10-15-2013

The Kalman filter for linear systems on time scales

Martin Bohner
Missouri University of Science and Technology

Nick Wintz
Lindenwood University, nwintz@lindenwood.edu

J. Math. Anal. Appl. 406 (2013) 419–436

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

The Kalman filter for linear systems on time scales

Martin Bohner [a], Nick Wintz [b,*]

[a] Missouri University of Science and Technology, Department of Mathematics and Statistics, 1870 Miner Circle, Rolla, MO 65409-0020, United States

[b] Lindenwood University, Department of Mathematics, 209 S. Kingshighway, St. Charles, MO 63301, United States

ARTICLE INFO

Article history:
Received 9 September 2012
Available online 9 May 2013
Submitted by Hans Zwart

Keywords:
Kalman filter
Time scale
Dynamic equation
Optimal estimation
Mean square error
Riccati equation

ABSTRACT

We introduce the Kalman filter for linear systems on time scales, which includes the discrete and continuous versions as special cases. When the system is also stochastic, we show that the Kalman filter is an observer that estimates the system when the state is corrupted by noisy measurements. Finally, we show that the duality of the Kalman filter and the Linear Quadratic Regulator (LQR) is preserved in their unification on time scales. A numerical example is provided.

1. Introduction

In our previous work [5–7], we considered a linear dynamic system where the given state was incomplete or had not been accurately measured. We then introduced a more desirable state which theoretically contained more information of our process. Taking the difference between these known states gave us a new state equation influenced by a known disturbance. The goal then was to determine an optimal control that not only minimized a cost functional, but this state difference (the error) as well. Such problems are referred to as disturbance/rejection models (see [12,16,6]). Now we consider a natural extension, where this disturbance is unknown.

In this paper, we consider the linear stochastic system

$$\begin{aligned} x^{\Delta}(t) &= Ax(t) + Bu(t) + Gw(t), \qquad x(t_0) = x_0, \\ y(t) &= Cx(t) + v(t), \end{aligned}$$

where $x \in \mathbb{R}^n$ represents the state, $u \in \mathbb{R}^m$ is a known input, $y \in \mathbb{R}^p$ represents the measurement, $w \in \mathbb{R}^l$ is the process noise, and $v \in \mathbb{R}^p$ is the measurement noise. In such a case, it is necessary to determine a filter that not only removes the noise but retains relevant information. Here we seek an estimate that can not only accurately estimate our unknown true state, but also ensures that the mean square error is as small as possible. This leads to finding an estimate $\hat{x}$ such that $\hat{x}$ satisfies the observer equation

$$\hat{x}^{\Delta}(t) = A\hat{x}(t) + Bu(t) + K(t)[y(t) - C\hat{x}(t)], \qquad \hat{x}(t_0) = \bar{x}(t_0),$$

where $\bar{x}$ is the expected value of our true state and K represents the Kalman gain.

The Kalman filter in the discrete case is usually attributed to Kalman in 1960 (see [13]), however this is subject to debate (see [18,20,21]). In this setting, the Kalman filter is essentially a predictor–corrector type estimator. Initially, there is a "time

* Corresponding author.
E-mail addresses: bohner@mst.edu (M. Bohner), nwintz@lindenwood.edu (N. Wintz).

http://dx.doi.org/10.1016/j.jmaa.2013.04.075

Table 1
The "predictive" form of the discrete Kalman filter.

System: $x(t+1) = Ax(t) + Bu(t) + Gw(t),\ x(t_0) = x_0$ Measurement: $y(t) = Cx(t) + v(t)$
Assumptions: $x_0 \sim (\overline{x}_0, P_0),\ w \sim (0, Q\delta(t-s)),\ v \sim (0, R\delta(t-s))$, which are mutually uncorrelated, $R > 0$
Initialization Initial estimate: $\hat{x}(t_0) = \overline{x}_0$ Error covariance: $P(t_0) = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] = P_0$
Estimate update: $\hat{x}(t+1\|t) = A\hat{x}(t\|t-1) + Bu(t) + K(t)[y(t) - C\hat{x}(t\|t-1)]$
Kalman gain: $K(t) = AP(t)C^T(R + CP(t)C^T)^{-1}$
Error covariance update: $P(t+1) = AP(t)A^T - AP(t)C^T(R + CP(t)C^T)^{-1}CP(t)A^T + GQG^T$

Table 2
The continuous Kalman–Bucy filter.

System: $\dot{x}(t) = Ax(t) + Bu(t) + Gw(t),\ x(t_0) = x_0$ Measurement: $y(t) = Cx(t) + v(t)$
Assumptions: $x_0 \sim (\overline{x}_0, P_0),\ w \sim (0, Q\delta(t-s)),\ v \sim (0, R\delta(t-s))$, which are mutually uncorrelated, $R > 0$
Initialization Initial estimate: $\hat{x}(t_0) = \overline{x}_0$ Error covariance: $P(t_0) = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] = P_0$
Estimate update: $\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K(t)[y(t) - C\hat{x}(t)]$
Kalman gain: $K(t) = P(t)C^TR^{-1}$
Error covariance update: $\dot{P}(t) = AP(t) + P(t)A^T - P(t)C^TR^{-1}CP(t) + GQG^T$

update" in which the algorithm predicts a state estimate based on a previous measurement. This prediction is then associated with an error covariance. Next there is a "measurement update" in which the algorithm calculates a correction of the state estimate based on the prediction and the new measurement along with its associated error covariance. The algorithm then repeats itself. While the discrete Kalman filter is usually written with separate time and measurement updates, this is not necessary to do so. In Table 1 we present the discrete case in the "predictive" form. Note that the error covariance is not in terms of either the measurement or the input (see [8,15]). As a result, it is possible that both the error covariance and the Kalman gain can be computed *a priori*.

In 1961, Kalman and Bucy created a corresponding filter in the continuous case [14]. It should be noted that the continuous (Kalman–Bucy) filter can be derived from the discrete filter. However, mathematically speaking, the filtering in continuous time is a more advanced problem than filtering in discrete time. Unlike the discrete case, the continuous Kalman filter cannot be decomposed into separate time and measurement updates. This is due to the fact that all of the error covariances in the discrete case tend to the same error covariance in limit. White noise also presents a problem as it forces the filter to be formulated and solved using Itô differentials and Brownian motion if certain precautions are not taken. Since we only consider the linear case, we can avoid this issue. The filter for the continuous case is found in Table 2.

Despite their differences, both filters look strikingly similar. When seeking to unify and extend the Kalman filter to dynamic equations, the following assumptions have been made.

Assumption 1.1. Throughout the paper, we assume the following.

a. The true state and state estimate belong to the same time scale.
b. The state and measurements are Gaussian.
c. The state measurement is being updated in "real-time". In other words, there is a measurement at the next available point in the time scale.
d. The error covariance of our hybrid filter is found through the integrator just as it is for the Kalman–Bucy filter.
e. There exists a term $\delta(\cdot,\cdot)$ such that

$$\int_{t_0}^{t_f}\int_{t_0}^{t_f} x^T(\tau_1)Q\delta(\tau_1,\tau_2)x(\tau_2)\Delta\tau_1\Delta\tau_2 = \int_{t_0}^{t_f} x^T(\tau)Qx(\tau)\Delta\tau. \tag{1}$$

Note that (1) is valid for isolated time scales as well as the reals with δ being the usual Dirac delta. Table 3 summarizes our results.

Table 3
The dynamic Kalman filter.

System: $x^{\Delta}(t) = Ax(t) + Bu(t) + Gw(t),\ x(t_0) = x_0$ Measurement: $y(t) = Cx(t) + v(t)$
Assumptions: $x_0 \sim (\overline{x}_0, P_0)$, $w \sim (0, Q\delta(t,s))$, $v \sim (0, R\delta(t,s))$, which are mutually uncorrelated, $R > 0$
Initialization Initial estimate: $\hat{x}(t_0) = \overline{x}_0$ Error covariance: $P(t_0) = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] = P_0$
Estimate update: $\hat{x}^{\Delta}(t) = A\hat{x}(t) + Bu(t) + K(t)[y(t) - C\hat{x}(t)]$
Kalman gain: $K(t) = (I + \mu(t)A)P(t)C^T(R + \mu(t)CP(t)C^T)^{-1}$
Error covariance update: $P^{\Delta}(t) = AP(t) + (I + \mu(t)A)P(t)A^T + GQG^T - (I + \mu(t)A)P(t)C^T(R + \mu(t)CP(t)C^T)^{-1}CP(t)(I + \mu(t)A^T)$

Table 4
Examples of time scales.

$\mathbb{T}$	$\mu(t)$	$\sigma(t)$
$\mathbb{R}$	0	t
$\mathbb{Z}$	1	$t+1$
$h\mathbb{Z}$	h	$t+h$
$\overline{q^{\mathbb{Z}}}$	$(q-1)t$	qt
$\mathbb{N}_0^2$	$1 + 2\sqrt{t}$	$\left(\sqrt{t}+1\right)^2$
$\mathbb{T} = \{H_n\}$	$\frac{1}{n+1}$	H_{n+1}
$\mathbb{P}_{a,b}$	$\begin{cases} 0, & t \in \bigcup_{k=0}^{\infty}[kc, kc+a) \\ b, & t \in \bigcup_{k=0}^{\infty}\{kc+a\}, \end{cases}$ where $c = a+b$	$\begin{cases} t, & t \in \bigcup_{k=0}^{\infty}[kc, kc+a) \\ t+b, & t \in \bigcup_{k=0}^{\infty}\{kc+a\} \end{cases}$

Over the years, the Kalman filter has proven to be a useful mathematical tool mainly for the simplicity of its design. While the Kalman filter is most famous for its role in putting a man on the moon, it has numerous other useful applications in engineering and analysis of economic systems (see [1,2]) as well. There has also been interest in comparing discrete with continuous measurements when the filter design is given as a continuous process [11,19,23]. Applications here include biomechanical models, particularly for cardiac kinetics estimation. Such filters are sometimes called "hybrid" filters, although this term is generic. Despite its various incarnations, each filter design looks remarkably like Kalman's original filter.

