On Kalman filtering for conditionally Gaussian systems with random matrices *

Han-Fu CHEN

Institute of Systems Science, Academia Sinica, Beijing, China 100080

P.R. KUMAR

Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory, C.S.L., University of Illinois, 1101 West Springfield Ave., Urbana, IL 61801, U.S.A.

J.H. van SCHUPPEN

Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Abstract: We consider linear stochastic systems with additive white Gaussian noise, with the added generality that the system matrices are random and adapted to the observation process. The main result of this paper is that in order for the standard Kalman filter to generate the conditional mean and conditional covariance of the conditionally Gaussian distributed state, it is sufficient for the random matrices to be finite with probability one at each time instant. This generalizes the best previous results available to date, to our knowledge, which require the more stringent hypothesis that the entries of the random matrices should possess finite second moments at each time instant.

A significant application of the results of this paper is to the problem of recursive identification of the unknown parameters of a controlled linear stochastic system. In such problems, the observation matrix is typically generated by complicated nonlinear feedback, as for example in adaptive control, and the finiteness of the second moments is difficult, if not impossible, to establish, while the finiteness with probability one has been established in many applications.

Keywords: Kalman filter; parameter estimation; system identification; adaptive control; linear filtering.

1. Introduction

Consider the system

$$x(t+1) = A(t)x(t) + B(t)u(t) + D(t)w(t+1) \quad \text{for } t \geq 0, \tag{1a}$$

$$y(t) = C(t-1)x(t) + G(t-1)v(t-1) + F(t-1)w(t) \quad \text{for } t \geq 1, \tag{1b}$$

where:

(A1) $\{w(t)\}$ is i.i.d. and $w(t) \sim N(0, I)$,

(A2) $\left[\begin{smallmatrix} x(0) \\ y(0) \end{smallmatrix}\right]$ is dependent of $\{w(t)\}$, and $x(0)$ is conditionally Gaussian given $y(0)$ with conditional mean $\hat{x}(0)$ and conditional covariance $P(0)$ (the precise definition of conditional Gaussianity is given in Section 2),

(A3) $A(t)$, $B(t)$, $u(t)$, $D(t)$, $C(t)$, $G(t)$, $v(t)$ and $F(t)$ are all $\sigma(y(0), \ldots, y(t))$-measurable,

(A4) the entries of $A(t)$, $B(t)$, $u(t)$, $D(t)$, $C(t)$, $G(t)$, $v(t)$ and $F(t)$ are all finite with probability one.

Thus we have a system with additive white Gaussian noise $w(t)$. The initial state $x(0)$ is conditionally Gaussian given $y(0)$. The inputs $\left[\begin{smallmatrix} u(t) \\ v(t) \end{smallmatrix}\right]$ may be chosen by feedback from $(y(0), \ldots, y(t))$. The system matrices $A(t)$, $B(t)$, $C(t)$, $D(t)$, $F(t)$ and $G(t)$ are also similarly allowed to be $\sigma(y(0), \ldots, y(t))$-measurable.

The key generality of this paper is that we do *not* assume that $A(t)$, $B(t)$, $C(t)$, $D(t)$, $F(t)$, $G(t)$, $u(t)$ and $v(t)$ are integrable. Rather, we only assume the much weaker condition (A4) that their entries are finite a.s. We show that this weak condition is sufficient for the conditional distribution of the state $x(t)$ given $(y(0), \ldots, y(t))$ to be Gaussian, and that the corresponding conditional mean and conditional covariance are generated by the well-known Kalman filtering algorithm (specified in Section 3).

The best previous results available to date, to our knowledge, require the finiteness of the second moments of the entries of $B(t)u(t)$, $G(t)v(t)$, $D(t)$, and $F(t)$, almost sure uniform boundedness of the entries of $A(t)$ and $C(t)$, and a finite second moment condition on $x(0)$ and $y(0)$; see Liptser and Shiryayev [7; Assumptions 1-4 on page 62, Theorem 13.4 on page 65].

