

MODULE 7

THE CONTINUOUS KALMAN FILTER

"Liebnitz" or "Leibniz" rule:

$$\frac{\partial}{\partial y} \int_{\alpha(y)}^{\beta(y)} f(x,y) dx = \int_{\alpha(y)}^{\beta(y)} \frac{\partial}{\partial y} f(x,y) dx + f(\beta(y),y) \frac{\partial}{\partial y} \beta(y) - f(\alpha(y),y) \frac{\partial}{\partial y} \alpha(y).$$

Continuous-Time State Vector Equation

$$\begin{cases} \dot{X}(t) = F(t)X(t) + G(t)u(t) \\ y(t) = H(t)X(t) + D(t)u(t) \end{cases}, \quad t \geq t_0, \quad X(t_0) \text{ known.}$$

- The solution for the state vector $X(t)$ is unique

$$X(t) = \Phi(t,t_0)X(t_0) + \int_{t_0}^t \Phi(t,\theta)G(\theta)u(\theta)d\theta$$

- So is the solution for $y(t)$:

$$y(t) = H(t)\Phi(t,t_0)X(t_0) + \int_{t_0}^t H(t)\Phi(t,\theta)G(\theta)u(\theta)d\theta + D(t)u(t)$$

Where $\Phi(t_2, t_1)$ is the state transition matrix defined by:

$$\textcircled{A} \quad \dot{\Phi}(t, t_0) = F(t) \Phi(t, t_0)$$

(Solves the homogeneous state vector equation.)

$$\textcircled{B} \quad \Phi(t, t) = I.$$

The solution to $\textcircled{A}, \textcircled{B}$ is given by the Peano-Baker series

$$\begin{aligned} \Phi(t, t_0) = I &+ \int_{t_0}^t F(\theta_1) d\theta_1 + \int_{t_0}^t F(\theta_1) \int_{t_0}^{\theta_1} F(\theta_2) d\theta_2 d\theta_1 \\ &+ \int_{t_0}^t F(\theta_1) \int_{t_0}^{\theta_1} F(\theta_2) \int_{t_0}^{\theta_2} F(\theta_3) d\theta_3 d\theta_2 d\theta_1 + \dots \end{aligned}$$

→ This is not usually very useful for solving Φ .

Properties of the State Transition Matrix

- 1) $\Phi(t, t) = I$
- 2) $\Phi(t, t_0) = \Phi(t, t_1) \Phi(t_1, t_0)$
- 3) $\{\Phi(t, t_0)\}^{-1} = \Phi(t_0, t)$
- 4) $\frac{d}{dt} \Phi(t_0, t) = -\Phi(t_0, t) F(t)$



$$5) \frac{d}{dt} \Phi(t, t_0) = F(t) \Phi(t, t_0)$$

$$6) \det \Phi(t, \tau) = \exp \left[\int_{\tau}^t \text{Tr } F(\theta) d\theta \right]$$

$$7) \frac{d}{dt} \det \Phi(t, \tau) = \text{Tr } F(t) \det \Phi(t, \tau)$$

Orthogonality Principle

- Let the observation be $z(t) = y(t) + v(t)$
 $= H(t)x(t) + v(t)$
- Suppose that $\hat{x}(t|t)$ is the optimal MMSE estimate of $x(t)$ given $z(t)$ up to and including time t .
- The error in $\hat{x}(t|t)$ is $\tilde{x}(t|t) = x(t) - \hat{x}(t|t)$.
→ This error is uncorrelated with any function of the observations $z(\theta)$ for $t_0 \leq \theta < t$.
 $E \{ h(z(\theta)) \tilde{x}(t|t) \} = 0, t_0 \leq \theta < t$.
- The optimal MMSE estimate of $y(t)$ given $z(\theta)$ for $t_0 \leq \theta \leq t$ is $\hat{y}(t|t) = H(t)\hat{x}(t|t)$.
- The error in $\hat{y}(t|t)$ is $\tilde{y}(t|t) = y(t) - \hat{y}(t|t)$.

- Again by the orthogonality principle, the error $\tilde{y}(t|t)$ is uncorrelated with (orthogonal to) any function of the observations $z(\theta)$ for $t_0 \leq \theta < t$.

Discrete Innovations Process

- The observations are $z(t) = H(t)x(t) + v(t)$
- The correlation structure of $v(t)$ is trivial (white noise), but that of $x(t)$ is not trivial.
- So there is generally correlation between the observations from different times.
- The innovations process is obtained by orthogonalizing or "whitening" the observations to remove the redundant information from each new observation as it arrives.
- The resulting "innovations" give new information only.
- We look at discrete innovations first because they are more intuitive than continuous innovations.

- Suppose the discrete observations occur at time instants $t_1, t_2, \dots, t_k, \dots$.
- The innovations process e_k is obtained by applying Gram-Schmidt orthogonalization to the observations $z(t_1) \dots z(t_{k-1})$.

$$e_1 = z(t_1) \quad (\text{first innovation})$$

→ obtain a unit vector in the same direction:

$$\tilde{e}_1 = \frac{z(t_1)}{\sqrt{E[z^2(t_1)]}}$$

$$\begin{aligned} e_2 &= z(t_2) - \hat{z}(t_2 | t_1) \\ &= z(t_2) - E[z(t_2) \tilde{e}_1] \tilde{e}_1 \end{aligned}$$

$$\tilde{e}_2 = \frac{e_2}{\sqrt{E[e_2^2]}}$$

:

$$e_k = z(t_k) - \sum_{j=1}^{k-1} E[z(t_k) \tilde{e}_j] \tilde{e}_j$$

$$\tilde{e}_k = \frac{e_k}{\sqrt{E[e_k^2]}}$$

NOTE : No information is added or lost by forming the innovations process.