The paper is organized as follows. In Section 2, we offer some basic properties of calculus on time scales. We also review our results for the linear quadratic regulator (LQR) [5] when the final state is free. In Section 3, we extend and unify the Kalman filter to dynamic equations on time scales. Finally, in Section 4, we examine the relationship between the LQR and the Kalman filter. This work is from the second author's dissertation [22].

2. Preliminaries

In this section, we provide a brief introduction to the theory of dynamic equations on time scales. For a more in-depth study of time scales, see Bohner and Peterson's books [3,4].

Definition 2.1. A *time scale* $\mathbb{T}$ is an arbitrary nonempty closed subset of the real numbers. We let $\mathbb{T}^{\kappa} = \mathbb{T} \setminus \{\max \mathbb{T}\}$ if $\max \mathbb{T}$ exists; otherwise $\mathbb{T}^{\kappa} = \mathbb{T}$.

Example 2.2. Some examples of time scales include $\mathbb{T} = \mathbb{R}, \mathbb{T} = \mathbb{Z}, \mathbb{T} = h\mathbb{Z}$ for $h > 0$, the quantum numbers $\mathbb{T} = \overline{q^{\mathbb{Z}}} = \{q^k : k \in \mathbb{Z}\} \cup \{0\}$ for $q > 1$, $\mathbb{T} = \mathbb{P}_{a,b} = \bigcup_{k=0}^{\infty}[k(a+b), k(a+b)+a]$, for $a, b > 0$, the so-called harmonic numbers $\mathbb{T} = \{H_n : n \in \mathbb{N}_0\}$ where $H_0 = 0$ and $H_n = \sum_{k=1}^{n} \frac{1}{k}$, and the Cantor set.

Definition 2.3. We define the *forward jump operator* $\sigma : \mathbb{T} \to \mathbb{T}$ and the *graininess function* $\mu : \mathbb{T} \to [0, \infty)$ by

$$\sigma(t) := \inf\{s \in \mathbb{T} : s > t\} \quad \text{and} \quad \mu(t) := \sigma(t) - t.$$

For any function $f : \mathbb{T} \to \mathbb{R}$, we define the function $f^{\sigma} : \mathbb{T} \to \mathbb{R}$ by $f^{\sigma} = f \circ \sigma$.

In Table 4, we give the forward jump operators and the graininess functions for some common time scales.

Next, we define the delta (or Hilger) derivative as follows.

Definition 2.4. Assume $f : \mathbb{T} \to \mathbb{R}$ and let $t \in \mathbb{T}^\kappa$. The *delta derivative* $f^\Delta(t)$ is the number (when it exists) such that given any $\varepsilon > 0$, there is a neighborhood U of t such that

$$\left|[f(\sigma(t)) - f(s)] - f^\Delta(t)[\sigma(t) - s]\right| \le \varepsilon|\sigma(t) - s| \quad \text{for all } s \in U.$$

In the next two theorems, we consider some properties of the delta derivative.

Theorem 2.5 (*See [3, Theorem* 1.16*]*)**.** *Let $f : \mathbb{T} \to \mathbb{R}$ be a function and $t \in \mathbb{T}^\kappa$. Then we have the following.*

a. *If f is differentiable at t, then f is continuous at t.*
b. *If f is continuous at t, where t is right-scattered, then f is differentiable at t and*

$$f^\Delta(t) = \frac{f(\sigma(t)) - f(t)}{\mu(t)}.$$

c. *If f is differentiable at t, where t is right-dense, then*

$$f^\Delta(t) = \lim_{s\to t} \frac{f(t) - f(s)}{t - s}.$$

d. *If f is differentiable at t, then*

$$f(\sigma(t)) = f(t) + \mu(t)f^\Delta(t). \tag{2}$$

Note that (2) is sometimes called the "simple useful formula".

Example 2.6. Note the following examples.

a. When $\mathbb{T} = \mathbb{R}$, then (if the limit exists)

$$f^\Delta(t) = \lim_{s\to t} \frac{f(t) - f(s)}{t - s} = f'(t).$$

b. When $\mathbb{T} = \mathbb{Z}$, then

$$f^\Delta(t) = f(t+1) - f(t) =: \Delta f(t).$$

c. When $\mathbb{T} = h\mathbb{Z}$ for $h > 0$, then

$$f^\Delta(t) = \frac{f(t+h) - f(t)}{h} =: \Delta_h f(t).$$

d. When $\mathbb{T} = q^{\mathbb{Z}}$ for $q > 1$, then

$$f^\Delta(t) = \frac{f(qt) - f(t)}{(q-1)t} =: D_q f(t).$$

Next we consider the linearity property as well as the product rules.

Theorem 2.7 (*See [3, Theorem* 1.20*]*)**.** *Let $f, g : \mathbb{T} \to \mathbb{R}$ be differentiable at $t \in \mathbb{T}^\kappa$. Then we have the following.*

a. *For any constants α and β, the sum $(\alpha f + \beta g) : \mathbb{T} \to \mathbb{R}$ is differentiable at t with*

$$(\alpha f + \beta g)^\Delta(t) = \alpha f^\Delta(t) + \beta g^\Delta(t).$$

b. *The product $fg : \mathbb{T} \to \mathbb{R}$ is differentiable at t with*

$$(fg)^\Delta(t) = f^\Delta(t)g(t) + f^\sigma(t)g^\Delta(t) = f(t)g^\Delta(t) + f^\Delta(t)g^\sigma(t).$$

Definition 2.8. A function $f : \mathbb{T} \to \mathbb{R}$ is said to be *rd-continuous* on $\mathbb{T}$ when f is continuous in points $t \in \mathbb{T}$ with $\sigma(t) = t$ and it has finite left-sided limits in points $t \in \mathbb{T}$ with $\sup\{s \in \mathbb{T} : s < t\} = t$. The class of rd-continuous functions $f : \mathbb{T} \to \mathbb{R}$ is denoted by $\mathrm{C_{rd}} = \mathrm{C_{rd}}(\mathbb{T}) = \mathrm{C_{rd}}(\mathbb{T}, \mathbb{R})$. The set of functions $f : \mathbb{T} \to \mathbb{R}$ that are differentiable and whose derivative is rd-continuous is denoted by $\mathrm{C^1_{rd}}$.

Theorem 2.9 (*See [3, Theorem* 1.74*]*)**.** *Any rd-continuous function $f : \mathbb{T} \to \mathbb{R}$ has an antiderivative F, i.e., $F^\Delta = f$ on $\mathbb{T}^\kappa$.*

Definition 2.10. Let $f \in \mathrm{C_{rd}}$ and let F be any function such that $F^\Delta(t) = f(t)$ for all $t \in \mathbb{T}^\kappa$. Then the Cauchy integral of f is defined by

$$\int_a^b f(t)\Delta t = F(b) - F(a) \quad \text{for all } a, b \in \mathbb{T}.$$

Example 2.11. Let $a, b \in \mathbb{T}$ with $a < b$ and assume that $f \in C_{rd}$.

a. When $\mathbb{T} = \mathbb{R}$, then
$$\int_a^b f(t)\Delta t = \int_a^b f(t)\mathrm{d}t.$$
b. When $\mathbb{T} = \mathbb{Z}$, then
$$\int_a^b f(t)\Delta t = \sum_{t=a}^{b-1} f(t).$$
c. When $\mathbb{T} = h\mathbb{Z}$ for $h > 0$, then
$$\int_a^b f(t)\Delta t = h\sum_{t=a/h}^{b/h-1} f(th).$$
d. When $\mathbb{T} = q^{\mathbb{N}_0}$ for $q > 1$, then
$$\int_a^b f(t)\Delta t = \int_a^b f(t)\mathrm{d}_q(t) := (q-1)\sum_{t\in[a,b)\cap\mathbb{T}} tf(t).$$

Definition 2.12. An $m \times n$ matrix-valued function A on $\mathbb{T}$ is rd-continuous if each of its entries are rd-continuous. Furthermore, if $m = n$, A is said to be *regressive* (we write $A \in \mathcal{R}$) if

$$I + \mu(t)A(t) \quad \text{is invertible for all } t \in \mathbb{T}^{\kappa},$$

where I is the identity matrix.

Theorem 2.13 (*See [3, Theorem* 5.8*]*)**.** *Suppose that A is regressive and rd-continuous. Then the initial value problem*

$$X^{\Delta}(t) = A(t)X(t), \qquad X(t_0) = I$$

has a unique $n \times n$ matrix-valued solution X.

Next, we present the matrix exponential and some of its properties.

Definition 2.14. The solution X from Theorem 2.13 is called the matrix exponential function on $\mathbb{T}$ and is denoted by $e_A(\cdot, t_0)$.

Theorem 2.15 (*See [3, Theorem* 5.21*]*)**.** *Let A be regressive and rd-continuous. Then for $r, s, t \in \mathbb{T}$,*

a. $e_A(t,s)e_A(s,r) = e_A(t,r)$,
b. $e_A(\sigma(t),s) = (I + \mu(t)A(t))e_A(t,s)$,
c. $e_A(t,\sigma(s)) = e_A(t,s)(I + \mu(s)A)^{-1}$,
d. $(e_A(\cdot,s))^{\Delta} = Ae_A(\cdot,s)$,
e. $(e_A(t,\cdot))^{\Delta} = -e_A^{\sigma}(t,\cdot)A = e_A(t,\cdot)(I + \mu(s)A)^{-1}A$.

Next we give the solution to our linear system using variation of parameters.