We have been drawn to this problem of removing integrability conditions because of its implications for system identification and adaptive control. Specifically, consider the system

$$y(t) = \sum_{i=1}^{p} \left[a_i y(t-i) + b_i u(t-i) \right] + w(t)$$

where $\{w(t)\}$ is a sequence of i.i.d. Gaussian random variables with mean 0 and variance 1, $\theta^0 := (a_1, \ldots, a_p, b_1, \ldots, b_p)$ is Gaussian, and w and θ^0 are independent. The control input $u(t)$ is assumed to be generated by feedback from $(y(0), \ldots, y(t))$, as for example in the important application area of adaptive control.

The problem of estimating θ^0 based on the observations $(y(0), \ldots, y(t))$ is the 'parameter estimation problem' of linear systems. In order to do so, one can rewrite the system as

$$\theta(t+1) = \theta(t), \quad \theta(0) = \theta^0,$$

$$y(t) = \phi^{\mathrm{T}}(t-1)\theta(t) + w(t),$$

where $\phi^{\mathrm{T}}(t-1) := (y(t-1), \ldots, y(t-p), u(t-1), \ldots, u(t-p))$ is the 'observation matrix'. Then one can estimate θ^0 by $E[\theta(t) \mid y(0), \ldots, y(t)]$.

This idea of using the Kalman filtering algorithm for parameter estimation dates back at least to Mayne [11]. Though it has since been part of the folklore (see Anderson and Moore [1], Caines [2], Kumar and Varaiya [6], Ljung and Söderstrom [8], etc.) that the Kalman filter generates the conditional mean $E[\theta(t) \mid y(0), \ldots, y(t)]$, the weakest assumption under which this appears to have been demonstrated is

$$E\|\phi(t)\|^2 < +\infty,$$

i.e., a finite second moment condition on the entries of $\phi(t)$; see Liptser and Shiryayev [7; Example 1, page 84].

This condition is however very difficult, if not impossible, to establish, when $u(t)$ is generated by nonlinear feedback from $(y(0), \ldots, y(t))$ as is typical in the important application area of adaptive control.

The results of this paper show that it is sufficient to have

$$\|\phi(t)\| < +\infty \quad \text{a.s.};$$

see condition (A4) above. This finiteness of the entries of $\phi(t)$ has been established in a variety of applications; see Meyn and Caines [12]. Thus, our results allow direct application to adaptive control; see Sternby [14], Rootzen and Sternby [15], and Kumar [5].

In any case, given the central role of the Kalman filter in a wide variety of applications, we feel that it is useful to state the precise result in its full generality.

2. Some preliminaries

Since the main thrust of this paper is to remove 'integrability' type assumptions, we will take as our starting point the well defined notion of conditional expectation for non-negative random variables; see Lemma I-2-9 of Neveu [13]. Proceeding from this, we define the conditional expectation used in this paper as follows.

Definition. Let (Ω, F, P) be a probability space, $G \subseteq F$ a sub-σ-algebra of F, and X a not necessarily integrable random variable. Let $X^+ := \max(X, 0)$ and $X^- := \max(-X, 0)$ be the positive and negative parts of X. If either $E(X^+ \mid G) < +\infty$ a.s. or $E(X^- \mid G) < +\infty$ a.s., then we will define (a version of) the condition expectation of X given G as

$$E(X \mid G) := E(X^+ \mid G) - E(X^- \mid G) \quad \text{a.s.}$$

Note that if X is a vector of random variables, then by $E(X \mid G)$, $|X|$, X^+ and X^- we shall denote the vector of corresponding components. By a vector inequality such as $X^+ < +\infty$, we shall mean the corresponding inequality for all its components.

It should be noted that one need *not* have $E[E[X \mid G] \mid H] = E[X \mid H]$ if $H \subset G$. For a trivial example, note that if $H = \{\phi, \Omega\}$, then the right-hand-side above need not even be defined. However, if X is integrable, then our definition of conditional expectation coincides with the usual one, and so all such properties hold.

However, the following corollary of Lemmas I-2-9 and I-2-10 of Neveu [13] shows that one such important property holds even without the integrability assumption.

