Lecture 1(b)

Kalman-Bucy filtering and duality (continuation)

Associated to the stochastic system

$$
\begin{aligned}
dx &= Ax\,dt + dv \\
dy &= Cx\,dt + de
\end{aligned}
\qquad (S)
$$

under suitable (hypotheses) assumptions about the driving noise and measurement noise, one derives a new, also linear stochastic dynamical system, known as the <u>Kalman-Bucy filter</u>,

$$
\begin{aligned}
d\hat{x} &= A\hat{x}\,dt + K\,(dy - C\hat{x}\,dt) \\
&= (A - KC)\hat{x}\,dt + K\,dy
\end{aligned}
\qquad (KB)
$$

driven by the measurement process; the coefficients A, C, and K are all time dependent; the initial condition

$$\hat{x}(t_0) = m_0$$

the gain

$$K = PC^T R_2^{-1}$$

where

$$
\begin{aligned}
\dot{P} &= AP + PA^T + R_1 - PC^T R_2^{-1} CP \\
P(t_0) &= R_0 \qquad\qquad t \geq t_0
\end{aligned}
$$

R_1 and R_2 are parameters respectively, associated with the processes v and e. The parameters m_0 and R_0 are associated to the initial condition $x(t_0)$ of the state process: $m_0 \triangleq E(x(t_0))$; $R_0 \triangleq Var(x(t_0))$

In the above formulas the superscript 'T' denotes matrix transpose.

We proceed by posing a problem that KB solves under <u>suitable hypotheses</u>. Along
(1) the way the meaning of the process $\hat{x}$ in relation to the process x becomes clear. The problem is one of <u>optimization</u>.
(2) By suitable <u>auxiliary system(s)</u>, the <u>cost</u> functional to be optimized can be expressed in convenient (quadratic) form.
(3) Appropriate hypotheses about the processes v and e make this representation feasible.
(4) The optimization problem is then solved by a dual, deterministic optimal control problem.
(5) The answer can be expressed in recursive form - the Kalman-Bucy filter. We will carry out all these steps in the order stated.

(1) Optimization Problem [Formulation]

Determine a process $\hat{x}(t)$ such that for any constant vector 'a' and a suitable constant 'b',

$$a^T \hat{x}(t_1) = -\int_{t_0}^{t_1} u^T(\sigma)\, dy(\sigma) + b^T m_0$$

is the minimum mean square error (MMSE) estimate of $a^T x(t)$, $t_1 > t_0$ arbitrary. Thus one has to decide

$$b, \quad u(t) \quad t_0 \le t \le t_1$$

to minimize

$$\text{COST} = E\left[\left(a^T x(t_1) - a^T \hat{x}(t_1)\right)^2\right]$$

Note that we have already restricted the form of the answer $a^T \hat{x}(t_1)$ to be linear in the observations y through an integral and the cost is quadratic.. This is made more explicit below

(2) Representation using dual system (deterministic)

We calculate :

$$a^T \hat{x}(t_1) = -\int_{t_0}^{t_1} u^T(\sigma) C(\sigma) x(\sigma) d\sigma - \int_{t_0}^{t_1} u^T(\sigma)\, de(\sigma) + b^T m_0$$

On the other hand, for <u>any</u> differentiable, deterministic function $z(\cdot)$ with $z(t_1) = a$,

$$a^T x(t_1) = z^T(t_1)\, x(t_1)$$

$$= z^T(t_0)\, x(t_0) + \int_{t_0}^{t_1} d\left(z^T(\sigma)\, x(\sigma)\right)$$

(fundamental theorem of calculus)

$$= z^T(t_0)\, x(t_0) + \int_{t_0}^{t_1} \dot{z}^T(\sigma)\, x(\sigma)\, d\sigma$$

$$+ \int_{t_0}^{t_1} z^T(\sigma)\, dx(\sigma)$$

Suppose z is defined by the <u>dual</u> system

$\boxed{\text{DUALSYS}}$ $$\frac{dz}{dt} = -A^T z - C^T u; \qquad z(t_1) = a$$

evolving backword from t_1.