Theorem 2.16 (*See [3, Theorem* 5.24*]*)**.** *Let $A \in \mathcal{R}$ be an $n \times n$ matrix-valued function on $\mathbb{T}$ and suppose that $f : \mathbb{T} \to \mathbb{R}^n$ is rd-continuous. Let $t_0 \in \mathbb{T}$ and $x_0 \in \mathbb{R}^n$. Then the solution of the initial value problem*

$$x^{\Delta}(t) = A(t)x(t) + f(t), \qquad x(t_0) = x_0$$

is given by

$$x(t) = e_A(t,t_0)x_0 + \int_{t_0}^{t} e_A(t,\sigma(\tau))f(\tau)\Delta\tau.$$

Here, we review our results regarding the linear quadratic regulator generalized to dynamic equations on time scales [5]. We considered the linear system

$$\begin{aligned} x^{\Delta}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \end{aligned} \tag{3}$$

associated with the quadratic performance index

$$J = \frac{1}{2}x^T(t_f)Fx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}(x^TQx + u^TRu)(\tau)\Delta\tau, \tag{4}$$

where $F, Q \geq 0$ and $R > 0$. Then a necessary condition for a minimum cost is that x, λ, u satisfy the state, costate, and stationary equations

$$x^{\Delta} = Ax - BR^{-1}B^T\lambda^{\sigma}, \tag{5a}$$

$$-\lambda^{\Delta} = Qx + A^T\lambda^{\sigma}, \tag{5b}$$

$$u = -R^{-1}B^T\lambda^{\sigma}. \tag{5c}$$

In the free final state case, we make the assumption that x and λ satisfy the linear relationship

$$\lambda(t) = S(t)x(t), \tag{6}$$

where S represents a solution to a Riccati equation

$$-S^{\Delta} = A^TS^{\sigma} + (I + \mu A^T)S^{\sigma}A - (I + \mu A^T)S^{\sigma}B(R + \mu B^TS^{\sigma}B)^{-1}B^TS^{\sigma}(I + \mu A) + Q \tag{7}$$

with $S(t_f) = F$. This condition (6) is called a "sweep condition", a term used by Bryson and Ho in [10].

Next, we offer the form of an optimal control in the free final state case.

Theorem 2.17 (*See [5, Theorem 6.10]*). *Let $R + \mu B^TS^{\sigma}B$ be invertible. Suppose that x, u, and λ solve (5) such that (6) holds. Then*

$$u(t) = -L(t)x(t),$$

where the matrix-valued function L represents the state feedback gain

$$L(t) = (R + \mu(t)B^TS^{\sigma}(t)B)^{-1}B^TS^{\sigma}(t)(I + \mu(t)A). \tag{8}$$

It was assumed in this paper that the state can be accurately measured. However, this is not always realistic. In the next section, we construct a system that estimates the state, an observer. This concept was first introduced in the discrete and continuous cases by Luenberger [17]. This leads us to the generalization of the Kalman filter on time scales. As such, the Kalman filter can be thought of as an observer in which the state is reconstructed from noisy measurements.

3. The Kalman filter

In this section, we provide additional background and assumptions to unify and extend the Kalman filter to time scales. Since the goal of the filter is to minimize the mean square error, it is also referred to as the Linear Quadratic Estimator (LQE).

Assumption 3.1. Throughout this paper, we consider the linear stochastic system

$$\begin{aligned} x^{\Delta}(t) &= Ax(t) + Bu(t) + Gw(t), \qquad x(t_0) = x_0, \\ y(t) &= Cx(t) + v(t), \end{aligned} \tag{9}$$

where the following assumptions have been made.

a. The state $x \in \mathbb{R}^n$ is a nonstationary random variable with mean $\bar{x}$ and covariance $P_x = \mathbb{E}[(x - \bar{x})(x - \bar{x})^T]$.
b. The input $u \in \mathbb{R}^m$ is deterministic.
c. The output $y \in \mathbb{R}^p$ is a nonstationary random variable with mean $\bar{y}$ and covariance $P_y = \mathbb{E}[(y - \bar{y})(y - \bar{y})^T]$.
d. The process noise $w \in \mathbb{R}^l$ is stationary white noise with mean 0 and covariance $\mathbb{E}[w(t)w^T(s)] = Q\delta(t, s)$.
e. The measurement noise $v \in \mathbb{R}^p$ is stationary white noise with mean 0 and covariance $\mathbb{E}[v(t)v^T(s)] = R\delta(t, s)$.
f. x_0, w, and v are assumed to be mutually uncorrelated.
g. $P(t_0) = P_0$, Q, and R are all positive definite.
h. We can interchange the expectation and integration operations, i.e.,

$$\mathbb{E}\left[\int Z(\tau)\Delta\tau\right] = \int \mathbb{E}[Z(\tau)]\Delta\tau.$$

Remark 3.2. Note that Assumption 3.1h is used throughout the paper to bring the expectation inside the integral (e.g., in the proofs of Corollary 3.7, Theorems 3.8 and 3.13). This is with the understanding that the integral $\int \mathbb{E}[Z(\tau)]\Delta\tau$ exists on the given time scale. This assumption is valid for the reals and isolated time scales. Since no research has been done to establish conditions that ensure this equality holds for a general time scale, we have made this assumption as a precaution.

For convenience, we consider the time-invariant case. However, our results are also valid for the time-varying case.

3.1. Propagation of means and variances

Definition 3.3. A random vector is said to be *stationary* if all of its statistical properties do not vary with time. Processes whose statistical properties do change are referred to as *nonstationary*.

The type of stationary processes we consider is white noise, which we define as follows.

Definition 3.4. A random vector v is said to be a white noise random vector if and only if

a. $\mathbb{E}(v(t)) = 0$;
b. $\mathbb{E}(v(t)v^T(s)) = R\delta(t, s)$.

Next, we consider the situation when two vectors are orthogonal to each other.

Definition 3.5. Two vector-valued functions $w, v : \mathbb{T} \to \mathbb{R}^m$ are said to be mutually uncorrelated if

$$\mathbb{E}[w(t)v^T(s)] = 0 \quad \text{for all } s, t \in \mathbb{T}.$$

Theorem 3.6 (*See [3, Theorem 5.24]*). *The solution of the initial value problem*

$$x^{\Delta}(t) = Ax(t) + Bu(t) + Gw(t), \qquad x(t_0) = x_0 \tag{10}$$

is given by

$$x(t) = e_A(t, t_0)x_0 + \int_{t_0}^{t} e_A(t, \sigma(\tau))Bu(\tau)\Delta\tau + \int_{t_0}^{t} e_A(t, \sigma(\tau))Gw(\tau)\Delta\tau.$$

Proof. The result is due to Theorem 2.16. □

Corollary 3.7. *The mean of the solution for (10) is given by*

$$\bar{x}(t) = e_A(t, t_0)\bar{x}_0 + \int_{t_0}^{t} e_A(t, \sigma(\tau))Bu(\tau)\Delta\tau, \tag{11}$$

and the difference between the solution of (10) and its mean is given by

$$x(t) - \bar{x}(t) = e_A(t, t_0)(x_0 - \bar{x}_0) + \int_{t_0}^{t} e_A(t, \sigma(\tau))Gw(\tau)\Delta\tau. \tag{12}$$

Proof. First, using Theorem 3.6 and Assumption 3.1a, b, d, and h, we have

$$\begin{aligned}
\mathbb{E}[x(t)] &= \mathbb{E}\left[e_A(t, t_0)x_0 + \int_{t_0}^{t} e_A(t, \sigma(\tau))Bu(\tau)\Delta\tau + \int_{t_0}^{t} e_A(t, \sigma(\tau))Gw(\tau)\Delta\tau\right] \\
&= e_A(t, t_0)\bar{x}_0 + \int_{t_0}^{t} \mathbb{E}[e_A(t, \sigma(\tau))Bu(\tau)]\,\Delta\tau + \int_{t_0}^{t} \mathbb{E}[e_A(t, \sigma(\tau))Gw(\tau)]\,\Delta\tau \\
&= e_A(t, t_0)\bar{x}_0 + \int_{t_0}^{t} e_A(t, \sigma(\tau))Bu(\tau)\Delta\tau.
\end{aligned}$$

Then taking the difference between the solution of (10) and (11), we have (12) as desired. □

Theorem 3.8. *The covariance of the solution of (10) is given by*

$$P(t) = e_A(t, t_0)\left[P_0 + \int_{t_0}^{t} e_A(t_0, \sigma(\tau))GQG^T e_A^T(t_0, \sigma(\tau))\Delta\tau\right] e_A^T(t, t_0). \tag{13}$$

Proof. Using the definition of covariance, Corollary 3.7, and Assumption 3.1a, b, d, and h, we have

$$\begin{aligned}
P(t) &= \mathbb{E}[(x(t) - \bar{x}(t))(x(t) - \bar{x}(t))^T] \\
&= \mathbb{E}\left[e_A(t, t_0)(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T e_A^T(t, t_0)\right] + \mathbb{E}\left[e_A(t, t_0)(x_0 - \bar{x}_0)\int_{t_0}^{t} w^T(\tau)G^T e_A^T(t, \sigma(\tau))\Delta\tau\right] \\
&\quad + \mathbb{E}\left[\int_{t_0}^{t} e_A(t, \sigma(\tau))Gw(\tau)(x_0 - \bar{x}_0)^T e_A^T(t, t_0)\Delta\tau\right] \\
&\quad + \mathbb{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} e_A(t, \sigma(\tau_1))Gw(\tau_1)w^T(\tau_2)G^T e_A^T(t, \sigma(\tau_2))\Delta\tau_1\Delta\tau_2\right] \\
&= e_A(t, t_0)P_0 e_A^T(t, t_0) + \int_{t_0}^{t}\int_{t_0}^{t} e_A(t, \sigma(\tau_1))GQ\delta(\tau_2, \tau_1)G^T e_A^T(t, \sigma(\tau_2))\Delta\tau_1\Delta\tau_2 \\
&= e_A(t, t_0)\left[P_0 + \int_{t_0}^{t} e_A(t_0, \sigma(\tau))GQG^T e_A^T(t_0, \sigma(\tau))\Delta\tau\right] e_A^T(t, t_0).
\end{aligned}$$

This shows (13) as desired. □

Note that the mean and covariance of the state should actually be considered as conditional. Next we find the propagation of the state covariance matrix.