Lemma 0. *If α is a G-measurable random variable, with $|\alpha| < +\infty$ a.s., and $E(|X| \mid G) < +\infty$ a.s., then*

$$E(\alpha X \mid G) = \alpha E(X \mid G) \quad \text{a.s.}$$

Proof. Note that

$$E((\alpha X)^+ \mid G) = \alpha^+ E(X^+ \mid G) + \alpha^- E(X^- \mid G) < +\infty \quad \text{a.s.}$$

from Lemmas I-2-9 and I-2-10 of Neveu [13]. Similarly,

$$E((\alpha X)^- \mid G) = \alpha^+ E(X^- \mid G) + \alpha^- E(X^+ \mid G) < +\infty \quad \text{a.s.}$$

Hence, the result follows. $\square$

We will use the following definition of conditional Gaussianity for not necessarily integrable random variables. In what follows, by $E(X \mid Y)$ we shall mean $E(X \mid \sigma(Y))$ where $\sigma(Y)$ is the sub-σ-algebra generated by the random variable (or vector) Y.

Definition. Let $|X| < +\infty$ a.s. We shall say that X is conditionally Gaussian given Y if there exist a Y-measurable random vector $\hat{X}$ and a Y-measurable random matrix $P = P^T \geq 0$ a.s. such that

$$E\left[\exp(i\lambda^T X) \mid Y\right] = \exp\left(i\lambda^T \hat{X} - \tfrac{1}{2}\lambda^T P \lambda\right) \quad \text{a.s.} \tag{2}$$

for every constant vector λ.

The following lemma shows that under the above definition the conditional distribution of X given Y is indeed Gaussian, and moreover that $\hat{X}$ and P are the conditional mean and covariance, respectively.

Lemma 1. *Suppose X is conditionally Gaussian given Y, i.e., $|X| < +\infty$ a.s. and (2) holds. Then $\hat{X} = E(X \mid Y)$ a.s., $P = E((X - \hat{X})(X - \hat{X})^T \mid Y)$ a.s. and for any $A \in B^n$, where B is the Borel σ-algebra on $\mathbf{R}$, $P(X \in A \mid Y)$ is the Gaussian measure, with mean $\hat{X}$ and covariance P, of the set A a.s.*

Proof. Since we are only concerned with the two random vectors X and Y, we can take the basic probability space (Ω, F, P) to be just the sample probability space, i.e., $\Omega = \mathbf{R}^{n+m}$ and $F = B^{n+m}$, assuming X is n-dimensional and Y is m-dimensional, and that X and Y are just the coordinate functions, i.e., for $\omega = (\omega_1, \dots, \omega_{n+m})$, $X(\omega) = (\omega_1, \dots, \omega_n)$ and $Y(\omega) = (\omega_{n+1}, \dots, \omega_{n+m})$. Since the range of X is a Borel set, it follows from Theorem 9.5, Chapter I of Doob [4] that there exists a conditional distribution $F(\cdot, \cdot)$ with the property that for every $A \in B^n$, $P(X \in A \mid Y)(\omega) = F(A, \omega)$ for a.e. ω. Moreover, by Theorem 9.1, Chapter I of Doob [4], for every vector λ, $E[\exp(i\lambda^T X) \mid Y](\omega) = \int \exp(i\lambda^T x) F(dx, \omega)$ for a.e. ω. Thus, by (2), for every vector λ,

$$\int_{\mathbf{R}^n} \exp(i\lambda^T x) F(dx, \omega) = \exp\left(i\lambda^T \hat{X}(\omega) - \tfrac{1}{2}\lambda^T P(\omega)\lambda\right) \quad \text{for a.e. } \omega.$$

Let $\tilde{R}^n = \{\lambda \in \mathbf{R}^n \mid \lambda = (\lambda_1, \dots, \lambda_n)^T$ and every λ_i is rational$\}$. Then there exists a set $N \subseteq \Omega$, $P(N) = 0$, such that

$$\int_{\mathbf{R}^n} \exp(i\lambda^T x) F(dx, \omega) = \exp\left(i\lambda^T \hat{X}(\omega) - \tfrac{1}{2}\lambda^T P(\omega)\lambda\right) \quad \text{for all } \lambda \in \tilde{R}^n, \omega \in N^c.$$