The optimal estimate of $X(t_k)$ or $Y(t_k)$ given $z(t_1), z(t_2), \dots, z(t_k)$ is exactly the same as the optimal estimate given e_1, e_2, \dots, e_k .

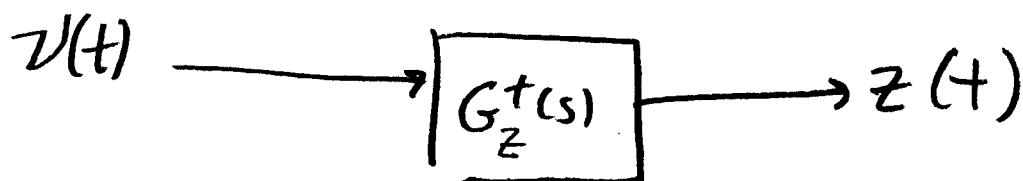
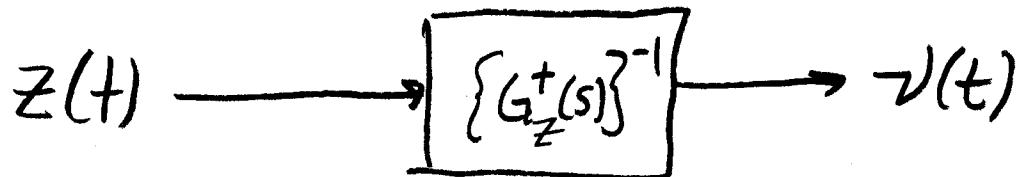
- Because of the orthogonal (and orthonormal) nature of $\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_k \dots$, it becomes very simple to compute the MMSE estimates of $Y(t_n)$ and $X(t_n)$ given $z(t_1) \dots z(t_n)$:

$$\begin{aligned}\hat{Y}(t_n | t_n) &= E[Y(t | z(t_1) \dots z(t_n))] \\ &= \sum_{k=1}^n E[Y(t_n) \tilde{e}_k] \tilde{e}_k\end{aligned}$$

$$\begin{aligned}\hat{X}(t_n | t_n) &= E[X(t | z(t_1) \dots z(t_n))] \\ &= \sum_{k=1}^n E[X(t_n) \tilde{e}_k] \tilde{e}_k.\end{aligned}$$

Continuous Innovations

- Intuitively, the continuous innovations $v(t)$ are obtained by applying a whitening filter to convert the colored noise process $z(t)$ into a white noise process.
- We will restrict our attention to the case where $z(t)$ is WSS (wide sense stationary).
- Let $G_z(s)$ be the PSD of $z(t)$.
- Then the innovations process is generated with an LTI system having transfer function $\{G_z^+(s)\}^{-1}$



$$\text{where } G_z(s) = \overline{G_z^+(s)} \overline{G_z^-(s)}$$

LHP poles
and zeros only

RHP poles and
zeros only

- The impulse response of the filter that generates $v(t)$ from $z(t)$ is given by $\mathcal{F}^{-1}\left\{\left[G_z^+(w)\right]^{-1}\right\}$. It is real-valued (and right-sided).
 - This implies that $\left[G_z^+(w)\right]^{-1}$ and $G_z^+(w)$ are conjugate symmetric.

- Applying the Wiener-Kinchine relation, we have that the power spectral density of $v(t)$ is

$$\begin{aligned} G_v(w) &= G_z(w) \left| \left[G^+(w)\right]^{-1} \right|^2 \\ &= G_z(w) \left[G^+(w) \right]^{-1} \left(\left[G^+(w) \right]^{-1} \right)^* \\ &= G_z(w) \left[G^+(w) \right]^{-1} \left[G^+(-w) \right]^{-1} \end{aligned}$$

→ But $G_z(w)$ is real and even.

$$\rightarrow \text{So } G^+(-w) = G^+(w)$$

We have

$$\begin{aligned} G_v(w) &= G_z(w) \left[G^+(w) \right]^{-1} \left[G^-(w) \right]^{-1} \\ &= G_z(w) \left[G^+(w) G^-(w) \right]^{-1} \\ &= G_z(w) \left[G_z(w) \right]^{-1} = 1 \end{aligned}$$

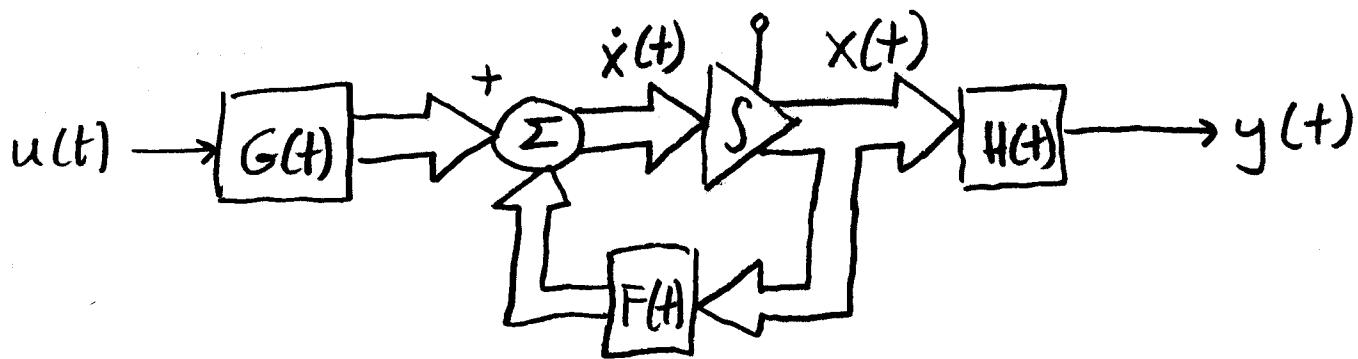
- Thus, we see that $v(t)$ has a flat power spectrum and is therefore a white noise.