Corollary 3.9. *The propagation of the covariance of the solution of* (10) *is given by*

$$P^{\Delta}(t) = AP(t) + (I + \mu(t)A)P(t)A^T + GQG^T. \tag{14}$$

Proof. Using Theorems 2.7b and 2.15, we differentiate (13) to get

$$\begin{aligned}
P^{\Delta}(t) &= \left\{e_A(t,t_0)\left[P_0 + \int_{t_0}^{t} e_A(t_0,\sigma(\tau))GQG^T e_A^T(t_0,\sigma(\tau))\Delta\tau\right]\right\}^{\Delta} e_A^T(\sigma(t),t_0) \\
&\quad + e_A(t,t_0)\left[P_0 + \int_{t_0}^{t} e_A(t_0,\sigma(\tau))GQG^T e_A^T(t_0,\sigma(\tau))\Delta\tau\right](e_A^{\Delta}(t,t_0))^T \\
&= e_A(\sigma(t),t_0)e_A(t_0,\sigma(t))GQG^T e_A^T(t_0,\sigma(t))e_A^T(\sigma(t),t_0) \\
&\quad + Ae_A(t,t_0)\left[P_0 + \int_{t_0}^{t} e_A(t_0,\sigma(\tau))GQG^T e_A^T(t_0,\sigma(\tau))\Delta\tau\right] e_A^T(t,t_0)(I+\mu(t)A)^T \\
&\quad + e_A(t,t_0)\left[P_0 + \int_{t_0}^{t} e_A(t_0,\sigma(\tau))GQG^T e_A^T(t_0,\sigma(\tau))\Delta\tau\right] e_A^T(t,t_0)A^T \\
&= AP(t)(I+\mu(t)A)^T + P(t)A^T + GQG^T.
\end{aligned}$$

This shows (14) and concludes the proof. □

3.2. The linear quadratic estimator

While the initial state is not known, we do assume that the statistics are known. Here we assume that x_0 has mean $\bar{x}_0$ and covariance P_0, where P_0 is given by

$$P_0 = \mathbb{E}[(x(t_0) - \hat{x}(t_0))(x(t_0) - \hat{x}(t_0))^T].$$

The goal of the filter is to determine a state $\hat{x}$ such that our observer is unbiased. Therefore it is a natural assumption to pick the initial state estimate to be $\hat{x}(t_0) = \bar{x}(t_0)$. This is referred to as the initialization of the filter.

Definition 3.10. Let the linear stochastic system be given by (9). Then the state estimate, $\hat{x}$, to (9) is given by the observer equation

$$\hat{x}^{\Delta}(t) = A\hat{x}(t) + Bu(t) + K(t)[y(t) - C\hat{x}(t)], \qquad \hat{x}(t_0) = \bar{x}(t_0), \tag{15}$$

where K represents the Kalman gain.

We derive the form of this gain later on in this section. Now since $\hat{x}$ is the state estimate, we can call $\hat{y} = C\hat{x}$ the output estimate. Just as we desire $\hat{x}$ to be close to x, our observer is working properly if $y - \hat{y}$ is made to be small. This difference is sometimes called the residual or output estimation error. The observer (15) can be thought as the measurement update of our state equation. As mentioned before, we assume that this measurement is being updated in real-time. Thus, the Kalman gain can be thought of as a "blending" factor that fuses the residual with the state estimate. A block diagram of the filter design is found in Fig. 1.

Lemma 3.11. *If x satisfies* (9) *and $\hat{x}$ solves* (15), *the state error $\tilde{x} = x - \hat{x}$ satisfies*

$$\tilde{x}^{\Delta}(t) = M(t)\tilde{x}(t) + Gw(t) - K(t)v(t), \tag{16}$$

where $M(t) = A - K(t)C$.

Proof. Taking the difference of (9) and (15), we have

$$\begin{aligned}
\tilde{x}^{\Delta}(t) &= x^{\Delta}(t) - \hat{x}^{\Delta}(t) \\
&= (A - K(t)C)\tilde{x}(t) + Gw(t) - K(t)v(t) \\
&= M(t)\tilde{x}(t) + Gw(t) - K(t)v(t).
\end{aligned}$$

This gives (16) as desired. □

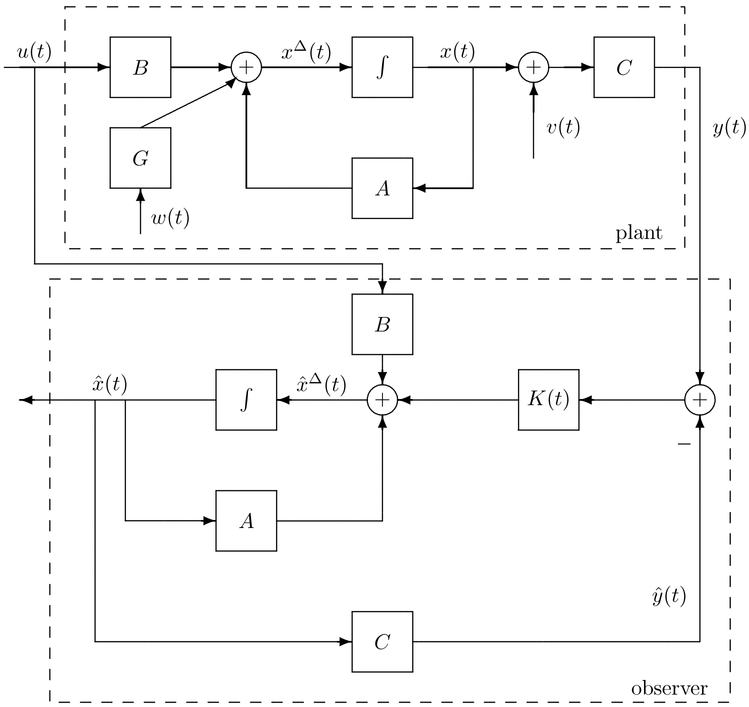


Fig. 1. Kalman filter design.

Remark 3.12. Note that by Theorem 2.16, the solution to (16) is of the form

$$\tilde{x}(t) = e_M(t, t_0)\tilde{x}_0 + \int_{t_0}^{t} e_M(t, \sigma(\tau))[Gw(\tau) - K(\tau)v(\tau)]\Delta\tau. \tag{17}$$

Next, we provide the error covariance.

Theorem 3.13. *The covariance of the solution of* (16) *is given by*

$$\begin{aligned} P(t) &= e_M(t, t_0)\left[P_0 + \int_{t_0}^{t} e_M(t_0, \sigma(\tau))GQG^T e_M^T(t_0, \sigma(\tau))\Delta\tau \right. \\ &\quad \left. + \int_{t_0}^{t} e_M(t_0, \sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t_0, \sigma(\tau))\Delta\tau\right] e_M^T(t, t_0). \end{aligned} \tag{18}$$

Proof. Expanding the definition of the error covariance and using Assumption 3.1a and d–h, we have

$$\begin{aligned} P(t) &= \mathbb{E}\left[\tilde{x}(t)\tilde{x}^T(t)\right] \\ &= \mathbb{E}\left[e_M(t, t_0)\tilde{x}_0\tilde{x}_0^T e_M^T(t, t_0)\right] + \mathbb{E}\left[e_M(t, t_0)\tilde{x}_0 \int_{t_0}^{t} [w^T(\tau)G^T - v^T(\tau)K^T(\tau)]e_M^T(t, \sigma(\tau))\Delta\tau\right] \\ &\quad + \mathbb{E}\left[\int_{t_0}^{t} e_M(t, \sigma(\tau))[Gw(\tau) - K(\tau)v(\tau)]\tilde{x}_0^T e_M^T(t, t_0)\Delta\tau\right] \\ &\quad + \mathbb{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} e_M(t, \sigma(\tau_1))[Gw - Kv](\tau_1)[w^T G^T - v^T K^T](\tau_2) \times e_M^T(t, \sigma(\tau_2))\Delta\tau_1\Delta\tau_2\right] \\ &= e_M(t, t_0)P_0 e_M^T(t, t_0) + \int_{t_0}^{t} e_M(t, \sigma(\tau))GQG^T e_M^T(t, \sigma(\tau))\Delta\tau + \int_{t_0}^{t} e_M(t, \sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t, \sigma(\tau))\Delta\tau \\ &= e_M(t, t_0)\left[P_0 + \int_{t_0}^{t} e_M(t_0, \sigma(\tau))GQG^T e_M^T(t_0, \sigma(\tau))\Delta\tau \right. \\ &\quad \left. + \int_{t_0}^{t} e_M(t_0, \sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t_0, \sigma(\tau))\Delta\tau\right] e_M^T(t, t_0). \end{aligned}$$

This shows (18) and concludes the proof. □

Now, we present the propagation of the error covariance.

