot{P}(t) = \left(A(t) - B(t)Q(t)D^T(t)R^{-1}(t)C(t)\right)P(t)$$
$$+ P(t)\left(A(t) - B(t)Q(t)D^T(t)R^{-1}(t)C(t)\right)^T + B(t)\left(Q(t) - Q(t)D(t)R^{-1}(t)D^T(t)Q(t)\right)B^T(t) .$$

Note that the above Riccati equation simplifies to

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - K(t)(t)R^{-1}(t)K^T(t) + B(t)Q(t)B^T(t) . \tag{65}$$

"No human investigation can be called real science if it cannot be demonstrated mathematically." *Leonardo di ser Piero da Vinci*

3.4 The Continuous-time Steady-State Minimum-Variance Filter

3.4.1 Riccati Differential Equation Monotonicity

This section sets out the simplifications for the case where the signal model is stationary (or time-invariant). In this situation the structure of the Kalman filter is unchanged but the gain is fixed and can be pre-calculated. Consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bw(t) , \tag{66}$$

$$y(t) = Cx(t) , \tag{67}$$

together with the observations

$$z(t) = y(t) + v(t) , \tag{68}$$

assuming that Re{$\lambda_i(A)$} < 0, $E\{w(t)\} = 0$, $E\{w(t)w^T(\tau)\} = Q(t)\delta(t - \tau)$, $E\{v(t)\} = 0$, $E\{v(t)v^T(\tau)\} = R$ and $E\{w(t)v^T(\tau)\} = 0$. It follows from the approach of Section 3 that the Riccati differential equation for the corresponding Kalman filter is given by

$$\dot{P}(t) = AP(t) + P(t)A^T - P(t)C^T R^{-1}CP(t) + BQB^T . \tag{69}$$

It will be shown that the solution for $P(t)$ monotonically approaches a steady-state asymptote, in which case the filter gain can be calculated before running the filter. The following result is required to establish that the solutions of the above Riccati differential equation are monotonic.

Lemma 5 [11], [19], [20]: *Suppose that X(t) is a solution of the Lyapunov differential equation*

$$\dot{X}(t) = AX(t) + X(t)A^T \tag{70}$$

over an interval $t \in [0, T]$. Then the existence of a solution $X(t_0) \geq 0$ implies $X(t) \geq 0$ for all $t \in [0, T]$.

Proof: *Denote the transition matrix of $\dot{x}(t)$ = - $A(t)x(t)$ by $\Phi^T(t,\tau)$, for which $\dot{\Phi}(t,\tau)$ = $-A^T(t)\Phi(t,\tau)$ and $\dot{\Phi}^T(t,\tau)$ = $-\Phi^T(t,\tau)A(t)$. Let $P(t)$ = $\Phi^T(t,\tau)X(t)\Phi(t,\tau)$, then from (70)*

$$\begin{aligned}0 &= \Phi^T(t,\tau)\left(\dot{X}(t) - AX(t) - X(t)A^T\right)\Phi(t,\tau)\\ &= \dot{\Phi}^T(t,\tau)X(t)\Phi(t,\tau) + \Phi^T(t,\tau)\dot{X}(t)\Phi(t,\tau) + \Phi^T(t,\tau)X(t)\dot{\Phi}(t,\tau)\\ &= \dot{P}(t) .\end{aligned}$$

Therefore, a solution $X(t_0) \geq 0$ of (70) implies that $X(t) \geq 0$ for all $t \in [0, T]$. □

The monotonicity of Riccati differential equations has been studied by Bucy [6], Wonham [23], Poubelle *et al* [19] and Freiling [20]. The latter's simple proof is employed below.