Continuous Kalman Filter (kalman-Bucy Filter)

- The system is given by

$$\left\{ \begin{array}{l} \dot{x}(t) = F(t)x(t) + G(t)u(t) \\ y(t) = H(t)x(t) \end{array} \right. \quad (7.1.1)$$

$$(7.1.2)$$



The observations are: $z(t) = y(t) + v(t)$.

Assumptions:

A) $E[u(t)] = E[v(t)] = 0$.

This implies also that $E[x(0)] = 0$.

B) The covariance of $x(0)$ is given by

$$\Pi_0 = E[x(0)x^T(0)] \text{ and is known.}$$

$$C) E[X(0)V(t)] = 0 \quad \forall t$$

$$D) E[X(0)U(t)] = 0 \quad \forall t$$

$$E) E[U(t)U(\tau)] = Q(t)\delta(t-\tau) \quad (7.1.3)$$

$$F) E[V(t)V(\tau)] = R(t)\delta(t-\tau) \quad (7.1.4)$$

$$G) E[U(t)V(\tau)] = 0 \quad (7.1.5)$$

NOTE: we are assuming here that $u(t)$ and $v(t)$ are scalars. In the book they are assumed to be vectors. It's not difficult to make the extension; it primarily involves notation.

Goal: Find optimal MMSE estimates $\hat{X}(t|t)$ of $X(t)$ given $Z(\theta)$, $0 \leq \theta \leq t$ and optimal MMSE estimates $\hat{y}(t|t) = H(t)\hat{X}(t|t)$ of $y(t)$.

- The error in the filtered state vector estimate is $\tilde{x}(t|t) = x(t) - \hat{x}(t|t)$.

- The innovations process is

$$\begin{aligned}
 v(t) &= z(t) - \hat{y}(t|t) \\
 &= z(t) - H(t) \hat{x}(t|t) && \text{compare (7.1.24)} \\
 &= y(t) + v(t) - H(t) \hat{x}(t|t) \\
 &= H(t) x(t) + v(t) - H(t) \hat{x}(t|t) \\
 &= H(t) [x(t) - \hat{x}(t|t)] + v(t) \\
 &= H(t) \tilde{x}(t|t) + v(t)
 \end{aligned}$$

- We will seek an estimator $\hat{x}(t|t)$ that is a linear function of the innovations process $v(\theta)$ for $0 \leq \theta \leq t$:

$$\hat{x}(t|t) = \int_0^t g(t, \tau) v(\tau) d\tau \quad (*)$$

- The orthogonality principle coincides with the "projection theorem" of Hilbert Spaces in this case. It says that, for the optimal MMSE estimator, the error is orthogonal to all linear functions of the observations ...

... and so also to all linear functions of the innovations.

→ In the special case where $u(t)$ and $v(t)$ are jointly Gaussian, the MMSE linear estimator coincides with the Maximum a posteriori probability (MAP) estimator. In this case, the MMSE linear estimator is the MMSE estimator.

- The orthogonality principle guarantees that

$$\tilde{X}(t|t) \perp v(\tau) \quad \forall 0 \leq \tau < t$$

i.e., $E[\tilde{X}(t|t)v(\tau)] = 0 \quad \forall 0 \leq \tau < t$

Now, $\tilde{X}(t|t) = x(t) - \hat{x}(t|t)$,

so $x(t) = \hat{x}(t|t) + \tilde{X}(t|t)$

We have, for $0 \leq \tau < t$:

$$\begin{aligned} E[x(t)v(\tau)] &= E[\{\hat{x}(t|t) + \tilde{X}(t|t)\}v(\tau)] \\ &= E[\hat{x}(t|t)v(\tau)] + \underbrace{E[\tilde{X}(t|t)v(\tau)]}_{\text{zero}} \\ &= E[\hat{x}(t|t)v(\tau)] \quad (***) \end{aligned}$$

Plugging (*) on 7.10 (assumed form of the optimal estimator) into (**) on 7.11, we have (for $0 \leq \tau < t$):

$$\begin{aligned}
 E[X(t)\nu(\tau)] &= E[\hat{X}(t|t)\nu(\tau)] \\
 &= E\left[\int_0^t g(t,\theta)\nu(\theta)d\theta \nu(\tau)\right] \\
 &= E\left[\int_0^t g(t,\theta)\nu(\theta)\nu(\tau)d\theta\right] \\
 &= \int_0^t g(t,\theta)E[\nu(\theta)\nu(\tau)]d\theta \quad (*)
 \end{aligned}$$

But $\nu(\theta) = H(\theta)\tilde{X}(\theta|\theta) + v(\theta)$ and

$\nu(\tau) = H(\tau)\tilde{X}(\tau|\tau) + v(\tau)$ from 7.10.