Corollary 3.14. *Let P be given by (18). Then P satisfies*

$$P^{\Delta} = AP + (I + \mu A)PA^T + K[R + \mu CPC^T]K^T - K[CP + \mu CPA^T] - [\mu APC^T + PC^T]K^T + GQG^T. \tag{19}$$

Proof. Using Theorems 2.7b and 2.15, we differentiate the error covariance (18) to get

$$\begin{aligned}
P^{\Delta}(t) &= \Bigg\{ e_M(t,t_0)\Bigg[P_0 + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))GQG^T e_M^T(t_0,\sigma(\tau))\Delta\tau \\
&\quad + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t_0,\sigma(\tau))\Delta\tau\Bigg]\Bigg\}^{\Delta} e_M^T(\sigma(t),t_0) \\
&\quad + e_M(t,t_0)\Bigg[P_0 + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))GQG^T e_M^T(t_0,\sigma(\tau))\Delta\tau \\
&\quad + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t_0,\sigma(\tau))\Delta\tau\Bigg](e_M^{\Delta}(t,t_0))^T \\
&= e_M(\sigma(t),t_0)e_M(t_0,\sigma(t))[GQG^T + K(t)RK^T(t)]e_M^T(t_0,\sigma(t))e_M^T(\sigma(t),t_0) \\
&\quad + M(t)e_M(t,t_0)\Bigg[P_0 + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))GQG^T e_M^T(t_0,\sigma(\tau))\Delta\tau \\
&\quad + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t_0,\sigma(\tau))\Delta\tau\Bigg]e_M^T(t,t_0)(I + \mu(t)M(t))^T \\
&\quad + e_M(t,t_0)\Bigg[P_0 + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))GQG^T e_M^T(t_0,\sigma(\tau))\Delta\tau \\
&\quad + \int_{t_0}^{t} e_M(t_0,\sigma(\tau))K(\tau)RK^T(\tau)e_M^T(t_0,\sigma(\tau))\Delta\tau\Bigg]e_M^T(t,t_0)M^T(t) \\
&= M(t)P(t)(I + \mu(t)M(t))^T + P(t)M^T(t) + K(t)RK^T(t) + GQG^T \\
&= (A - K(t)C)P(t)(I + \mu(t)(A - K(t)C))^T + P(t)(A - K(t)C)^T + K(t)RK^T(t) + GQG^T \\
&= AP(t) + (I + \mu(t)A)P(t)A^T + K(t)[R + \mu(t)CP(t)C^T]K^T(t) \\
&\quad - K(t)CP(t)[I + \mu(t)A^T] - [I + \mu(t)A]P(t)C^TK^T(t) + GQG^T.
\end{aligned}$$

This gives us (19) as desired. □

Remark 3.15. In this remark, we find the "correct" form of the Kalman gain using a technique similar to the one found by Sorenson [20] and Brown [9] for the discrete case. First note that the term

$$K(t)CP(t)[I + \mu(t)A^T] + [I + \mu(t)A]P(t)C^TK^T(t)$$

is linear in $K(t)$ while $K(t)[R+\mu(t)CP(t)C^T]K^T(t)$ is quadratic in $K(t)$. Now assume that $R+\mu(t)CP(t)C^T$ is positive definite. Thus there exists an invertible matrix $D(t)$ such that

$$D(t)D^T(t) = R + \mu(t)CP(t)C^T.$$

Then (19) can be rewritten as

$$\begin{aligned}
P^{\Delta}(t) &= AP(t) + (I + \mu(t)A)P(t)A^T + K(t)D(t)D^T(t)K^T(t) \\
&\quad - K(t)CP(t)[I + \mu(t)A^T] - [I + \mu(t)A]P(t)C^TK^T(t) + GQG^T \\
&= AP(t) + (I + \mu(t)A)P(t)A^T + (K(t)D(t) - N(t))(K(t)D(t) - N(t))^T - N(t)N^T(t) + GQG^T
\end{aligned} \tag{20}$$

provided that

$$[KCP(I + \mu A^T)](t) + [(I + \mu A)PC^TK^T](t) = (KDN^T)(t) + (ND^TK^T)(t),$$

which holds if

$$N(t) = (I + \mu(t)A)P(t)C^T(D^T(t))^{-1}.$$

Then to minimize the diagonal terms of $P^{\Delta}(t)$, we want the middle terms of (20) to be zero. Setting

$$K(t)D(t) = N(t),$$

we get

$$\begin{aligned}K(t) &= N(t)D^{-1}(t)\\ &= (I+\mu(t)A)P(t)C^T(D^T(t))^{-1}D^{-1}(t)\\ &= (I+\mu(t)A)P(t)C^T(R+\mu(t)CP(t)C^T)^{-1}.\end{aligned}$$

Definition 3.16. Assume that $R+\mu CPC^T > 0$. Then we define the *Kalman gain* by

$$K(t) = (I+\mu(t)A)P(t)C^T(R+\mu(t)CP(t)C^T)^{-1}. \tag{21}$$

Now that our Kalman gain is in the form (21), it is possible to write our error propagation as a Riccati equation.

Theorem 3.17. *Assume that $R+\mu CPC^T > 0$ and define K by (21). Then P solves (19) if and only if*

$$P^\Delta = AP + (I+\mu A)PA^T - (I+\mu A)PC^T(R+\mu CPC^T)^{-1}CP(I+\mu A^T) + GQG^T. \tag{22}$$

Proof. Plugging (21) into (19), we get

$$\begin{aligned}P^\Delta(t) &= AP(t) + (I+\mu(t)A)P(t)A^T + (I+\mu(t)A)P(t)C^T(R+\mu(t)CP(t)C^T)^{-1}(R+\mu(t)CP(t)C^T)K^T(t)\\ &\quad - (I+\mu(t)A)P(t)C^T(R+\mu(t)CP(t)C^T)^{-1}CP(t)(I+\mu(t)A^T) - (I+\mu(t)A)P(t)C^TK^T(t) + GQG^T\\ &= AP(t) + (I+\mu(t)A)P(t)A^T - (I+\mu(t)A)P(t)C^T(R+\mu(t)CP(t)C^T)^{-1}CP(t)(I+\mu(t)A^T) + GQG^T.\end{aligned}$$

This gives us (22) as desired. □

Note that the term $-(I+\mu A)PC^T(R+\mu CPC^T)^{-1}CP(I+\mu A^T)$ represents the decrease in P due to measurement. We illustrate this in the following example.

Example 3.18. Assume that $C = 0$ such that there are no measurements. Then the propagation of the error covariance of the linear stochastic system

$$x^\Delta = Ax + Bu + Gw$$

is given by

$$P^\Delta = AP + (I+\mu A)PA^T + GQG^T.$$

Since there are no measurements, the observer equation becomes

$$\hat{x}^\Delta = A\hat{x} + Bu.$$

Therefore the estimator propagates according to the deterministic version of the system. Next define $W(t) = \mathbb{E}(x(t)x^T(t))$. Now assume that $u = 0$. Then the optimal estimate is given as $\hat{x} = \bar{x}$. Now it follows that

$$P = \mathbb{E}[(x-\bar{x})(x-\bar{x})^T] = W - \overline{xx^T},$$

such that

$$\begin{aligned}P^\Delta &= W^\Delta - \overline{x^\Delta x^T} - \overline{x^\sigma x^{\Delta T}}\\ &= W^\Delta - A\overline{xx^T} - \overline{(I+\mu A)x(Ax)^T}\\ &= W^\Delta - AW - (I+\mu A)WA^T.\end{aligned}$$

Now comparing equations, we have

$$W^\Delta = AW + (I+\mu A)WA^T + GQG^T. \tag{23}$$

Thus, with no measurement and under a deterministic input, W and P must satisfy the same Lyapunov equation.

Example 3.19. Consider the scalar stochastic system

$$\begin{aligned}&x^\Delta(t) = ax(t) + bu(t) + gw(t), \qquad x(t_0) = x_0,\\ &y(t) = cx(t) + v(t).\end{aligned}$$

The solution to the state equation is given by

$$x(t) = e_a(t,t_0)x_0 + \int_{t_0}^t e_a(t,\sigma(\tau))(bu+gw)(\tau)\Delta\tau.$$

Then the covariance is written as

$$\begin{aligned} p(t) &= \mathbb{E}[(x(t)-\bar{x}(t))^2] \\ &= \mathbb{E}\left[\left(e_a(t,t_0)(x_0-\bar{x}_0)+\int_{t_0}^{t} e_a(t,\sigma(\tau))gw(\tau)\Delta\tau\right)^2\right] \\ &= \mathbb{E}\left[(e_a(t,t_0)(x_0-\bar{x}_0))^2\right]+\mathbb{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} e_a(t,\sigma(\tau_1))g^2w(\tau_1)w(\tau_2)e_a(t,\sigma(\tau_2))\Delta\tau_1\Delta\tau_2\right] \\ &= e_a^2(t,t_0)p_0+e_a^2(t,t_0)\int_{t_0}^{t} e_a^2(t_0,\sigma(\tau))g^2q(\tau)\Delta\tau \\ &= e_{2\odot a}(t,t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot a}(t_0,\sigma(\tau))g^2q(\tau)\Delta\tau\right], \end{aligned}$$

where $2\odot a = 2a+\mu a^2$. As a result, the propagation of the covariance can be found to be

$$\begin{aligned} p^{\Delta}(t) &= e_{2\odot a}^{\Delta}(t,t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot a}(t_0,\sigma(\tau))g^2q(\tau)\Delta\tau\right]+e_{2\odot a}(\sigma(t),t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot a}(t_0,\sigma(\tau))g^2q(\tau)\Delta\tau\right]^{\Delta} \\ &= (2\odot a)(t)e_{2\odot a}(t,t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot a}(t_0,\sigma(\tau))g^2q(\tau)\Delta\tau\right]+e_{2\odot a}(\sigma(t),t_0)e_{2\odot a}(t_0,\sigma(\tau))g^2q \\ &= (2\odot a)(t)p(t)+g^2q. \end{aligned}$$

Next, we estimate our state equation with the observer

$$\hat{x}^{\Delta}(t) = a\hat{x}(t)+bu(t)+k(t)[cy(t)-c\hat{x}(t)], \qquad \hat{x}(t_0)=\bar{x}_0.$$