"Today's scientists have substituted mathematics for experiments, and they wander off through equation after equation, and eventually build a structure which has no relation to reality." *Nikola Tesla*

***Lemma 6 [19], [20]:** Suppose for a $t \ge 0$ and a $\delta_t > 0$ there exist solutions $P(t) \ge 0$ and $P(t + \delta_t) \ge 0$ of the Riccati differential equations*

$$\dot{P}(t) = AP(t) + P(t)A^T - P(t)C^T R^{-1} CP(t) + BQB^T \tag{71}$$

and

$$\dot{P}(t+\delta_t) = AP(t+\delta_t) + P(t+\delta_t)A^T - P(t+\delta_t)C^T R^{-1} CP(t+\delta_t) + BQB^T , \tag{72}$$

respectively, such that $P(t) - P(t + \delta_t) \ge 0$. Then the sequence of matrices $P(t)$ is monotonic nonincreasing, that is,

$$P(t) - P(t + \delta_t) \ge 0, \text{ for all } t \ge \delta_t. \tag{73}$$

***Proof:** The conditions of the Lemma are the initial step of an induction argument. For the induction step, denote $\dot{P}(\delta_t) = \dot{P}(t) - \dot{P}(t+\delta_t)$, $P(\delta_t) = P(t) - P(t+\delta_t)$ and $\bar{A} = AP(t_t)C^T R^{-1} C - 0.5P(\delta_t)$. Then*

$$\begin{aligned}\dot{P}(\delta_t) &= AP(\delta_t) + P(\delta_t)A^T - P(t+\delta_t)C^T R^{-1} CP(t+\delta_t) + P(t)C^T R^{-1} CP(t) \\ &= \bar{A}P(\delta_t) + P(\delta_t)\bar{A}^T ,\end{aligned}$$

which is of the form (70), and so the result (73) follows. □

A monotonic nondecreasing case can be established similarly – see [20].

3.4.2 Observability

The continuous-time system (66) – (67) is termed completely observable if the initial states, $x(t_0)$, can be uniquely determined from the inputs and outputs, $w(t)$ and $y(t)$, respectively, over an interval [0, T]. A simple test for observability is is given by the following lemma.

***Lemma 7 [10], [21].** Suppose that $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{p\times n}$. The system is observable if and only if the observability matrix $O \in \mathbb{R}^{np\times n}$ is of rank n, where*

$$O = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}. \tag{74}$$

"You can observe a lot by just watching." *Lawrence Peter (Yogi) Berra*

Proof: *Recall from Chapter 2 that the solution of (66) is*

$$x(t) = e^{At}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}Bw(\tau)d\tau \ . \tag{75}$$

Since the input signal w(t) within (66) is known, it suffices to consider the unforced system $\dot{x}(t) = Ax(t)$ and $y(t) = Cx(t)$, *that is, Bw(t) = 0, which leads to*

$$y(t) = Ce^{At}x(t_0) \ . \tag{76}$$

The exponential matrix is defined as

$$\begin{aligned} e^{At} &= I + At + \frac{A^2t^2}{2} + \cdots + \frac{A^Nt^N}{N!} \\ &= \sum_{k=0}^{N-1}\alpha_k(t)A^k \ , \end{aligned} \tag{77}$$

where $\alpha_k(t) = t^k/k!$. *Substituting (77) into (76) gives*

$$\begin{aligned} y(t) &= \sum_{k=0}^{N-1}\alpha_k(t)CA^k x(t_0) \\ &= \alpha_0(t)Cx(t_0) + \alpha_1(t)CAx(t_0) + \ldots + \alpha_{N-1}(t)CA^{N-1}x(t_0) \ . \\ &= \begin{bmatrix} \alpha_0(t) & \alpha_1(t) & \cdots & \alpha_{N-1}(t) \end{bmatrix} \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{N-1} \end{bmatrix} x(t_0) \ . \end{aligned} \tag{78}$$

From the Cayley-Hamilton Theorem [22],

$$rank\left(\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{N-1} \end{bmatrix}\right) = rank\left(\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}\right)$$

for all N ≥ n. Therefore, we can take N = n within (78). Thus, equation (78) uniquely determines $x(t_0)$ *if and only if O has full rank n.* □

A system that does not satisfy the above criterion is said to be unobservable. An alternate proof for the above lemma is provided in [10]. If a signal model is not observable then a Kalman filter cannot estimate all the states from the measurements.