$$\begin{aligned}
 \text{So } E[\nu(\theta)\nu(\tau)] &= E\left\{\left[H(\theta)\tilde{X}(\theta|\theta) + v(\theta)\right]\right. \\
 &\quad \times \left.\left[H(\tau)\tilde{X}(\tau|\tau) + v(\tau)\right]\right\} \\
 &= E\left\{H(\theta)\tilde{X}(\theta|\theta)\tilde{X}^T(\tau|\tau)H^T(\tau) + v(\theta)H(\tau)\tilde{X}(\tau|\tau)\right. \\
 &\quad \left.+ v(\tau)H(\theta)\tilde{X}(\theta|\theta) + v(\theta)v(\tau)\right\}
 \end{aligned}$$

$0 \leq \tau < t$

$0 \leq \theta \leq t$

"It can be shown" that $v(t)$ has the same correlation structure as $v(t)$, i.e., that

$$E[v(\theta)v(\tau)] = R(\theta)\delta(\theta-\tau)$$

- plugging this into (*) on 7.12, we have

$$\begin{aligned} E[X(t)v(\tau)] &= \int_0^t g(t,\theta) R(\theta)\delta(\theta-\tau) d\theta \\ &= g(t,\tau)R(\tau), \quad 0 \leq \tau < t. \end{aligned}$$

- Solving for the kernel $g(t,\tau)$ of the optimal estimator, we have

$$g(t,\tau) = E[X(t)v(\tau)] R^{-1}(\tau), \quad 0 \leq \tau < t.$$

- Plugging this into (*) on 7.10, we have

$$\begin{aligned} \hat{x}(t|t) &= \int_0^t g(t,\tau)v(\tau) d\tau \\ &= \int_0^t E[X(t)v(\tau)] R^{-1}(\tau)v(\tau) d\tau \\ &\quad (*) \end{aligned}$$

- Now apply Leibniz rule to both sides of (4) on 7.13 to get a differential equation that can be solved for $\hat{x}(t|t)$:

$$\dot{\hat{x}}(t|t) = \int_0^t \frac{\partial}{\partial t} \left\{ E[x(t)v(\tau)] R^{-1}(\tau) v(\tau) \right\} d\tau$$

$$+ 1 \cdot E[x(t)v(t)] R^{-1}(t) v(t) + 0$$

$$= E[x(t)v(t)] R^{-1}(t) v(t)$$

$$+ \int_0^t E[\dot{x}(t)v(\tau)] R^{-1}(\tau) v(\tau) d\tau$$

$$= E[x(t)v(t)] R^{-1}(t) v(t)$$

$$+ \int_0^t E \left\{ [F(t)x(t) + G(t)u(t)] v(\tau) \right\} R^{-1}(\tau) v(\tau) d\tau$$

$$= E[x(t)v(t)] R^{-1}(t) v(t)$$

$$+ F(t) \int_0^t E[x(t)v(\tau)] R^{-1}(\tau) v(\tau) d\tau$$

$$+ G(t) \int_0^t E[u(t)v(\tau)] R^{-1}(\tau) v(\tau) d\tau$$

zero

because for $\tau < t$ because $v(\tau)$ depends on $x(0)$, $u(\theta)$ for $0 \leq \theta \leq \tau$ and $v(\theta)$ for $0 \leq \theta \leq t$. When $\tau < t$, these are all uncorrelated with $u(t)$. PAGE 7.14

$$\begin{aligned}
 \dot{\hat{x}}(t|t) &= \dots \\
 &= E[x(t)v(t)]R^{-1}(t)v(t) \\
 &\quad + F(t) \underbrace{\int_0^t E[x(t)v(\tau)]R^{-1}(\tau)v(\tau)d\tau}_{\hat{x}(t|t)}
 \end{aligned}$$