In turn, we have the error system

$$\begin{aligned} \tilde{x}^{\Delta}(t) &= (a-k(t)c)\tilde{x}(t)+gw-kv \\ &= m(t)\tilde{x}(t)+gw-kv. \end{aligned}$$

Then the error covariance can be written as

$$\begin{aligned} p(t) &= \mathbb{E}[\tilde{x}^2(t)] \\ &= \mathbb{E}\left[\left(e_m(t,t_0)\tilde{x}_0+\int_{t_0}^{t} e_m(t,\sigma(\tau))(gw-kv)(\tau)\Delta\tau\right)^2\right] \\ &= \mathbb{E}\left[(e_m(t,t_0)\tilde{x}_0)^2\right]+\mathbb{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} e_m(t,\sigma(\tau_1))g^2w(\tau_1)w(\tau_2)e_m(t,\sigma(\tau_2))\Delta\tau_1\Delta\tau_2\right] \\ &\quad +\mathbb{E}\left[\int_{t_0}^{t}\int_{t_0}^{t} e_m(t,\sigma(\tau_1))k^2v(\tau_1)v(\tau_2)e_m(t,\sigma(\tau_2))\Delta\tau_1\Delta\tau_2\right] \\ &= e_m^2(t,t_0)p_0+e_m^2(t,t_0)\int_{t_0}^{t} e_m^2(t_0,\sigma(\tau))(g^2q+k^2r)(\tau)\Delta\tau \\ &= e_{2\odot m}(t,t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot m}(t_0,\sigma(\tau))(g^2q+k^2r)(\tau)\Delta\tau\right]. \end{aligned}$$

It follows that the propagation of the error covariance is found to be

$$\begin{aligned} p^{\Delta}(t) &= e_{2\odot m}^{\Delta}(t,t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot m}(t_0,\sigma(\tau))(g^2q+k^2r)(\tau)\Delta\tau\right] \\ &\quad +e_{2\odot m}(\sigma(t),t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot m}(t_0,\sigma(\tau))(g^2q+k^2r)(\tau)\Delta\tau\right]^{\Delta} \\ &= (2\odot m)(t)e_{2\odot m}(t,t_0)\left[p_0+\int_{t_0}^{t} e_{2\odot m}(t_0,\sigma(\tau))(g^2q+k^2r)(\tau)\Delta\tau\right] \\ &\quad +e_{2\odot m}(\sigma(t),t_0)e_{2\odot m}(t_0,\sigma(\tau))(g^2q+k^2r)(t) \\ &= (2\odot(a-ck))(t)p(t)+g^2q+rk^2(t) \\ &= (2\odot a)(t)p(t)-2(1+\mu(t)a)cp(t)k(t)+(r+\mu(t)c^2p(t))k^2(t)+g^2q. \end{aligned}$$

Now completing the square, we have

$$\begin{aligned} p^{\Delta}(t) &= (2\odot a)(t)p(t) + g^2 q - \frac{(1+\mu(t)a)^2}{(r+\mu(t)c^2p(t))}(cp(t))^2 + (r+\mu(t)c^2p(t))k^2(t) \\ &\quad + (r+\mu(t)c^2p(t))\left[-\frac{2(1+\mu(t)a)cp(t)}{(r+\mu(t)c^2p(t))}k(t) + \left(\frac{1+\mu(t)a}{r+\mu(t)c^2p(t)}\right)^2 (cp(t))^2\right] \\ &= (2\odot a)(t)p(t) + g^2 q - \frac{(1+\mu(t)a)^2}{(r+\mu(t)c^2p(t))}(cp(t))^2 + \left[k(t) - \frac{(1+\mu(t)a)cp(t)}{(r+\mu(t)c^2p(t))}\right]^2. \end{aligned}$$

To zero out the last term, we set (see (20))

$$k(t) = \frac{(1+\mu(t)a)cp(t)}{(r+\mu(t)c^2p(t))}.$$

Finally, we can write the dynamic equation for p as a Riccati equation (see (22)) as

$$p^{\Delta}(t) = (2\odot a)(t)p(t) - \frac{(1+\mu(t)a)^2}{(r+\mu(t)c^2p(t))}(cp(t))^2 + g^2 q.$$

Next, we consider a simple numerical example.

Example 3.20. Consider the spring–mass system

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\Delta} &= \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w, \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v, \end{aligned}$$

where x_1 represents the position of an object, x_2 is its velocity, y is the measurement, $w \sim N(0, 1)$ is the process noise, and $v \sim N(0, 2)$ represents the measurement noise. The initial estimate for position is 1 ft below equilibrium position while the initial velocity is estimated to be 1 ft/s. In applying the Kalman filter, it is assumed that the measurements are Gaussian with variances $\sigma_{x_1}^2 = 2$ and $\sigma_{x_2}^2 = 3$. As a result, we can initialize our filter by assuming that

$$\begin{aligned} \hat{x}_0 &= \bar{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \\ P_0 &= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}. \end{aligned}$$

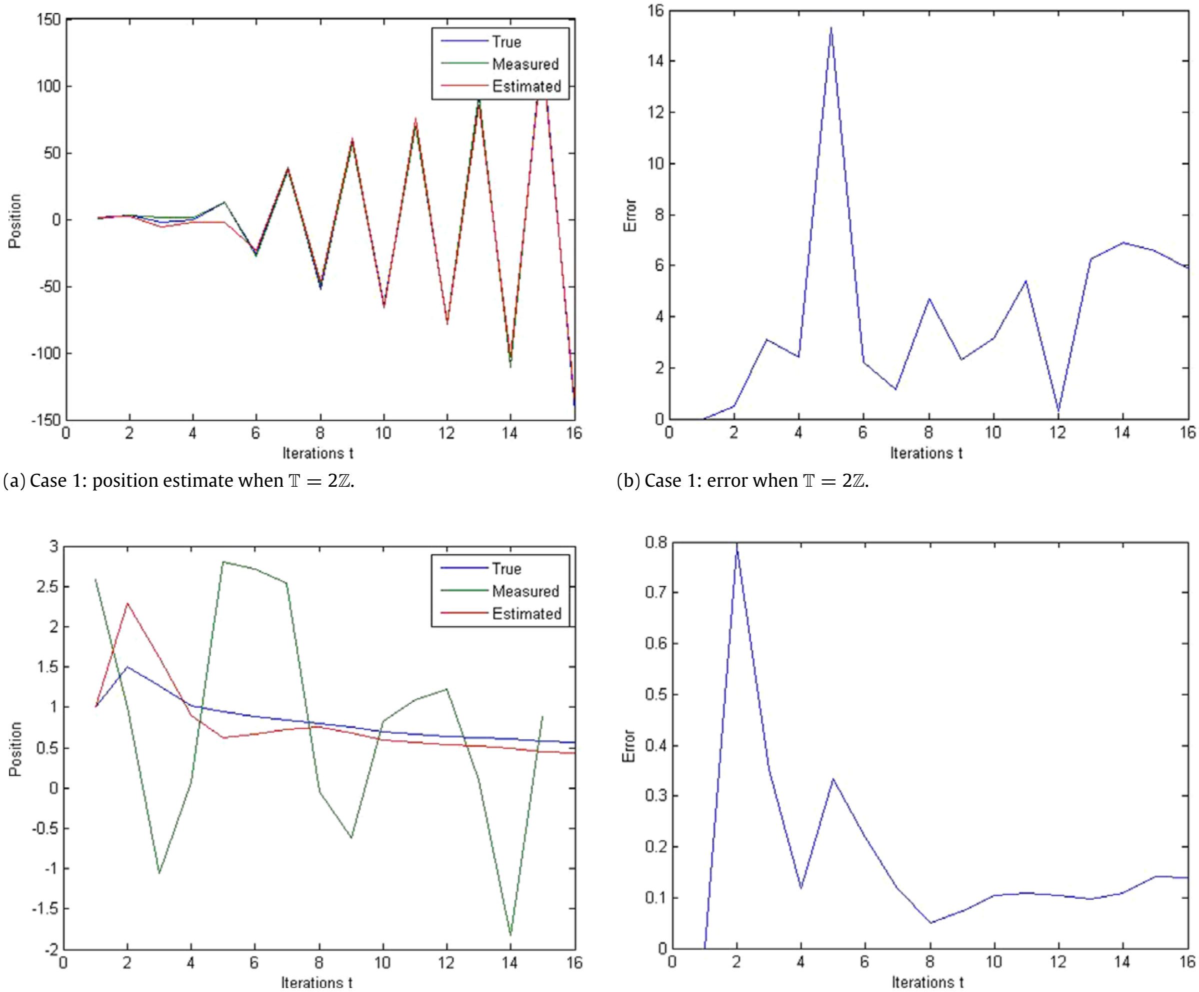
For convenience, we consider only time scales with bounded graininess, where it is assumed that the time scale is known *a priori*. In estimating the object's true position, we implemented our filter for 15 iterations. Note that since the Riccati equation and gain do not depend on the current state, these equations can be pre-computed and stored offline. In the first three cases, we use the same time scale throughout the entire iteration ($\mathbb{T} = 2\mathbb{Z}$, $\mathbb{T} = \{H_n\}$, and $\mathbb{T} = \mathbb{P}_{1,2}$, respectively). In the fourth case, we let $\mathbb{T} = 2\mathbb{Z}$ for $t \le 8$ and $\mathbb{T} = \{H_n\}$ for $t > 8$. As a result, the Riccati and estimator equations are altered midway through the implementation of the filter. It follows that the gain is also changed as the time scale changes. This is an example of a useful engineering technique called gain scheduling. In Fig. 2, we plot the true position, measurement, and position estimate as well as the error for each case. Note that the absolute value in the error appears to be bounded in each case.