For fixed $\omega \in N^c$, since $\int_{\mathbf{R}^n} \exp(i\lambda^T x) F(dx, \omega)$ is a continuous function of λ (see Chung [3, page 143]), it follows that

$$\int_{\mathbf{R}^n} \exp(i\lambda^T x) F(dx, \omega) = \exp\left(i\lambda^T \hat{X}(\omega) - \tfrac{1}{2}\lambda^T P(\omega)\lambda\right) \quad \text{for all } \lambda \in \mathbf{R}^n, \omega \in N^c.$$

Due to the one-to-one correspondence between characteristic functions and distribution functions (see Theorem 6.6.2 of Chung [3]), it follows that for every $\omega \in N^c$, $F(\cdot, \omega)$ is the Gaussian distribution with mean $\hat{X}(\omega)$ and covariance $P(\omega)$. Thus $P(X \in A \mid Y)(\omega) = F(A, \omega)$ a.s., as claimed. Moreover, by the Restricted Integration Theorem of Loéve [9, page 359], it follows that $E(X^+ \mid Y)(\omega) = \int_{\mathbf{R}^n} x^+ F(dx, \omega)$ for a.e. ω, and the right hand side is finite for a.e. ω. Similarly, $E(X^- \mid Y)(\omega) = \int_{\mathbf{R}^n} x^- F(dx, \omega)$ for a.e. ω, and is also finite for a.e. ω. Thus $E(X \mid Y) = \int_{\mathbf{R}^n} x F(dx, \omega) = \hat{X}$ a.s. Similarly,

$$E\left[(X - \hat{X})(X - \hat{X})^T \mid Y\right](\omega) = \int_{\mathbf{R}^n} (x - \hat{X}(\omega))(x - \hat{X}(\omega))^T F(dx, \omega) = P(\omega) \quad \text{for a.e. } \omega,$$

proving the claim. □

The following lemma shows that the usual properties of conditionally Gaussian random variables are valid for our definition of conditional Gaussianity.

Lemma 2. (i) *If X is conditionally Gaussian given Z, and $A(\cdot)$ and $b(\cdot)$ are Borel measurable functions with $\|A(Z)\| < +\infty$ a.s. and $\|b(Z)\| < +\infty$ a.s., then $A(Z)X + b(Z)$ is also conditionally Gaussian given Z.*

(ii) *If $\begin{bmatrix} X \\ Y \end{bmatrix}$ is conditionally Gaussian given Z, then X and Y are conditionally independent given Z if and only if $E[(X - E(X \mid Z))(Y - E(Y \mid Z))^T \mid Z] = 0$ a.s.*

Proof. (i) If X is n-dimensional and Z is m-dimensional, as earlier, we take the basic probability space (Ω, F, P) to be just the sample space, and X, Z to be the coordinate random variables, i.e., $\Omega := \mathbf{R}^{n+m}$, $F := B^{n+m}$, $X(\omega) := (\omega_1, \dots, \omega_n)$, and $Z(\omega) := (\omega_{n+1}, \dots, \omega_{n+m})$, for $\omega = (\omega_1, \dots, \omega_{n+m})$. From Loéve [9, Section 27.b, page 363] it follows that there exists a conditional distribution $F(A, (\omega_{n+1}, \dots, \omega_{n+m}))$ such that for every $\lambda \in \mathbf{R}^n$,

$$E\left[\exp(i\lambda^T A(Z) X) \mid Z\right](\omega) = \int_{\mathbf{R}^{n+m}} \exp\left(i\lambda^T A((\omega_{n+1}, \dots, \omega_{n+m}))x\right) F(dx, \omega_{n+1}, \dots, \omega_{n+m})$$

for a.e. $\omega = (\omega_1, \dots, \omega_{n+m})$.