“Who will observe the observers ?” *Arthur Stanley Eddington*

Example 3. The pair $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$ is expected to be unobservable because one of the two states appears as a system output whereas the other is hidden. By inspection, the rank of the observability matrix, $\begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, is 1. Suppose instead that $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, namely measurements of both states are available. Since the observability matrix $\begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ is of rank 2, the pair (A, C) is observable, that is, the states can be uniquely reconstructed from the measurements.

3.4.3 The Algebraic Riccati Equation

Some pertinent facts concerning the Riccati differential equation (69) are:

- Its solutions correspond to the covariance of the state estimation error.
- From Lemma 6, if it is suitably initialised then its solutions will be monotonically nonincreasing.
- If the pair (A, C) is observable then the states can be uniquely determined from the outputs.

In view of the above, it is not surprising that if the states can be estimated uniquely, in the limit as t approaches infinity, the Riccati differential equation will have a unique steady state solution.

Lemma 8 [20], [23], [24]: *Suppose that $Re\{\lambda_i(A)\} < 0$, the pair (A, C) is observable, then the solution of the Riccati differential equation (69) satisfies*

$$\lim_{t \to \infty} P(t) = P \, , \tag{79}$$

where P is the solution of the algebraic Riccati equation

$$0 = AP + PA^T - PC^T R^{-1} CP + BQB^T \, . \tag{80}$$

A proof that the solution P is in fact unique appears in [24]. A standard way for calculating solutions to (80) arises by finding an appropriate set of Schur vectors for the Hamiltonian matrix $H = \begin{bmatrix} A & -C^T R^{-1} C \\ BQB^T & -A^T \end{bmatrix}$, see [25] and the Hamiltonian solver within *Matlab*™.

"Stand firm in your refusal to remain conscious during algebra. In real life, I assure you, there is no such thing as algebra." *Francis Ann Lebowitz*

t	$P(t)$	$\dot{P}(t)$
1	0.9800	−2.00
10	0.8316	−1.41
100	0.4419	$-8.13*10^{-2}$
1000	0.4121	$-4.86*10^{-13}$

Table 1. Solutions of (69) for Example 4.

Example 4. Suppose that $A = -1$ and $B = C = Q = R = 1$, for which the solution of the algebraic Riccati equation (80) is $P = 0.4121$. Using Euler's integration method (see Chapter 1) with $\delta_t = 0.01$ and $P(0) = 1$, the calculated solutions of the Riccati differential equation (69) are listed in Table 1. The data in the table demonstrate that the Riccati differential equation solution converges to the algebraic Riccati equation solution and $\lim_{t\to\infty} \dot{P}(t) = 0$.

The so-called infinite-horizon (or stationary) Kalman filter is obtained by substituting time-invariant state-space parameters into (34) - (35) to give

$$\dot{\hat{x}}(t \mid t) = (A - KC)\hat{x}(t \mid t) + Kz(t), \tag{81}$$

$$\hat{y}(t \mid t) = C\hat{x}(t \mid t), \tag{82}$$

where

$$K = PC^T R^{-1}, \tag{83}$$

in which P is calculated by solving the algebraic Riccati equation (80). The output estimation filter (81) – (82) has the transfer function

$$H_{OE}(s) = C(sI - A + KC)^{-1} K. \tag{84}$$

Example 5. Suppose that a signal $y(t) \in \mathbb{R}$ is generated by the system

$$y(t) = \left(\frac{b_m \frac{d^m}{dt^m} + b_{m-1} \frac{d^{m-1}}{dt^{n-1}} + \ldots + b_1 \frac{d}{dt} + b_0}{a_n \frac{d^n}{dt^n} + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} + \ldots + a_1 \frac{d}{dt} + a_0} \right) w(t).$$

This system's transfer function is

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0},$$

which can be realised in the controllable canonical form [10]

"If you think dogs can't count, try putting three dog biscuits in your pocket and then giving Fido two of them." *Phil Pastoret*

$$A = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \dots & -a_1 & -a_0 \\ 1 & 0 & & \dots & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \text{ and } C = \begin{bmatrix} b_m & b_{m-1} & \dots & b_1 & b_0 \end{bmatrix}.$$

The optimal filter for estimating $y(t)$ from noisy measurements (29) is obtained by using the above state-space parameters within (81) – (83). It has the structure depicted in Figs. 3 and 4. These figures illustrate two features of interest. First, the filter's model matches that within the signal generating process. Second, designing the filter is tantamount to finding an optimal gain.