4. A comparison of the LQR and Kalman filter

In this section, we establish a mathematical relationship between the LQR and Kalman filter on time scales. First note the Riccati equation and Kalman gain for the LQR as given in (7) and (8), respectively. These equations are in fact mathematically dual to the Riccati equation (22) and gain (21) associated with the Kalman filter, preserving the dual relationship between the LQR and the Kalman filter in their unification. Intuitively, we can see that both mirror two concepts we have shown are dual in [7]. The LQR mirrors the concept of controllability in that we are seeking an optimal control such that some cost function is minimized. Similarly, the Kalman filter mirrors the concepts of observability in that we are seeking an optimal estimate based on previously observed measurements. A comparison of the Riccati equations and gains is given in Table 5.

Finally, we show that when the final time t_f is fixed, the optimal estimator associated with

$$\begin{aligned} x^{\Delta}(t) &= Ax(t) + Gw(t), \qquad x(t_0) = x_0, \\ y(t) &= Cx(t) + v(t) \end{aligned} \tag{24}$$



(a) Case 1: position estimate when $\mathbb{T} = 2\mathbb{Z}$.

(b) Case 1: error when $\mathbb{T} = 2\mathbb{Z}$.

(c) Case 2: position estimate when $\mathbb{T} = \{H_n\}$.

(d) Case 2: error when $\mathbb{T} = \{H_n\}$.

Fig. 2. Kalman filter schemes to estimate position.

Table 5
A comparison of the LQR and Kalman filter.

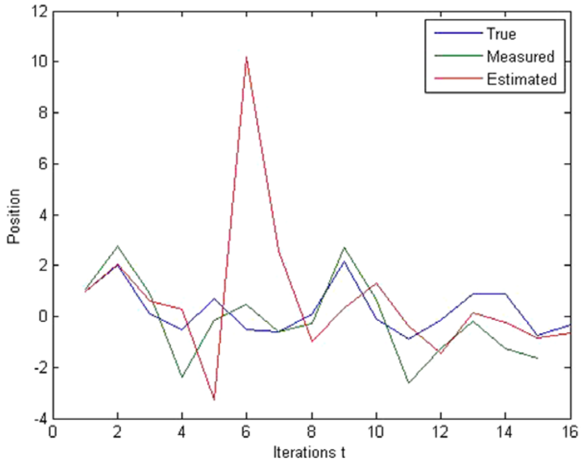
LQR	Kalman filter
$-S^{\Delta}(t)$	$P^{\Delta}(t)$
$S^{\sigma}(t)$	$P(t)$
A^T	A
B^T	C
$R > 0$	$R > 0$
Q	GQG^T

can be rewritten as an optimal regulator process. Now in estimating $x(t_f)$, we want to find a number β from the measurement such that

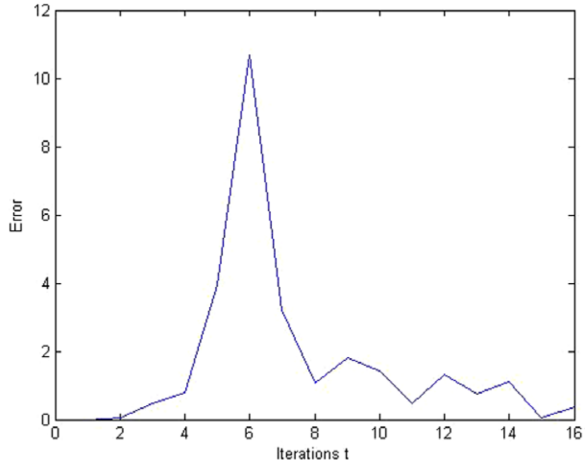
$$J := \mathbb{E}[(\alpha^T x(t_f) - \beta)^2],$$

is a minimum, where α is some constant vector. We refer to β as the minimum variance estimate of $\alpha^T x(t_f)$. Note that since all random variables of (24) are assumed to be Gaussian, we can derive β from linear operations of y. In other words, we assume that there exists some function $s(\cdot; \alpha, t_f)$ for $[t_0, t_f]$ such that

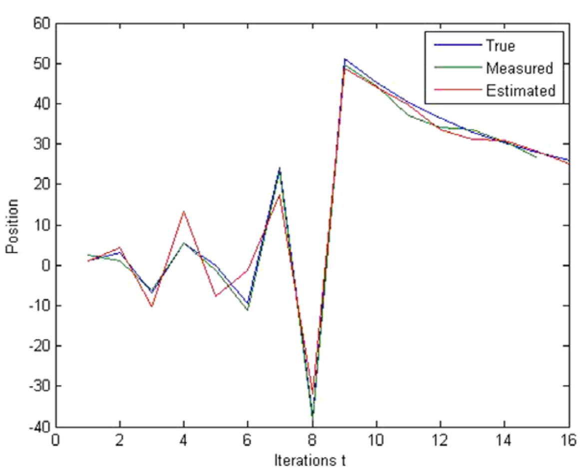
$$\beta = \int_{t_0}^{t_f} s^T(\tau; \alpha, t_f) y(\tau) \Delta\tau. \tag{25}$$



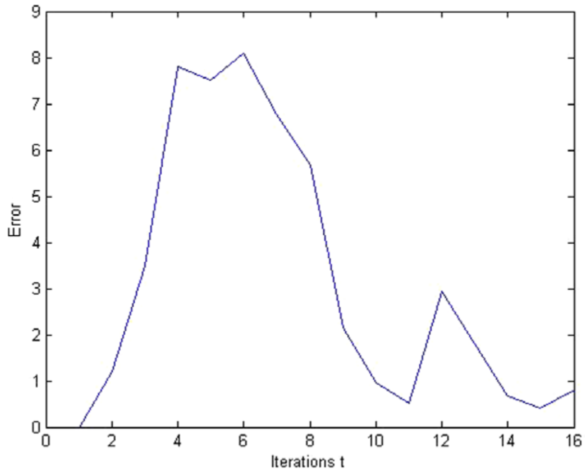
(e) Case 3: position estimate when $\mathbb{T} = \mathbb{P}_{1,2}$.



(f) Case 3: error when $\mathbb{T} = \mathbb{P}_{1,2}$.



(g) Case 4: position estimate when $\mathbb{T} = 2\mathbb{Z}$ for $t \le 8$ and $\mathbb{T} = \{H_n\}$ for $t > 8$.



(h) Case 4: error when $\mathbb{T} = 2\mathbb{Z}$ for $t \le 8$ and $\mathbb{T} = \{H_n\}$ for $t > 8$.

Fig. 2. (*continued*)

Now plugging in (25), we can show that

$$J = \mathbb{E}\left[\left(\alpha^T x(t_f) - \int_{t_0}^{t_f} s^T(\tau; \alpha, t_f) y(\tau) \Delta\tau\right)^2\right] \tag{26}$$

can be rewritten as the quadratic performance index. This brings us to our next theorem.

Theorem 4.1. *Suppose that x and y solve (24) and define J by (26). If $z \in \mathbb{R}^n$ and $s \in \mathbb{R}^m$ solve the deterministic terminal value problem*

$$z^\Delta(t) = -A^T z^\sigma(t) + C^T s(t), \qquad z(t_f) = \alpha, \tag{27}$$

then we have

$$J = z^T(t_0) P_0 z(t_0) + \int_{t_0}^{t_f} [(z^\sigma)^T G Q G^T z^\sigma + s^T R s](\tau) \Delta\tau. \tag{28}$$

Proof. First note that z is of the same dimension as x. Then using Theorem 2.7b, (24) and (27), we have

$$\begin{aligned} (z^T x)^\Delta &= (z^\Delta)^T x + (z^\sigma)^T x^\Delta \\ &= -(z^\sigma)^T A x + s^T C x + (z^\sigma)^T A x + (z^\sigma)^T G w \\ &= s^T y - s^T v + (z^\sigma)^T G w. \end{aligned}$$

Integrating both sides from t_0 to t_{f}, we have

$$z^T(t_{\mathrm{f}})x(t_{\mathrm{f}}) - z^T(t_0)x(t_0) = \int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)y(\tau)\Delta\tau - \int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)v(\tau)\Delta\tau + \int_{t_0}^{t_{\mathrm{f}}} (z^\sigma)^T(\tau)Gw(\tau)\Delta\tau.$$

Now, rearranging terms we have

$$\alpha^T x(t_{\mathrm{f}}) - \int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)y(\tau)\Delta\tau = z^T(t_0)x(t_0) - \int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)v(\tau)\Delta\tau + \int_{t_0}^{t_{\mathrm{f}}} (z^\sigma)^T(\tau)Gw(\tau)\Delta\tau.$$

Recalling that x_0, w, v are mutually uncorrelated of each other by Assumption 3.1f, we can write

$$\begin{aligned} J &= \mathbb{E}\left[\left(\alpha^T x(t_{\mathrm{f}}) - \int_{t_0}^{t_{\mathrm{f}}} s^T(\tau;\alpha,t_{\mathrm{f}})y(\tau)\Delta\tau\right)^2\right] \\ &= \mathbb{E}[(z^T(t_0)x(t_0))^2] + \mathbb{E}\left[\left(\int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)v(\tau)\Delta\tau\right)^2\right] + \mathbb{E}\left[\left(\int_{t_0}^{t_{\mathrm{f}}} (z^\sigma)^T(\tau)Gw(\tau)\Delta\tau\right)^2\right]. \end{aligned}$$

Next, we calculate the expectations of each term on the right-hand side separately. Thus,

$$\begin{aligned} \mathbb{E}[(z^T(t_0)x(t_0))^2] &= \mathbb{E}[z^T(t_0)x(t_0)x^T(t_0)z(t_0)] \\ &= z^T(t_0)\mathbb{E}[x(t_0)x^T(t_0)]z(t_0) \\ &= z^T(t_0)P_0 z(t_0). \end{aligned}$$