However, as shown in Lemma 1, except for ω in a null set, $F(\cdot, \omega)$ is a Gaussian distribution with mean (say) $\hat{X}((\omega_{n+1}, \ldots, \omega_{n+m}))$ and covariance $P((\omega_{n+1}, \ldots, \omega_{n+m}))$. Hence the evaluation of the integral yields

$$
\begin{aligned}
E\left[\exp(i\lambda^{T}[A(Z)X+b(Z)]) \mid Z\right] &= \exp(i\lambda^{T}b(Z))E\left[\exp(i\lambda^{T}A(Z)X) \mid Z\right] \\
&= \exp(i\lambda^{T}b(Z))\left[\exp(i\lambda^{T}A(Z)\hat{X} - \tfrac{1}{2}\lambda^{T}A(Z)PA^{T}(Z)\lambda)\right] \\
&= \exp(i\lambda^{T}[A(Z)\hat{X} + b(Z)] - \tfrac{1}{2}\lambda^{T}[A(Z)PA^{T}(Z)]\lambda) \quad \text{a.s.}
\end{aligned}
$$

This proves that $A(Z)X + b(Z)$ is conditional Gaussian given Z, and moreover that the conditional mean is $A(Z)\hat{X} + b(Z)$, while the conditional covariance is $A(Z)PA^{T}(Z)$.

(ii) This follows by straightforward computation which shows that

$$
E\left[\exp(i\lambda^{T}X + i\mu^{T}Y) \mid Z\right] = E\left[\exp(i\lambda^{T}X) \mid Z\right] E\left[\exp(i\mu^{T}Y) \mid Z\right] \quad \text{a.s.}
$$

if and only if

$$
E\left[(X - E(X \mid Z))(Y - E(Y \mid Z))^{T} \mid Z\right] = 0 \quad \text{a.s.} \qquad \square
$$

The next lemma shows the Gaussian 'updating' formula to be valid for random variables which are conditionally Gaussian.

Lemma 3. *Suppose that $[\begin{smallmatrix}X\\Y\end{smallmatrix}]$ is conditionally Gaussian given Z with conditional covariance*

$$
\begin{bmatrix} P_{xx\mid z} & P_{xy\mid z} \\ P_{yx\mid z} & P_{yy\mid z} \end{bmatrix} \quad \text{a.s.}
$$

Then:

(i) *Given (Z, Y), X is conditionally Gaussian with conditional mean*

$$
E[X \mid Z, Y] = E[X \mid Z] + P_{x,y\mid z}P_{y,y\mid x}^{+}(Y - E(Y \mid Z)) \quad \text{a.s.,}
$$

and conditional covariance $P_{xx\mid zy} := P_{xx\mid z} - P_{xy\mid z}P_{yy\mid z}^{+}P_{yx\mid z}$. Here P^{+} is the pseudo-inverse of P which (uniquely) satisfies $PP^{+}P = P$, as well as $P^{+} = UP^{T} = P^{T}V$, for some matrices U and V.

(ii) *Given Z, $X - E(X \mid Z, Y)$ is conditionally Gaussian and independent of Y.*

Proof. Note first since $P_{yy\mid z}$ is Z-measurable one can choose a Z-measurable version of $P_{yy\mid z}^{+}$. Let us define $W := X - E(X \mid Z) - P_{xy\mid z}P_{yy\mid z}^{+}[Y - E(Y \mid Z)]$. It follows readily from Lemma 2(i) that $[\begin{smallmatrix}W\\Y\end{smallmatrix}]$ is conditionally Gaussian given Z. Moreover, a straightforward computation which consists of verifying the conditions of Lemma 2(ii) using the properties of the pseudo-inverse (see Marsaglia [10]) shows that W and Y are conditionally independent given Z. Hence $[\begin{smallmatrix}W^{+}\\W^{-}\end{smallmatrix}]$ and Y are alsod conditionally independent given Z. Thus $E[W^{+} \mid Z, Y] = E[W^{+} \mid Z]$ and $E[W^{-} \mid Z, Y] = E[W^{-} \mid Z]$ a.s., and so $E[W \mid Z, Y] = E[W \mid Z] = 0$ a.s., where the last equality follows by straightforward computation. Thus, using the definition of W, we obtain $E[X \mid Z, Y] = E[X \mid Z] + P_{xy\mid z}P_{yy\mid z}^{+}[Y - E(Y \mid Z)]$ as claimed. Consequently, we also have $W = X - E(X \mid Z, Y)$, and so the claim (ii) is proved. To complete the proof of (i) note that

$$
\begin{aligned}
E(\exp(i\lambda^{T}X) \mid Z, Y) &= \exp(i\lambda^{T}E(X \mid Z, Y))W\left[\exp(i\lambda^{T}(X - E(X \mid Z, Y))) \mid Z, Y\right] \\
&= \exp(i\lambda^{T}E(X \mid Z, Y))E\left[\exp(i\lambda^{T}W) \mid Z, Y\right] \\
&= \exp(i\lambda^{T}E(X \mid Z, Y))E\left[\exp(i\lambda^{T}W) \mid Z\right]
\end{aligned}
$$