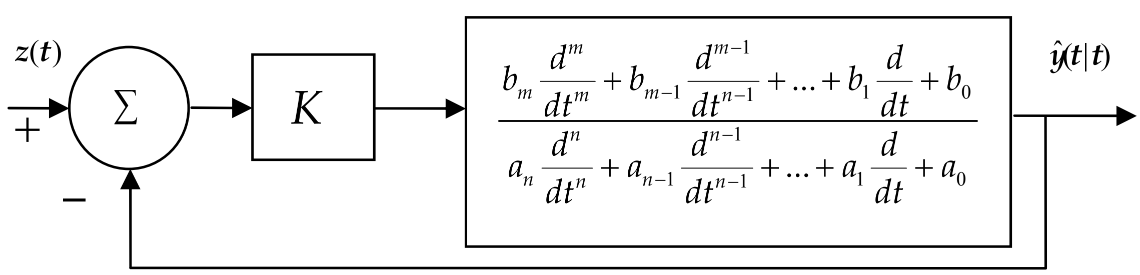


Figure 4. The optimal filter for Example 5.

3.4.4 Equivalence of the Wiener and Kalman Filters

When the model parameters and noise statistics are time-invariant, the Kalman filter reverts to the Wiener filter. The equivalence of the Wiener and Kalman filters implies that spectral factorisation is the same as solving a Riccati equation. This observation is known as the Kalman-Yakubovich-Popov Lemma (or Positive Real Lemma) [15], [26], which assumes familiarity with the following Schur complement formula.

For any matrices Φ_{11}, Φ_{12} and Φ_{22}, where Φ_{11} and Φ_{22} are symmetric, the following are equivalent.

(i) $\begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^T & \Phi_{22} \end{bmatrix} \ge 0$.

(ii) $\Phi_{11} \ge 0$, $\Phi_{22} \ge \Phi_{12}^T \Phi_{11}^{-1} \Phi_{12}$.

(iii) $\Phi_{22} \ge 0$, $\Phi_{11} \ge \Phi_{12} \Phi_{22}^{-1} \Phi_{12}^T$.

"Mathematics is the queen of sciences and arithmetic is the queen of mathematics." *Carl Friedrich Gauss*

The Kalman-Yakubovich-Popov Lemma is set out below. Further details appear in [15] and a historical perspective is provided in [26]. A proof of this Lemma makes use of the identity

$$-PA^T - AP = P(-sI - A^T) + (sI - A)P\,. \tag{85}$$

Lemma 9 [15], [26]: *Consider the spectral density matrix*

$$\Delta\Delta^H(s) = \begin{bmatrix} C(sI - A)^{-1} & I \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} (-sI - A^T)^{-1}C^T \\ I \end{bmatrix}. \tag{86}$$

Then the following statements are equivalent:

(i) $\Delta\Delta^H(j\omega) \geq 0$ *for all* $\omega \in (-\infty,\infty)$.

(ii) $\begin{bmatrix} BQB^T + AP + PA^T & PC^T \\ CP & R \end{bmatrix} \geq 0\,.$

(iii) *There exists a nonnegative solution P of the algebraic Riccati equation (80).*

Proof: *To establish equivalence between (i) and (iii), use (85) within (80) to obtain*

$$P(-sI - A^T) + (sI - A)P = BQB^T - PC^TRCP\,. \tag{87}$$

Premultiplying and postmultiplying (87) by $C(sI - A)^{-1}$ *and* $(-sI - A^T)^{-1}C^T$, *respectively, results in*

$$C(sI - A)^{-1}PC^T + CP(-sI - A^T)^{-1}C^T = C(sI - A)^{-1}(BQB^T - PC^TRCP)(-sI - A^T)^{-1}C^T\,. \tag{88}$$

Hence,

$$\begin{aligned} \Delta\Delta(s) &= GQG(s) + R \\ &= C(sI - A)^{-1}BQB^T(-sI - A^T)^{-1}C^T + R \\ &= C(sI - A)^{-1}PC^TRCP(-sI - A^T)^{-1}C^T + C(sI - A)^{-1}PC^T + CP(-sI - A^T)^{-1}C^T + R \\ &= \left(C(sI - A)^{-1}KR^{1/2} + R^{1/2}\right)\left(R^{1/2}K^T(-sI - A^T)^{-1}C^T + R^{1/2}\right) \\ &\geq 0\,. \end{aligned} \tag{89}$$