Then

$$a^T x(t_1) = z^T(t_0)\, x(t_0) + \int_{t_0}^{t_1} \left(-z^T(\sigma) A(\sigma) x(\sigma) - u^T(\sigma) C(\sigma) x(\sigma)\right) d\sigma$$

$$+ \int_{t_0}^{t_1} \left(z^T(\sigma) A(\sigma) X(\sigma)\, d\sigma + z^T(\sigma)\, dV(\sigma)\right)$$

$$= Z^T(t_0)\,x(t_0) \quad + \int_{t_0}^{t_1} \left(-u^T(\sigma)C(\sigma)x(\sigma)\,d\sigma + Z^T(\sigma)\,dv(\sigma)\right)$$

We now compute the <u>error</u>:

$\cancel{a^T(x(t_1) - x^T(x}$

$$a^T\left(x(t_1) - \hat{x}(t_1)\right) \quad = \quad \left(Z^T(t_0)\,x(t_0) - b^T m_0\right)$$

$$+ \int_{t_0}^{t_1} \left[Z^T(\sigma)\,dv(\sigma) + u^T(\sigma)\,de(\sigma)\right]$$

The error is linear in the processes v and e. <u>Suppose we now make</u>

(3)

* [HYPOTHESIS 1] $\qquad E(v(t)) = E(e(t)) \equiv 0;$

then the expectations of <u>increments</u> in v and e will be also <u>vanishing</u>.

Hence

$$E\left(a^T(x(t_1) - \hat{x}(t_1))\right) = E\left(Z^T(t_0)\,x(t_0) - b^T m_0\right)$$

$$= \left(Z(t_0) - b\right)^T m_0$$

Since

$$E\left(a^T(x(t_1) - \hat{x}(t_1))\right)^2$$

$$= \operatorname{Var}\left(a^T(x(t_1) - \hat{x}(t_1))\right) + \left(E\left(a^T(x(t_1) - \hat{x}(t_1))\right)\right)^2,$$

one reduces the optimization problem to that of minimizing a variance if we set:

* [CHOICE 1] $\qquad b = z(t_0)$

Since $z(t_0)$ depends on the solution to the differential equation for x driven by u, it follows that u dependence is encoded in b. One needs to compute the variance of $a^T(x(t) - \hat{x}(t_1))$. suppose

* [Hypothesis 2]

$v(t)$ and $e(t)$ are <u>orthogonal increment</u> processes, i.e. for $t_1 < t_2 < t_3 < t_4$

$$(v(t_2) - v(t_1)) \perp (v(t_4) - v(t_3))$$
$$(e(t_2) - e(t_1)) \perp (e(t_4) - e(t_3))$$

(where the symbol '$\perp$' stands for
$X \perp Y$ if $E(XY^T) = 0$
X is uncorrelated to Y $\qquad$ matrix)

<u>and</u> $\qquad v(t) \perp e(s) \qquad \forall\, t, s$

<u>and</u> $\qquad v(t) \perp x(t_0) \qquad \forall\, t \geq t_0$

and

$$E\,(v(t_2)-v(t_1))(v(t_2)-v(t_1))^T$$

$$= \int_{t_1}^{t_2} R_1(\sigma)\,d\sigma \qquad t_2 > t_1$$

$$E\,(e(t_2)-e(t_1))(e(t_2)-e(t_1))^T$$

$$= \int_{t_1}^{t_2} R_2(\sigma)\,d\sigma \qquad t_2 > t_1$$

with $R_1(\sigma) = R_1^T(\sigma) \geq 0$ and

$R_2(\sigma) = R_2^T(\sigma) > 0$,

Under these conditions the incremental covariances satisfy

$$E\,(dv\,dv^T) = R_1\,dt$$

$$E\,(de\,de^T) = R_2\,dt$$

highlighting the 'impulsive character' of any notion of $\dot{v}$ and $\dot{e}$ – a good reason to avoid these derivatives.

With HYPOTHESES 1 and 2 and CHOICE 1, the COST

$$\eta = E\left(a^T(x(t) - \hat{x}(t_1))\right)^2$$

$$= Var\left(a^T(x(t) - \hat{x}(t_1))\right)$$

$$(3) \quad \left\{ = z^T(t_0)\, R_0\, z(t_0) + \int_{t_0}^{t_1} z^T(\sigma)\cdot R_1(\sigma)\cdot z(\sigma)\, d\sigma + \int_{t_0}^{t_1} u^T(\sigma)\cdot R_2(\sigma)\cdot u(\sigma)\, d\sigma \right.$$

to be minimized, subject to the deterministic dynamics

$$\left\{ \begin{aligned} \dot{z} &= -A^T z - C^T u \qquad 0 < t < t_1 \\ z(t_1) &= a \end{aligned} \right.$$

4. Dual Optimal Control Problem

The above dual optimal control problem is very much what one encounters in ENEE 664 OPTIMAL CONTROL, except for the fact that here the dynamics runs backward in time and the cost to be optimized contains an initial cost

$z^T(t_0)\, R_0\, z(t_0)$ added to path cost and control cost. (In 664 we speak of terminal cost.)