(since W and Y are conditionally independent given Z)

$$
= \exp(i\lambda^{T}E(X \mid Z, Y) - \tfrac{1}{2}\lambda^{T}[P_{xx\mid z} - P_{xy\mid z}P_{yy\mid z}^{+}P_{yx\mid z}]\lambda)
$$

(by straightforward computation).

This proves (i), thus completing the proof. $\square$

3. The main result

Consider the system (1) satisfying the assumptions (A1)–(A4). The main result of this paper is the following theorem which establishes the validity of the Kalman filter.

Theorem. *The state $x(t)$ is conditionally Gaussian given $(y(0), y(1), \ldots, y(t))$, with conditional mean $\hat{x}(t)$ and conditional covariance $P(t)$ which are given recursively by the following system of equations*:

$$\hat{x}(t+1) = A(t)\hat{x}(t) + B(t)u(t)$$
$$+ K(t)[y(t+1) - C(t)A(t)\hat{x}(t) - C(t)B(t)u(t) - G(t)v(t)], \quad \hat{x}(0), \qquad (3)$$

$$P(t+1) = R(t) - K(t)[C(t)R(t) + F(t)D^{\mathsf{T}}(t)], \quad P(0), \qquad (4)$$

where

$$K(t) := \left[(R(t)C^{\mathsf{T}}(t) + D(t)F^{\mathsf{T}}(t))\right]\left[C(t)R(t)C^{\mathsf{T}}(t) + F(t)F^{\mathsf{T}}(t)\right.$$
$$\left. + C(t)D(t)F^{\mathsf{T}}(t) + F(t)D^{\mathsf{T}}(t)C^{\mathsf{T}}(t)\right]^{+},$$

$$R(t) := A(t)P(t)A^{\mathsf{T}}(t) + D(t)D^{\mathsf{T}}(t).$$

Proof. Let us denote $Y(t) := (y(0), \ldots, y(t))$. First we show by induction that given $Y(t)$, both $x(t)$ and $\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix}$ are conditionally Gaussian.

Consider $t = 0$. By assumption, $x(0)$ is conditionally Gaussian given $Y(0) = y(0)$, and its conditional mean and covariance are finite a.s. Moreover, since

$$\begin{bmatrix} x(1) \\ y(1) \end{bmatrix} = \begin{bmatrix} A(0) & D(0) \\ C(0)A(0) & C(0)D(0) + F(0) \end{bmatrix} \begin{bmatrix} x(0) \\ w(1) \end{bmatrix} + \begin{bmatrix} B(0)u(0) \\ C(0)B(0)u(0) + G(0)v(0) \end{bmatrix},$$

in order to prove that $\begin{bmatrix} x(1) \\ y(1) \end{bmatrix}$ is conditionally Gaussian given $y(0)$, it suffices by Lemma 2(i) to show that $\begin{bmatrix} x(0) \\ w(1) \end{bmatrix}$ is conditionally Gaussian given $y(0)$. But this follows from (A2) since

$$E\left[\exp(i\lambda^{\mathsf{T}}x(0) + i\mu^{\mathsf{T}}w(1)) \mid y(0)\right] = E\left[\exp(i\lambda^{\mathsf{T}}x(0))E\left[\exp(i\mu^{\mathsf{T}}w(1)) \mid x(0), y(0)\right] \mid y(0)\right]$$

$$= E\left[\exp(i\mu^{\mathsf{T}}w(1))\right] E\left[\exp(i\lambda^{\mathsf{T}}x(0)) \mid y(0)\right]$$

$$= \exp\left(i\lambda^{\mathsf{T}}\hat{x}(0) - \tfrac{1}{2}\lambda^{\mathsf{T}}P(0)\lambda - \tfrac{1}{2}\mu^{\mathsf{T}}\mu\right) \quad \text{a.s.}$$

It should also be noted that the conditional mean and covariance are finite a.s.