$$\Rightarrow \dot{\hat{x}}(t|t) = F(t)\hat{x}(t|t) + E[x(t)v(t)]R^{-1}(t)v(t)$$

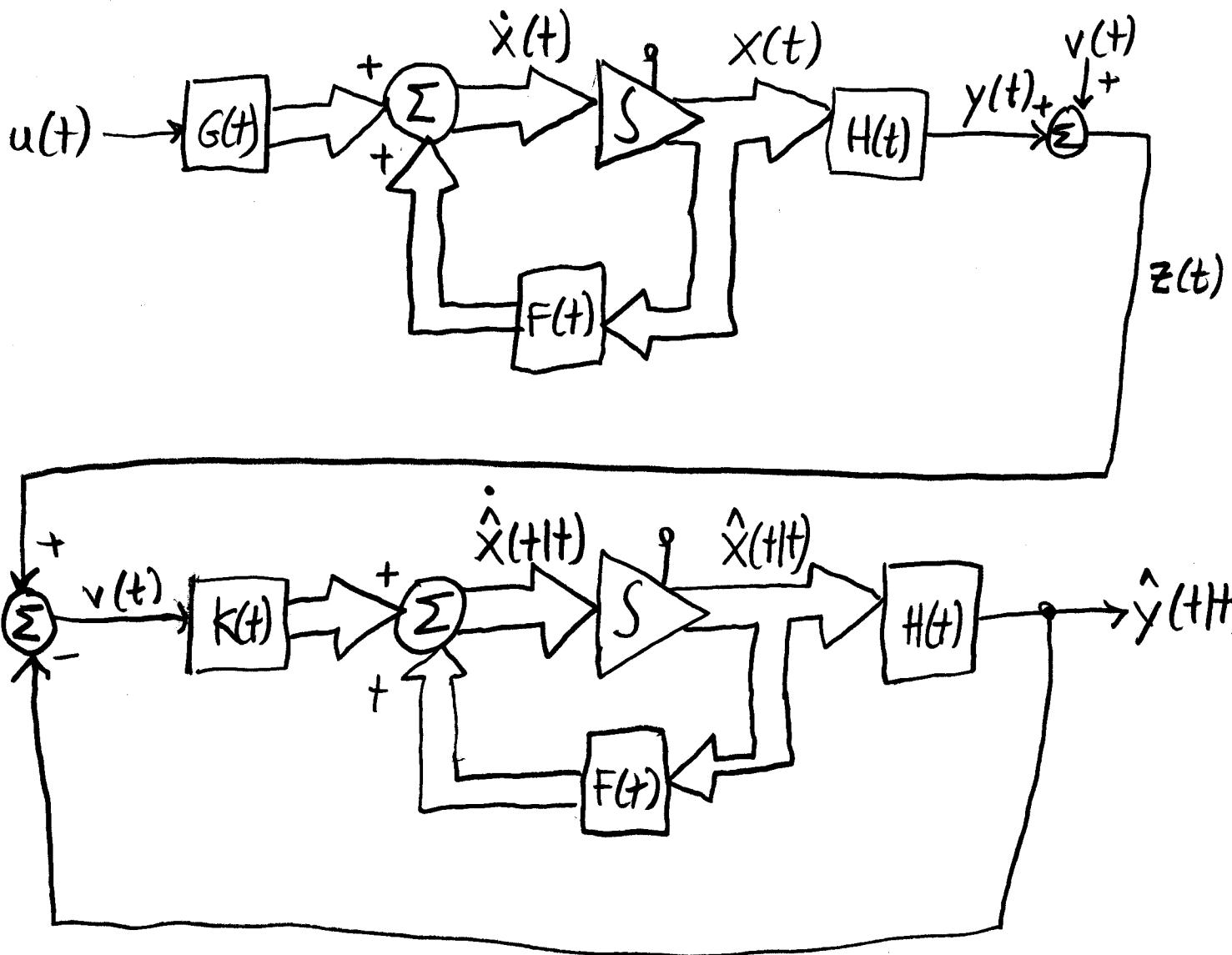
Let $K(t) = E[x(t)v(t)]R^{-1}(t)$. Then

$$\begin{cases} \dot{\hat{x}}(t|t) = F(t)\hat{x}(t|t) + K(t)v(t) & (*) \\ \hat{y}(t|t) = H(t)\hat{x}(t|t) & (**) \end{cases}$$

This is the continuous-time Kalman filter,
or "Kalman-Bucy" filter.

$$\begin{aligned}
 \text{Recall from 7.10: } v(t) &= z(t) - \hat{y}(t|t) \\
 &= z(t) - H(t)\hat{x}(t|t).
 \end{aligned}$$

→ If this is plugged into (*) above for $v(t)$,
it is the same as (7.1.24)



- Now we need to solve for $k(t)$.

$$K(t) \equiv E[x(t)v(t)]R^{-1}(t)$$

$$= E\{x(t)[H(t)\hat{x}(t|t) + v(t)]\} R^{-1}(t) \quad \text{see p-7.10}$$

$$= \underbrace{\{E[x(t)\hat{x}^T(t|t)H^T(t)] + E[x(t)v(t)]\}}_{\text{zero because } x(t) \text{ depends on } x(0) \text{ and } u(\theta) \text{ for } 0 \leq \theta < t, \text{ and these are uncorrelated with } v(t).} R^{-1}(t)$$

zero because $x(t)$ depends on $x(0)$ and $u(\theta)$ for $0 \leq \theta < t$, and these are uncorrelated with $v(t)$.

- So we have $K(t) = E[x(t)\tilde{x}^T(t|t)]H^T(t)R^{-1}(t)$. (*)
- But $\hat{X}(t|t)$ is uncorrelated with the error $\tilde{x}(t|t)$.
Also, $x(t) = \hat{X}(t|t) + \tilde{x}(t|t)$. So we have

$$\begin{aligned} E[x(t)\tilde{x}^T(t|t)] &= E\left\{\left[\hat{X}(t|t) + \tilde{x}(t|t)\right]\tilde{x}^T(t|t)\right\} \\ &= \underbrace{E\left[\hat{X}(t|t)\tilde{x}^T(t|t)\right]}_{\text{zero}} + E\left[\tilde{x}(t|t)\tilde{x}^T(t|t)\right] \\ &= E[\tilde{x}(t|t)\tilde{x}^T(t|t)] \equiv P(t|t). \end{aligned}$$

- Plugging this into (*) above, we have

$$K(t) = P(t|t)H^T(t)R^{-1}(t) \quad (7.1.16)$$

- Now we need to solve for $P(t|t)$ subject to the initial condition

$$P(0|0) = \Pi_0 = E[x(0)x^T(0)].$$

→ This is done by solving the famous "matrix Riccati equation".

$$\begin{aligned}
P(t|t) &= E[\tilde{x}(t|t)\tilde{x}^T(t|t)] \\
&= E\{[x(t) - \hat{x}(t|t)][x(t) - \hat{x}(t|t)]^T\} \\
&= E\{[x(t) - \hat{x}(t|t)][x^T(t) - \hat{x}^T(t|t)]\} \\
&= E[x(t)x^T(t) - x(t)\hat{x}^T(t|t) - \hat{x}(t|t)x^T(t) \\
&\quad + \hat{x}(t|t)\hat{x}^T(t|t)] \\
&= E[x(t)x^T(t)] - E[x(t)\hat{x}^T(t|t)] \\
&\quad - \underbrace{E[\hat{x}(t|t)x^T(t)]}_{\text{plug in } x(t) = \tilde{x}(t|t) + \hat{x}(t|t) \text{ and}} + E[\tilde{x}(t|t)\hat{x}^T(t|t)] \\
&\quad \text{observe that } E[\tilde{x}(t|t)\hat{x}^T(t|t)] = 0. \\
&= E[x(t)x^T(t)] - E[\hat{x}(t|t)\hat{x}^T(t|t)] \\
&\triangleq \pi(t) - \Sigma(t|t) \tag{*}
\end{aligned}$$

→ To find a differential equation for $P(t|t)$,
 ⇒ find separate differential equations for
 $\pi(t)$ and $\Sigma(t|t)$.