The Schur complement formula can be used to verify the equivalence of (ii) and (iii). □

In Chapter 1, it is shown that the transfer function matrix of the optimal Wiener solution for output estimation is given by

$$H_{OE}(s) = I - R^{1/2}\Delta^{-1}(s)\,, \tag{90}$$

where $s = j\omega$ and

"Arithmetic is being able to count up to twenty without taking off your shoes." *Mickey Mouse*

$$\Delta\Delta^{H}(s) = GQG^{H}(s) + R\ . \tag{91}$$

is the spectral density matrix of the measurements. It follows from (91) that

$$\Delta(s) = C(sI - A)^{-1}KR^{1/2} + R^{1/2}\ . \tag{92}$$

The Wiener filter (90) requires the spectral factor inverse, $\Delta^{-1}(s)$, which can be found from (92) and using $[I + C(sI - A)^{-1}K]^{-1} = I + C(sI - A + KC)^{-1}K$ to obtain

$$\Delta^{-1}(s) = R^{-1/2} - R^{-1/2}C(sI - A + KC)^{-1}K\ . \tag{93}$$

Substituting (93) into (90) yields

$$H_{OE}(s) = C(sI - A + KC)^{-1}K\ , \tag{94}$$

which is identical to the minimum-variance output estimator (84).

Example 5. Consider a scalar output estimation problem where $G(s) = (s + 1)^{-1}$, $Q = 1$, $R = 0.0001$ and the Wiener filter transfer function is

$$H(s) = 99(s + 100)^{-1}\ . \tag{95}$$

Applying the bilinear transform yields $A = -1$, $B = C = 1$, for which the solution of (80) is $P = 0.0099$. By substituting $K = PC^{T}R^{-1} = 99$ into (90), one obtains (95).

3.5 Conclusion

The Kalman-Bucy filter which produces state estimates $\hat{x}(t\,|\,t)$ and output estimates $\hat{y}(t\,|\,t)$ from the measurements $z(t) = y(t) + v(t)$ at time t is summarised in Table 2. This filter minimises the variances of the state estimation error, $E\{(x(t) - \hat{x}(t\,|\,t))(x(t) - \hat{x}(t\,|\,t))^{T}\} = P(t)$ and the output estimation error, $E\{(y(t) - \hat{y}(t\,|\,t))(y(t) - \hat{y}(t\,|\,t))^{T}\} = C(t)P(t)C^{T}(t)$.

When the model parameters and noise covariances are time-invariant, the gain is also time-invariant and can be precalculated. The time-invariant filtering results are summarised in Table 3. In this stationary case, spectral factorisation is equivalent to solving a Riccati equation and the transfer function of the output estimation filter, $H_{OE}(s) = C(sI - A + KC)^{-1}K$, is identical to that of the Wiener filter. It is not surprising that the Wiener and Kalman filters are equivalent since they are both derived by completing the square of the error covariance.

"Mathematics consists in proving the most obvious thing in the least obvious way." *George Polya*

	ASSUMPTIONS	MAIN RESULTS
Signals and system	$E\{w(t)\} = E\{v(t)\} = 0$. $E\{w(t)w^T(t)\} = Q(t)$ and $E\{v(t)v^T(t)\} = R(t)$ are known. $A(t)$, $B(t)$ and $C(t)$ are known.	$\dot{x}(t) = A(t)x(t) + B(t)w(t)$ $y(t) = C(t)x(t)$ $z(t) = y(t) + v(t)$
Filtered state and output factorisation		$\dot{\hat{x}}(t \mid t) = A(t)\hat{x}(t \mid t) + K(t)(z(t) - C(t)\hat{x}(t \mid t))$ $\hat{y}(t \mid t) = C(t)\hat{x}(t \mid t)$
Filter gain and Riccati differential equation	$Q(t) > 0$ and $R(t) > 0$.	$K(t) = P(t)C(t)R^{-1}(t)$ $\dot{P}(t) = A(t)P(t) + P(t)A^T(t)$ $-P(t)C^T(t)R^{-1}(t)C(t)P(t) + B(t)Q(t)B^T(t)$

Table 2. Main results for time-varying output estimation.