We now proceed by induction, and suppose that, given $Y(t-1)$, both $x(t-1)$ and $\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$ are conditionally Gaussian with a.s. finite conditional means and conditional covariances. The conditional Gaussianity of $x(t)$ given $Y(t)$, with a.s. finite conditional mean and conditional covariance, follows immediately from Lemma 3(i). Note also that

$$E\left[\exp(i\lambda^{\mathsf{T}}x(t) + i\mu^{\mathsf{T}}w(t+1)) \mid Y(t)\right] = E\left[\exp(i\lambda^{\mathsf{T}}x(t))E\left[\exp(i\mu^{\mathsf{T}}w(t+1)) \mid x(t), Y(t)\right] \mid Y(t)\right]$$

$$= E\left[\exp(i\mu^{\mathsf{T}}w(t+1))\right] E\left[\exp(i\lambda^{\mathsf{T}}x(t)) \mid Y(t)\right] \quad \text{a.s.,}$$

and by the same argument as for $t = 0$, we deduce that $\begin{bmatrix} x(t) \\ w(t+1) \end{bmatrix}$ is conditionally Gaussian given $Y(t)$, with a.s. finite conditional mean and conditional covariance. Finally, since

$$\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} = \begin{bmatrix} A(t) & D(t) \\ C(t)A(t) & C(t)D(t) + F(t) \end{bmatrix} \begin{bmatrix} x(t) \\ w(t+1) \end{bmatrix} + \begin{bmatrix} B(t)u(t) \\ C(t)B(t)u(t) + G(t)v(t) \end{bmatrix},$$

the induction is completed by Lemma 2(i).

It remains only to show that the formulas for the conditional mean and conditional covariance of $x(t)$ given $Y(t)$ are as claimed. By Lemma 3(i), by identifying Z with $Y(t)$, Y with $y(t+1)$, and X with $x(t+1)$, we obtain

$$E[x(t+1) \mid Y(t+1)] = E[x(t+1) \mid Y(t)] + P_{xy\mid z}P_{yy\mid z}^+[y(t+1) - E(y(t+1) \mid Y(t))] \quad \text{a.s.,} \quad (5)$$

where

$$P_{xy\mid z} := E\left[(x(t+1) - E(x(t+1) \mid Y(t)))(y(t+1) - E(y(t+1) \mid Y(t)))^T \mid Y(t)\right], \quad (6)$$

$$P_{yy\mid z} := E\left[(y(t+1) - E(y(t+1) \mid Y(t)))(y(t+1) - E(y(t+1) \mid Y(t)))^T \mid Y(t)\right]. \quad (7)$$

Note first that,

$$E(x(t+1) \mid Y(t)) = E[A(t)x(t) + B(t)u(t) + D(t)w(t+1) \mid Y(t)]$$
$$= A(t)E(x(t) \mid Y(t)) + B(t)u(t) \quad \text{a.s.,} \quad (8)$$

$$E[y(t+1) \mid Y(t)] = E[C(t)x(t+1) + G(t)v(t) + F(t)w(t+1) \mid Y(t)]$$
$$= C(t)E(x(t+1) \mid Y(t)) + G(t)v(t) \quad \text{a.s.} \quad (9)$$

Also, by substitution of (5), (8), (9) in (6),

$$P_{xy\mid z} = E\left[((A(t)(x(t) - E(x(t) \mid Y(t))))\right.$$
$$+ D(t)w(t+1))(C(t)A(t)(x(t) - E(x(t) \mid Y(t)))$$
$$+ (F(t) + C(t)D(t))w(t+1))^T \mid Y(t)\right]$$
$$= A(t)E\left[(x(t) - E(x(t) \mid Y(t)))(x(t) - E(x(t) \mid Y(t)))^T \mid Y(t)\right]A^T(t)C^T(t)$$
$$+ D(t)(F(t) + C(t)D(t))^T \quad \text{a.s.} \quad (10)$$

where the last equality arises because the cross-term can be evaluated as follows:

$$E\left[D(t)w(t+1)(C(t)A(t))(x(t) - E(x(t) \mid Y(t)))^T \mid Y(t)\right]$$
$$= E\left[D(t)E[w(t+1) \mid Y(t), x(t)](C(t)A(t)(x(t) - E(x(t) \mid Y(t))))^T \mid Y(t)\right]$$
$$= 0 \quad \text{a.s.}$$