- Differential equation for $\Pi(t)$:

→ From page 7.1:

$$X(t) = \Phi(t, 0) X(0) + \int_0^t \Phi(t, \theta) G(\theta) u(\theta) d\theta$$

So,

$$\Pi(t) = E[X(t) X^T(t)]$$

$$= E \left[\left\{ \Phi(t, 0) X(0) + \int_0^t \Phi(t, \alpha) G(\alpha) u(\alpha) d\alpha \right\} \right]$$

$$\times \left\{ \Phi(t, 0) X(0) + \int_0^t \Phi(t, \beta) G(\beta) u(\beta) d\beta \right\}^T \right]$$

$$= E \left[\left\{ \Phi(t, 0) X(0) + \int_0^t \Phi(t, \alpha) G(\alpha) u(\alpha) d\alpha \right\} \right]$$

$$\times \left\{ X^T(0) \Phi^T(t, 0) + \int_0^t u(\beta) G^T(\beta) \Phi^T(t, \beta) d\beta \right\} \right]$$

$$= \Phi(t, 0) E[X(0) X^T(0)] \Phi^T(t, 0)$$

$$+ \Phi(t, 0) \int_0^t \underbrace{E[X(\alpha) u(\beta)]}_{\text{zero}} G^T(\beta) \Phi^T(t, \beta) d\beta$$

$$+ \int_0^t \Phi(t, \alpha) G(\alpha) \underbrace{E[u(\alpha) X^T(0)]}_{\text{zero}} d\alpha \Phi^T(t, 0)$$

$$+ \int_0^t \int_0^t \Phi(t, \alpha) G(\alpha) E[u(\alpha) u(\beta)] G^T(\beta) \Phi^T(t, \beta) d\beta d\alpha$$

$$\dots \pi(t) = \bar{\Phi}(t,0) \pi_0 \bar{\Phi}^T(t,0) + \int_0^t \bar{\Phi}(t,\alpha) G(\alpha) \left[\int_0^\alpha Q(\alpha) \delta(\alpha-\beta) G^T(\beta) \bar{\Phi}^T(t,\beta) d\beta \right] d\alpha$$

$$= \bar{\Phi}(t,0) \pi_0 \bar{\Phi}^T(t,0) + \int_0^t \bar{\Phi}(t,\alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t,\alpha) d\alpha \quad (\star)$$

→ Use Leibnitz rule to differentiate this with respect to t .

→ Note from top of page 7.3: $\frac{d}{dt} \bar{\Phi}(t,0) = F(t) \bar{\Phi}(t,0)$

$$\dot{\pi}(t) = F(t) \bar{\Phi}(t,0) \pi_0 \bar{\Phi}^T(t,0) + \bar{\Phi}(t,0) \pi_0 \bar{\Phi}^T(t,0) F^T(t)$$

$$+ \int_0^t \left\{ F(t) \bar{\Phi}(t,\alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t,\alpha) \right.$$

$$+ \bar{\Phi}(t,\alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t,\alpha) F^T(t) \left. \right\} d\alpha$$

$$+ \underbrace{\bar{\Phi}(t,t)}_{I} G(t) Q(t) G^T(t) \underbrace{\bar{\Phi}^T(t,t)}_{I}$$

$$= F(t) \bar{\Phi}(t,0) \pi_0 \bar{\Phi}^T(t,0) + \bar{\Phi}(t,0) \pi_0 \bar{\Phi}^T(t,0) F^T(t)$$

$$+ F(t) \int_0^t \bar{\Phi}(t,\alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t,\alpha) d\alpha$$

$$+ \int_0^t \bar{\Phi}(t,\alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t,\alpha) d\alpha F^T(t)$$

$$+ G(t) Q(t) G^T(t)$$

$$\begin{aligned}\dot{\Pi}(t) &= \dots = \text{collect terms w/ } F(t) \text{ and w/ } F^T(t) \\ &= F(t) \left[\bar{\Phi}(t, 0) \Pi_0 \bar{\Phi}^T(t, 0) + \int_0^t \bar{\Phi}(t, \alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t, \alpha) d\alpha \right] \\ &\quad + \left[\bar{\Phi}(t, 0) \Pi_0 \bar{\Phi}^T(t, 0) + \int_0^t \bar{\Phi}(t, \alpha) G(\alpha) Q(\alpha) G^T(\alpha) \bar{\Phi}^T(t, \alpha) d\alpha \right] F^T(t) \\ &\quad + G(t) Q(t) G^T(t)\end{aligned}$$

→ but, from ~~(***)~~ on p. 7.20, the bracketed quantities are each equal to $\Pi(t)$.