Also,

$$\begin{aligned} \mathbb{E}\left[\left(\int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)v(\tau)\Delta\tau\right)^2\right] &= \mathbb{E}\left[\int_{t_0}^{t_{\mathrm{f}}}\int_{t_0}^{t_{\mathrm{f}}} s^T(\tau_1)v(\tau_1)v^T(\tau_2)s(\tau_2)\Delta\tau_1\Delta\tau_2\right] \\ &= \int_{t_0}^{t_{\mathrm{f}}}\int_{t_0}^{t_{\mathrm{f}}} s^T(\tau_1)\mathbb{E}[v(\tau_1)v^T(\tau_2)]s(\tau_2)\Delta\tau_1\Delta\tau_2 \\ &= \int_{t_0}^{t_{\mathrm{f}}}\int_{t_0}^{t_{\mathrm{f}}} s^T(\tau_1)R\delta(\tau_1,\tau_2)s(\tau_2)\Delta\tau_1\Delta\tau_2 \\ &= \int_{t_0}^{t_{\mathrm{f}}} s^T(\tau)Rs(\tau)\Delta\tau. \end{aligned}$$

Similarly,

$$\mathbb{E}\left[\left(\int_{t_0}^{t_{\mathrm{f}}} (z^\sigma)^T(\tau)Gw(\tau)\Delta\tau\right)^2\right] = \int_{t_0}^{t_{\mathrm{f}}} (z^\sigma)^T(\tau)GQG^T z^\sigma(\tau)\Delta\tau.$$

Hence (28) holds. □

Note that all of the terms on the right-hand side of (28) are deterministic. Then as a regulator problem, the goal would be to determine an optimal control s such that (28) is minimized. Unlike (3), (27) is associated with a quadratic cost functional with initial weighting function rather than a terminal weighting function. This is due to the fact that the equations that describe the optimal estimator operate "forward in time", whereas the equations that describe the optimal regulator operate "backward in time".

Remark 4.2. Now that the LQR and LQE problems have been unified and extended to dynamic equations on time scales, we can introduce another fundamental problem in optimal control. The linear quadratic Gaussian (LQG) problem concerns stochastic linear systems disturbed by white noise, corrupted measurements of the state, and associated with a quadratic cost function. This problem is essentially a combination of the LQR and LQE. Consider the linear system

$$\begin{aligned} &x^\Delta(t) = Ax(t) + Bu(t) + Gw(t), \qquad x(t_0) = x_0, \\ &y(t) = Cx(t) + v(t), \end{aligned}$$

associated with the cost functional

$$J = \mathbb{E}\left[\frac{1}{2}x^T(t_{\mathrm{f}})Fx(t_{\mathrm{f}}) + \frac{1}{2}\int_{t_0}^{t_{\mathrm{f}}} (x^T Mx + u^T Nu)(\tau)\Delta\tau\right], \tag{29}$$

where $F, M \geq 0$ and $N > 0$ and Assumption 3.1a–h are assumed. The control objective then is to find an optimal control that minimizes (29). Next we introduce the controller equations

$$\begin{aligned} \hat{x}^{\Delta}(t) &= A\hat{x}(t) + Bu(t) + K(t)[y(t) - C\hat{x}(t)], \qquad \hat{x}(t_0) = \bar{x}_0 \\ u(t) &= -L(t)\hat{x}(t). \end{aligned} \tag{30}$$

Now using the state-estimate feedback in (30), the closed-loop can be written as

$$\begin{aligned} x^{\Delta}(t) &= Ax(t) - BL(t)\hat{x}(t) + Gw(t) \\ &= (A - BL(t))x(t) + BL(t)\tilde{x}(t) + Gw(t). \end{aligned}$$

Similarly, the error system is as given in (16). Then the closed-loop dynamics can be described by the system

$$\begin{bmatrix} x \\ \tilde{x} \end{bmatrix}^{\Delta}(t) = \begin{bmatrix} A - BL(t) & BL(t) \\ 0 & A - K(t)C \end{bmatrix} \begin{bmatrix} x \\ \tilde{x} \end{bmatrix}(t). \tag{31}$$

The system (31) describes the *separability principle* of the LQR and the Kalman filter. As a result, the Kalman filter and LQR can be designed and computed independent of each other. Now the Kalman filter estimates the state based previous measurements and is associated with the Riccati equation

$$\begin{aligned} P^{\Delta} &= AP + (I + \mu A)PA^T - (I + \mu A)PC^T(R + \mu CPC^T)^{-1}CP(I + \mu A^T) + GQG^T, \\ P(t_0) &= P_0. \end{aligned}$$

Using the solution P, the Kalman gain is given by (21). Similarly, the Riccati equation that solves the LQR problem is given by

$$\begin{aligned} -S^{\Delta} &= A^T S^{\sigma} + (I + \mu A^T)S^{\sigma}A - (I + \mu A^T)S^{\sigma}B(N + \mu B^T S^{\sigma} B)^{-1}B^T S^{\sigma}(I + \mu A) + M, \\ S(t_f) &= F. \end{aligned}$$

As a result, the feedback gain becomes

$$L = (N + \mu B^T S^{\sigma} B)^{-1} B^T S^{\sigma}(I + \mu A).$$

5. Future plans

In the numerical example provided, we implemented our filter when the system was exponentially stable and the graininess was bounded. From our results, we saw that the error could be bounded. However, finding necessary and sufficient conditions that ensure that the error is bounded remains an ongoing project.

Throughout this paper, we considered an initial state that was corrupted or missing data. However, it is possible that no information is known about the initial state. It is also possible that the dimension of the measurement is large compared to that of the process noise. As a result, initial error covariance can be infinitely large. To side-step this issue in the discrete and continuous cases, the inverse of the error covariance is used instead. To this end, we are extending our results to generalize the so-called information filter. In turn, these results can be used to generalize an optimal smoother on time scales.

It is also our intention to study other variations of our generalized filter. Namely, we are interested in extending our results to nonlinear dynamic systems. For the discrete and continuous cases, such a filter is called an extended Kalman filter. Here, the filter design is derived using Itô differentials and Brownian motion. In this setting, the underlying system can be estimated by an observer with either discrete or continuous measurements. As a result, it may be possible to estimate the true state when the system studied and the observer are on two different time scales.

Acknowledgments

We would like to thank the referee(s) for their thoughtful review of this manuscript and their insightful comments used to improve the quality of this paper.

References

[1] M. Athans, The importance of Kalman filtering methods for economic systems, Ann. Econ. Soc. Meas. 3 (1) (1974) 49–64.
[2] M. Bahmani-Oskooee, F. Brown, Kalman filter approach to estimate the demand for international reserves, Appl. Econ. 36 (15) (2004) 1655–1668.
[3] M. Bohner, A. Peterson, Dynamic Equations on Time Scales, Birkhäuser Boston Inc., Boston, MA, 2001.
[4] M. Bohner, A. Peterson (Eds.), Advances in Dynamic Equations on Time Scales, Birkhäuser Boston Inc., Boston, MA, 2003.
[5] M. Bohner, N. Wintz, The linear quadratic regulator on time scales, Int. J. Differ. Equ. 5 (2) (2010) 149–174.
[6] M. Bohner, N. Wintz, The linear quadratic tracker on time scales, Int. J. Dyn. Syst. Differ. Equ. 3 (4) (2011) 423–447.
[7] M. Bohner, N. Wintz, Controllability and observability of time-invariant linear dynamic systems, Math. Bohem. 137 (2) (2012) 149–163.
[8] K. Brammer, G. Siffling, Kalman–Bucy Filters, Artech House, Inc., Norwood, MA, 1989.
[9] R.G. Brown, Introduction to Random Signal Analysis and Kalman Filtering, John Wiley & Sons Inc., New York, NY, 1983.
[10] A.E. Bryson Jr., Y.C. Ho, Applied Optimal Control, Hemisphere Publishing Corp., Washington, D. C, 1975, Optimization, estimation, and control, Revised printing.

[11] G. De Nicolao, S. Strada, Kalman filtering with mixed discrete-continuous observations, Internat. J. Control 70 (1) (1998) 71–84.
[12] P. Dorato, V. Cerone, C. Abdallah, Linear-Quadratic Control: An Introduction, Simon & Schuster, 1994.
[13] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME D 82 (1960) 35–45.
[14] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, Trans. ASME D 83 (1961) 95–108.
[15] F.L. Lewis, Optimal Estimation, A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1986, With an introduction to stochastic control theory.
[16] F.L. Lewis, V.L. Syrmos, Optimal Control, Second ed., 1995.
[17] D.G. Luenberger, Observers for multivariable systems, IEEE Trans. Automat. Control 11 (2) (1966) 190–197.
[18] P.S. Maybeck, Stochastic Models, Estimation, and Control, in: Mathematics in Science and Engineering, vol. 141, 1979.
[19] P. Shi, Robust Kalman filtering for continuous-time systems with discrete-time measurements, IMA J. Math. Control Inform. 16 (3) (1999) 221–232.
[20] H.W. Sorenson, Least-squares estimation: from Gauss to Kalman, IEEE Spectr. 7 (1970) 63–68.
[21] P. Swerling, First-order error propagation in a stagewise smoothing procedure for satellite observations, J. Astronaut. Sci. 6 (1959) 46–52.
[22] N. Wintz, The Kalman filter on time scales, Ph.D. thesis, Missouri University of Science and Technology, Rolla, Missouri, USA, 2009.
[23] H. Zhang, M. Basin, M. Skliar, Optimal state estimation for continuous stochastic state-space system with hybrid measurements, Int. J. Innov. Comput. Inf. Control 2 (2) (2006) 357–370.