	ASSUMPTIONS	MAIN RESULTS
Signals and system	$E\{w(t)\} = E\{v(t)\} = 0$. $E\{w(t)w^T(t)\} = Q$ and $E\{v(t)v^T(t)\} = R$ are known. A, B and C are known. The pair (A, C) is observable.	$\dot{x}(t) = Ax(t) + Bw(t)$ $y(t) = Cx(t)$ $z(t) = y(t) + v(t)$
Filtered state and output factorisation		$\dot{\hat{x}}(t \mid t) = A\hat{x}(t \mid t) + K(z(t) - C\hat{x}(t \mid t))$ $\hat{y}(t \mid t) = C\hat{x}(t \mid t)$
Filter gain and algebraic Riccati equation	$Q > 0$ and $R > 0$.	$K = PCR^{-1}$ $0 = AP + PA^T - PC^TR^{-1}CP + BQB^T$

Table 3. Main results for time-invariant output estimation.

"There are two ways to do great mathematics. The first is to be smarter than everybody else. The second way is to be stupider than everybody else - but persistent." *Raoul Bott*

3.6 Problems

Problem 1. Show that $\dot{x}(t) = A(t)x(t)$ has the solution $x(t) = \Phi(t,0)x(0)$ where $\dot{\Phi}(t,0) = A(t)\Phi(t,0)$ and $\Phi(t,t) = I$. Hint: use the approach of [13] and integrate both sides of $\dot{x}(t) = A(t)x(t)$.

Problem 2. Given that:

(i) the Lyapunov differential equation for the system $\dot{x}(t) = F(t)x(t) + G(t)w(t)$ is $\frac{d}{dt}E\{x(t)x^T(t)\} = A(t)E\{x(t)x^T(t)\} + E\{x(t)x^T(t)\}F^T(t) + G(t)Q(t)G^T(t)$;

(ii) the Kalman filter for the system $\dot{x}(t) = A(t)x(t) + B(t)w(t)$, $z(t) = C(t)x(t) + v(t)$ has the structure $\dot{\hat{x}}(t|t) = A(t)\hat{x}(t|t) + K(t)(z(t) - C(t)\hat{x}(t|t))$;

write a Riccati differential equation for the evolution of the state error covariance and determine the optimal gain matrix $K(t)$.

Problem 3. Derive the Riccati differential equation for the model $\dot{x}(t) = A(t)x(t) + B(t)w(t)$, $z(t) = C(t)x(t) + v(t)$ with $E\{w(t)w^T(\tau)\} = Q(t)\delta(t - \tau)$, $E\{v(t)v^T(\tau)\} = R(t)\delta(t - \tau)$ and $E\{w(t)v^T(\tau)\} = S(t)\delta(t - \tau)$. Hint: consider $\dot{x}(t) = A(t)x(t) + B(t)w(t) + B(t)S(t)R^{-1}(t)(z(t) - C(t)x(t) - v(t))$.

Problem 4. For output estimation problems with $B = C = R = 1$, calculate the algebraic Riccati equation solution, filter gain and transfer function for the following.

(a)	$A = -1$ and $Q = 8$.	(b)	$A = -2$ and $Q = 12$.
(c)	$A = -3$ and $Q = 16$.	(d)	$A = -4$ and $Q = 20$.
(e)	$A = -5$ and $Q = 24$.	(f)	$A = -6$ and $Q = 28$.
(g)	$A = -7$ and $Q = 32$.	(h)	$A = -8$ and $Q = 36$.
(i)	$A = -9$ and $Q = 40$.	(j)	$A = -10$ and $Q = 44$.