→ Substituting this, we have

$$\dot{\Pi}(t) = F(t) \Pi(t) + \Pi(t) F^T(t) + G(t) Q(t) G^T(t) \quad (*)$$

- Now, comparing the system in the top half of the figure on p. 7.16 w/ the filter in the bottom half, the differences are:

	System	Filter
input	$u(t)$	$v(t)$
input covariance structure	$Q(t)$	$R(t)$
input gain	$G(t)$	$K(t)$
State Vector	$X(t)$	$\hat{X}(t t)$

-Plugging these changes into the differential equation we just developed for $\dot{\pi}(t)$ transforms it into the equation for $\dot{\Sigma}(t|t)$:

$$\dot{\Sigma}(t|t) = F(t)\Sigma(t|t) + \Sigma(t|t)F^T(t) + K(t)R(t)K^T(t) \quad (***)$$

-Plugging (*) on p. 7.21 and (**) above into (*) on p. 7.18,

$$\begin{aligned} \dot{P}(t|t) &= \dot{\pi}(t|t) - \dot{\Sigma}(t|t) && \left\{ \begin{array}{l} \text{differentiate} \\ (*) \text{ on 7.18} \end{array} \right. \\ &= F(t)\pi(t) + \pi(t)F^T(t) + G(t)Q(t)G^T(t) \\ &\quad - F(t)\Sigma(t|t) - \Sigma(t|t)F^T(t) - K(t)R(t)K^T(t) \\ &= F(t)[\pi(t) - \Sigma(t|t)] + [\pi(t) - \Sigma(t|t)]F^T(t) \\ &\quad + G(t)Q(t)G^T(t) - K(t)R(t)K^T(t) \\ &= F(t)P(t|t) + P(t|t)F^T(t) + G(t)Q(t)G^T(t) \\ &\quad - K(t)R(t)K^T(t) \end{aligned}$$

- From (7.1.16) on p. 7.17, we have

$$K(t) = P(t|t) H^T(t) R^{-1}(t)$$

$$K^T(t) = \underbrace{R^{-1}(t)}^{\text{symmetric}} H(t) \underbrace{P^T(t|t)}_{\text{symmetric}} = R^{-1}(t) H(t) P(t|t)$$

- Plugging this into the last result on p. 7.22,

$$\dot{P}(t|t) = F(t)P(t|t) + P(t|t)F^T(t)$$

$$+ G(t)Q(t)G^T(t)$$

$$- P(t|t)H^T(t)R^{-1}(t)R(t)R^{-1}(t)H(t)P(t|t)$$

- or

$$\dot{P}(t|t) = F(t)P(t|t) + P(t|t)F^T(t) + G(t)Q(t)G^T(t)$$

$$- P(t|t)H^T(t)R^{-1}(t)H(t)P(t|t) \quad (7.2.1)$$

This is the "Matrix Riccati Equation" that must be solved for $P(t|t)$.

→ NOTE: it is peculiar and unexpected that the error covariance matrix diverges $\rightarrow \infty$ if the input noise vanishes.

- Before solving the Riccati equation, let's address continuous-time prediction and summarize the filtering/prediction equations.

- From (k) on p. 7.13, we obtain

$$\hat{X}(t+\Delta|t) = \int_0^t E[X(t+\Delta)Z(\tau)] R^{-1}(\tau) Z(\tau) d\tau \quad (*)$$

- From p. 7.1, we have

$$X(t+\Delta) = \Phi(t+\Delta, t) X(t) + \int_t^{t+\Delta} \Phi(t+\Delta, \alpha) G(\alpha) u(\alpha) d\alpha \quad (**)$$

- Post multiplying (**) by $Z(\tau)$ and taking the expected value, we have

$$E[X(t+\Delta)Z(\tau)] = \Phi(t+\Delta, t) E[X(t)Z(\tau)]$$

$$+ \int_t^{t+\Delta} \Phi(t+\Delta, \alpha) G(\alpha) E[u(\alpha)Z(\tau)] d\alpha$$

where $0 \leq \tau \leq t$.

\Rightarrow For $t \leq \alpha \leq t+\Delta$, $u(\alpha)$ and $Z(\tau)$ are uncorrelated -

$$- \text{So } E[X(t+\Delta)u(\tau)] = \Phi(t+\Delta, t) E[X(t)u(\tau)].$$

-Plugging this into (*) on p 7.24, we have

$$\hat{X}(t+\Delta|t) = \int_0^t \Phi(t+\Delta, \tau) E[X(\tau)u(\tau)] R^{-1}(\tau) u(\tau) d\tau$$

$$= \Phi(t+\Delta, t) \underbrace{\int_0^t E[X(\tau)u(\tau)] R^{-1}(\tau) u(\tau) d\tau}_{}$$

from (*) on p. 7.13, this
is $\hat{X}(t|t)$

$$\Rightarrow \hat{X}(t+\Delta|t) = \Phi(t+\Delta, t) \hat{X}(t|t) \quad (*)$$

-It follows immediately that

$$\hat{y}(t+\Delta|t) = H(t+\Delta) \Phi(t+\Delta, t) \hat{X}(t|t) \quad (**)$$

Summary : Continuous-time Kalman Filter

innovations: $\nu(t) = z(t) - \hat{y}(t|t)$
 $= H(t) \hat{x}(t|t) + v(t)$

$$E[\nu(t)\nu(t)] = R(t) \delta(t-t)$$

Filter: $\begin{cases} \dot{\hat{x}}(t|t) = F(t) \hat{x}(t|t) + K(t) \nu(t) \\ \hat{y}(t|t) = H(t) \hat{x}(t|t) \end{cases}$

Gain: $K(t) = P(t|t) H^T(t) R^{-1}(t)$

state Vector
 Error Covariance : $\dot{P}(t|t) = F(t)P(t|t) + P(t|t)F^T(t)$
 Matrix X (Riccati Eq)
 $+ G(t)Q(t)G^T(t)$
 $- P(t|t)H^T(t)R^{-1}(t)H(t)P(t|t)$