(and similarly for the other cross-term). By a similar computation,

$$P_{yy\mid z} = C(t)A(t)E\left[(x(t) - E(x(t)) \mid Y(t))(x(t) - E(x(t) \mid Y(t)))^T \mid Y(t)\right]A^T(t)C^T(t)$$
$$+ (F(t) + C(t)D(t))(F(t) + C(t) + C(t)D(t))^T \quad \text{a.s.} \quad (11)$$

Finally, by the formula for the covariance 'update' $P_{xx\mid zy}$ in Lemma 3(i),

$$E\left[(x(t+1) - E(x(t+1) \mid Y(t+1)))(x(t+1) - E(x(t+1) \mid Y(t+1)))^T \mid Y(t+1)\right]$$
$$= E\left[(x(t+1) - E(x(t+1) \mid Y(t)))(x(t+1) - E(x(t+1) \mid Y(t)))^T\right] - P_{xy\mid z}P_{yy\mid z}^+P_{xy\mid z}$$
$$= E\left[(A(t)(x(t) - E(x(t) \mid Y(t)))) + D(t)w(t+1))\right.$$
$$\left.\cdot(A(t)(x(t) - E(x(t) \mid Y(t)))) + D(t)w(t+1))^T \mid Y(t)\right] - P_{xy\mid z}P_{yy\mid z}^+P_{yx\mid z}$$
$$= A(t)E\left[(x(t) - E(x(t) \mid Y(t)))(x(t) - E(x(t) \mid Y(t)))^T \mid Y(t)\right]A^T(t) + D(t)D^T(t)$$
$$- P_{xy\mid z}P_{yy\mid z}^+P_{yx\mid z} \quad \text{a.s.} \quad (12)$$

Substituting (10), (11) in (12) gives (4), while substituting (8)–(11) in (5) gives (3), completing the proof.
□

4. Concluding remarks

We have shown that integrability conditions on the entries of the random matrices can be removed, and replaced by the much weaker conditions of a.s. finiteness. This relaxation of conditions for the validity of the Kalman filter is very useful, particularly in identification and adaptive control, as the recent paper of Kumar [5] shows.

References

[1] B.D.O. Anderson and J. Moore, *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, NJ, 1979).

[2] P.E. Caines, *Linear Stochastic Systems* (John Wiley, New York, 1988).

[3] K. Chung, *A Course in Probability Theory* (Academic Press, New York, 1974).

[4] J. Doob, *Stochastic Processes* (John Wiley, New York, 1953).

[5] P.R. Kumar, Convergence of adaptive control schemes using least-squares parameter estimates, University of Illinois (Dec. 1988).

[6] P.R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control* (Prentice Hall, New York, 1986).

[7] R.S. Liptser and A.N. Shiryayev, *Statistics of Random Processes, II: Applications* (Springer-Verlag, New York, 1977).

[8] L. Ljung and T. Söderstrom, *Theory and Practice of Recursive Identification* (M.I.T. Press, Cambridge, MA, 1985).

[9] M. Loéve, *Probability Theory* (Van Nostrand, New York, 1963).

[10] G. Marsaglia, Conditional means and covariances of normal variables with singular covariance matrix, *J. Amer. Statist. Assoc.* **59** (1964) 1203–1204.

[11] D. Mayne, Optimal non-stationary estimation of the parameters of a linear system with Gaussian inputs, *J. Electron. Control* **14** (1963) 101–112.

[12] S. Meyn and P. Caines, The zero divisor problem of multivariable stochastic adaptive control, *Systems Control Lett.* **6** (1985) 235–238.

[13] J. Neveu, *Discrete-Parameter Martingales* (North-Holland, Amsterdam–New York, 1975).

[14] J. Sternby, On consistency for the method of least squares using martingale theory, *IEEE Trans. Automat. Control* **22** (1977) 346–352.

[15] H. Rootzen and J. Sternby, Consistency in least squares: A Bayesian approach, *Automatica* **20** (1984) 471–475.