Problem 5. Prove the Kalman-Yakubovich-Popov Lemma for the case of

$E\left\{\begin{bmatrix} w(t) \\ v(t) \end{bmatrix}\begin{bmatrix} w^T(\tau) & v^T(\tau) \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\delta(t-\tau)$, *i.e.*, show

$$\Delta\Delta^H(s) = \begin{bmatrix} C(sI - A)^{-1} & I \end{bmatrix}\begin{bmatrix} Q & S \\ S & R \end{bmatrix}\begin{bmatrix} (-sI - A^T)^{-1}C^T \\ I \end{bmatrix}.$$

Problem 6. Derive a state space formulation for minimum-mean-square-error equaliser using $\Delta^{-1}(s) = R^{-1/2} - R^{-1/2}C(sI - A + KC)^{-1}K$.

"Mathematics is a game played according to certain simple rules with meaningless marks on paper." *David Hilbert*

3.7 Glossary

In addition to the terms listed in Section 1.6, the following have been used herein.

$\mathcal{G}: \mathbb{R}^p \rightarrow \mathbb{R}^q$	A linear system that operates on a p-element input signal and produces a q-element output signal.
$A(t)$, $B(t)$, $C(t)$, $D(t)$	Time-varying state space matrices of appropriate dimension. The system $\mathcal{G}$ is assumed to have the realisation $\dot{x}(t) = A(t)x(t) + B(t)w(t)$, $y(t) = C(t)x(t) + D(t)w(t)$.
$Q(t)$ and $R(t)$	Covariance matrices of the nonstationary stochastic signals $w(t)$and $v(t)$, respectively.
$\Phi(t,0)$	State transition matrix which satisfies $\dot{\Phi}(t,0) = \dfrac{d\Phi(t,0)}{dt} = A(t)\Phi(t,0)$ with the boundary condition $\Phi(t,t) = I$.
$\mathcal{G}^H$	Adjoint of $\mathcal{G}$. The adjoint of a system having the state-space parameters $\{A(t), B(t), C(t), D(t)\}$ is a system parameterised by $\{-A^T(t), -C^T(t), B^T(t), D^T(t)\}$.
$E\{.\}$, $E\{x(t)\}$	Expectation operator, expected value of $x(t)$.
$E\{x(t) \mid y(t)\}$	Conditional expectation, namely the estimate of $x(t)$ given $y(t)$.
$\hat{x}(t \mid t)$	Conditional mean estimate of the state $x(t)$ given data at time t.
$\tilde{x}(t \mid t)$	State estimation error which is defined by $\tilde{x}(t \mid t) = x(t) - \hat{x}(t \mid t)$.
$K(t)$	Time-varying filter gain matrix.
$P(t)$	Time-varying error covariance, *i.e.*, $E\{\tilde{x}(t)\tilde{x}^T(t)\}$, which is the solution of a Riccati differential equation.
A, B, C, D	Time-invariant state space matrices of appropriate dimension.
Q and R	Time-invariant covariance matrices of the stationary stochastic signals $w(t)$ and $v(t)$, respectively.
O	Observability matrix.
SNR	Signal to noise ratio.
K	Time-invariant filter gain matrix.
P	Time-invariant error covariance which is the solution of an algebraic Riccati equation.
H	Hamiltonian matrix.
$G(s)$	Transfer function matrix of the signal model.
$H(s)$	Transfer function matrix of the minimum-variance solution.
$H_{OE}(s)$	Transfer function matrix of the minimum-variance solution specialised for output estimation.

"A mathematician is a device for turning coffee into theorems." *Paul Erdos*

3.8 References

[1] R. W. Bass, "Some reminiscences of control theory and system theory in the period 1955 – 1960: Introduction of Dr. Rudolf E. Kalman, *Real Time*, Spring/Summer Issue, The University of Alabama in Huntsville, 2002.

[2] M. S. Grewal and A. P. Andrews, "Applications of Kalman Filtering in Areospace 1960 to the Present", *IEEE Control Systems Magazine*, vol. 30, no. 3, pp. 69 – 78, June 2010.