Subject to $P(0|0) = P_0$

predictor: $\begin{cases} \hat{x}(t+\Delta|t) = \Phi(t+\Delta, t) \hat{x}(t|t) \\ \hat{y}(t+\Delta|t) = H(t+\Delta) \hat{x}(t+\Delta|t) \end{cases}$

MATRIX RICCATI EQUATIONS

- The matrix Riccati equation is a nonlinear differential equation.
- The key to solving it is a tricky manipulation that converts it into a pair of linear differential equations that can be solved using routine methods.
- We will need a formula for the derivative of the inverse of a matrix.
→ Let B be a square nonsingular matrix. Then

$$BB^{-1} = I \quad (\#)$$

→ Differentiate both sides:

$$\frac{d}{dt} I = 0 \quad (\text{**})$$

$$\frac{d}{dt} BB^{-1} = \left(\frac{d}{dt} B \right) B^{-1} + B \left(\frac{d}{dt} B^{-1} \right) \quad (\text{****})$$

- Plug (**) and (****) into $(\#)$:

$$0 = \left(\frac{d}{dt} B \right) B^{-1} + B \left(\frac{d}{dt} B^{-1} \right)$$

$$B \left(\frac{d}{dt} B^{-1} \right) = - \left(\frac{d}{dt} B \right) B^{-1} \quad \rightarrow$$

$$\boxed{\frac{d}{dt} B^{-1} = -B^{-1} \left(\frac{d}{dt} B \right) B^{-1}}$$

STRATEGY:

1) Factor the Riccati equation to show that $P(t|t)$ may be written as

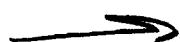
$$P(t|t) = A(t) B^{-1}(t)$$

2) Obtain a pair of simultaneous linear differential equations for $A(t)$ and $B(t)$.

- The Riccati equation (from p. 7.23) is

$$\dot{P}(t|t) = F(t)P(t|t) + P(t|t)F^T(t) + G(t)Q(t)G^T(t) \\ - P(t|t)H^T(t)R^{-1}(t)H(t)P(t|t)$$

- Let us suppose that $P(t|t) = A(t)B^{-1}(t)$ and see if this leads to solutions for $A(t)$ and $B(t)$.



$$\dot{P}(t|t) = \frac{d}{dt} \{ A(t) B^{-1}(t) \}$$

$$= \dot{A}(t) B^{-1}(t) - A(t) B^{-1}(t) \dot{B}(t) B^{-1}(t)$$

- Plug this into the Riccati equation:

$$\dot{A}(t) B^{-1}(t) - A(t) B^{-1}(t) \dot{B}(t) B^{-1}(t) = F(t) P(t|t) + P(t|t) F^T(t)$$

$$+ G(t) Q(t) G^T(t) - P(t|t) H^T(t) R^{-1}(t) H(t) P(t|t)$$

$$= F(t) A(t) B^{-1}(t) + A(t) B^{-1}(t) F^T(t)$$

$$- A(t) B^{-1}(t) H^T(t) R^{-1}(t) H(t) A(t) B^{-1}(t) + G(t) Q(t) G^T(t)$$

- Post multiply both sides by $B(t)$:

$$\dot{A}(t) - A(t) B^{-1}(t) \dot{B}(t) = F(t) A(t) + A(t) B^{-1}(t) F^T(t) B(t)$$

$$- A(t) B^{-1}(t) H^T(t) R^{-1}(t) H(t) A(t) + G(t) Q(t) G^T(t) B(t)$$

- Factor:

$$\dot{A}(t) - \underbrace{\{A(t) B^{-1}(t)\}}_{\text{Factor}} \dot{B}(t) = F(t) A(t) + \underbrace{G(t) Q(t) G^T(t)}_{\text{Factor}} B(t)$$

$$- \underbrace{\{A(t) B^{-1}(t)\}}_{\text{Factor}} \left[\underbrace{H^T(t) R^{-1}(t) H(t) A(t)}_{\text{Factor}} - \underbrace{F^T(t) B(t)}_{\text{Factor}} \right]$$



- This is equivalent to two simultaneous linear equations:

$$\dot{A}(t) = F(t)A(t) + G(t)Q(t)G^T(t)B(t) \quad (*)$$

$$\dot{B}(t) = H^T(t)R^{-1}(t)H(t)A(t) - F(t)B(t)$$

- The decomposition of $P(t|t)$ into $A(t)$ and $B^{-1}(t)$ is not unique, and neither are the solutions $A(t)$, $B(t)$ to (*).

- By imposing some reasonable initial conditions on (*), we can formulate a problem that will be uniquely solvable for $A(t)$ and $B(t)$, and will therefore give us a solution for $P(t|t)$ according to $P(t|t) = A(t)B^{-1}(t)$.

- Let us require that $B(0) = I$.

→ We have then that

$$P(0|0) = A(0)B^{-1}(0) = A(0)I = A(0).$$

→ Then $A(0) = T C_0$.

- (k) on p- 7.30 can be written in matrix form together with the initial conditions:

$$\frac{d}{dt} \begin{bmatrix} A(t) \\ B(t) \end{bmatrix} = \begin{bmatrix} F(t) & G(t)Q(t)G^T(t) \\ H^T(t)R^{-1}(t)H(t) & -F(t) \end{bmatrix} \begin{bmatrix} A(t) \\ B(t) \end{bmatrix}$$

Subject to

$$\begin{bmatrix} A(0) \\ B(0) \end{bmatrix} = \begin{bmatrix} T_0 \\ I \end{bmatrix}.$$

⇒ This pair of simultaneous linear equations can now be solved by standard techniques.