[3] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME, Series D, Journal of Basic Engineering*, vol 82, pp. 35 – 45, 1960.

[4] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory", *Transactions of the ASME, Series D, Journal of Basic Engineering*, vol 83, pp. 95 – 107, 1961.

[5] R. E. Kalman, "New Methods in Wiener Filtering Theory", *Proc. First Symposium on Engineering Applications of Random Function Theory and Probability*, Wiley, New York, pp. 270 – 388, 1963.

[6] R. S. Bucy, "Global Theory of the Riccati Equation", *Journal of Computer and System Sciences*, vol. 1, pp. 349 – 361, 1967.

[7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, Inc., New York, 1970.

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

[9] A. Gelb, *Applied Optimal Estimation*, The Analytic Sciences Corporation, USA, 1974.

[10] T. Kailath, *Linear Systems*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1980.

[11] H. W. Knobloch and H. K. Kwakernaak, *Lineare Kontrolltheorie*, Springer-Verlag, Berlin, 1980.

[12] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, Inc., USA, 1983.

[13] P. A. Ruymgaart and T. T. Soong, *Mathematics of Kalman-Bucy Filtering*, Second Edition, Springer-Verlag, Berlin, 1988.

[14] M. S. Grewal and A. P. Andrews, *Kalman Filtering, Theory and Practice*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1993.

[15] T. Kailath, A. H. Sayed and B. Hassibi, *Linear Estimation*, Prentice-Hall, Upper Saddle River, New Jersey, 2000.

[16] D. Simon, *Optimal State Estimation, Kalman H_∞ and Nonlinear Approaches*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[17] 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.

[18] T. Söderström, *Discrete-time Stochastic Systems: estimation and control*, Springer-Verlag London Ltd., 2002.

[19] M. – A. Poubelle, R. R. Bitmead and M. R. Gevers, "Fake Algebraic Riccati Techniques and Stability", *IEEE Transactions on Automatic Control*, vol. 33, no. 4, pp. 379 – 381, 1988.

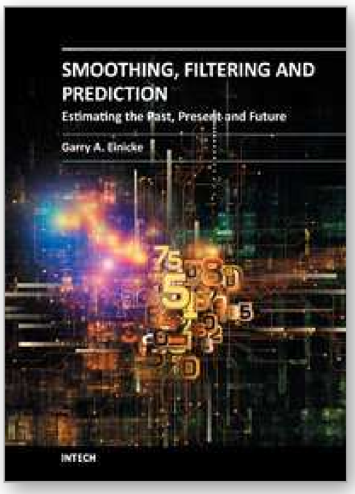
[20] G. Freiling, V. Ionescu, H. Abou-Kandil and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*, Birkhauser, Boston, 2003.

[21] K. Ogata, *Matlab for Control Engineers*, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2008.

"But mathematics is the sister, as well as the servant, of the arts and is touched with the same madness and genius." *Harold Marston Morse*

[22] T. Kaczorek, "Cayley-Hamilton Theorem" in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, 2001.
[23] W. M. Wonham, "On a Matrix Riccati Equation of Stochastic Control", *SIAM Journal on Control*, vol. 6, no. 4, pp. 681 – 697, 1968.
[24] M. – A. Poubelle, I. R. Petersen, M. R. Gevers and R. R. Bitmead, "A Miscellany of Results on an Equation of Count J. F. Riccati", *IEEE Transactions on Automatic Control*, vol. 31, no. 7, pp. 651 – 654, 1986.
[25] A. J. Laub, "A Schur Method for Solving Algebraic Riccati Equations", *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 913 – 921, 1979.
[26] S. V. Gusev and A. L. Likhtarnikov, "Kalman-Popov-Yakubovich Lemma and the S-procedure: A Historical Essay", *Automation and Remote Control*, vol. 67, no. 11, pp. 1768 – 1810, 2006.

"Mathematics are like Frenchmen: whatever you say to them they translate into their own language, and forthwith it is something entirely different." *Johann Wolfgang von Goethe*



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.

InTech Europe

University Campus STeP Ri

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai

www.intechopen.com

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