This is solved by completion of squares.

step (i) Observe, under the deterministic dynamics of z, for any differentiable matrix valued function, $P(t) = P^T(t)$

$$z^T(t_1) P(t_1) z(t_1) - z^T(t_0) P(t_0) z(t_0)$$

$$= \int_{t_0}^{t_1} \frac{d}{dt}\left(z^T(t) P(t) z(t) \right) dt$$

$$= \int_{t_0}^{t} \left(u^T(t) \;\; z^T(t) \right) \begin{pmatrix} 0 & -C(t) P(t) \\ -P(t) C^T(t) & \dot{P}(t) - A(t) P(t) - P(t) A^T(t) \end{pmatrix} \begin{pmatrix} u(t) \\ z(t) \end{pmatrix} dt$$

step (ii) From $\eta = \eta + 0$ and step (i), we get

$$\eta = z^T(t_0) R_0 z(t_0) + \int_{t_0}^{t_1} \left(u^T(\sigma) \;\; z^T(\sigma) \right) \begin{pmatrix} R_2(\sigma) & 0 \\ 0 & R_1(\sigma) \end{pmatrix} \begin{pmatrix} u(\sigma) \\ z(\sigma) \end{pmatrix} d\sigma$$

$$+ z^T(t_1) P(t_1) z(t_1) - z^T(t_0) P(t_0) z(t_0) +$$

$$- \int_{t_0}^{t_1} \left(u^T(t) \;\; z^T(t) \right) \begin{pmatrix} 0 & -C(t)P(t) \\ -P^T(t)\,C^T(t) & \dot{P}(t) - A(t)P(t) - P(t)A^T(t) \end{pmatrix} \begin{pmatrix} u(t) \\ z(t) \end{pmatrix} dt$$

$$= z^T(t_1)\, P(t_1)\, z(t_1) + \int_{t_0}^{t_1} \left(u^T(\sigma) \;\; z^T(\sigma) \right) \begin{pmatrix} R_2(\sigma) & C(\sigma)P(\sigma) \\ P(\sigma)C^T(\sigma) & P(\sigma)C^T(\sigma)R_2^{-1}(\sigma)C(\sigma)P(\sigma) \end{pmatrix} \begin{pmatrix} u(\sigma) \\ z(\sigma) \end{pmatrix} d\sigma$$

provided we set $P = P^T$ to be solution to

$$\begin{cases} \dot{P} = AP + PA^T + R_1 - PC^T R_2^{-1} C P \\ P(t_0) = R_0 \end{cases} \qquad \begin{pmatrix}\text{Riccati} \\ \text{Equation}\end{pmatrix}$$

on the entire interval $[t_0, t_1]$.

$$\Rightarrow \eta = z^T(t_1)\, P(t_1)\, z(t_1) + \int_{t_0}^{t_1} \left(u(\sigma) + R_2^{-1}(\sigma)\, C(\sigma) P(\sigma) z(\sigma) \right)^T R_2(\sigma) \cdot \left(u(\sigma) + R_2^{-1}(\sigma)\, C(\sigma) P(\sigma) z(\sigma) \right) d\sigma$$

which is minimized when

$$u(t) = -R_2^{-1}(t)\, C(t)\, P(t)\, z(t)$$

$$t_0 \le t \le t_1$$

since $z(t_1) = a$ fixed, and $P(t_1)$ does not depend on $u(\cdot)$.

This clean minimization is achieved by the use of step (i) to turn the integrand of the integral term in η to be a <u>perfect square</u>, when the Riccati equation holds, on the entire interval $[t_0, t_1]$.

Summarizing, the <u>linear</u> minimum mean square error estimate of $a^T x(t_1)$ is given by

$$a^T \hat{x}(t_1) = -\int_{t_0}^{t_1} u^T(\sigma)\, dy(\sigma) + b^T m_0$$

where,

$$u(t) = -R_2^{-1}(t)\, C(t)\, P(t)\, z(t)$$
$$\dot{z}(t) = -A^T(t)\, z(t) - C^T(t)\, u(t)$$
$$z(t_1) = a$$
$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) + R_1(t) - P(t)\, C^T(t)\, R_2^{-1}(t)\, C(t)\, P(t)$$
$$P(t_0) = R_0 \quad \text{and} \quad b = z(t_0)$$

Let $\dot{z} = -A^T z - C^T(-R_2^{-1} C P z)$

$$= (-A^T + C^T K^T)\, z$$

where $K \triangleq P C^T R_2^{-1}$ (Kalman gain)

$\Rightarrow \quad \dot{z} = -(A - KC)^T z$.

Let $\Psi(t, t_1)$ be solution to the matrix equation

$$\frac{d\Psi}{dt} = (A - KC)\,\Psi$$

with $\Psi(t_1, t_1) = \mathbb{1}$ the identity matrix. Then

$$z(t) = \Psi^T(t_1, t)\, z(t_1)$$

check: $\Psi^T(t_1, t)\, z(t_1) = \left(\Psi^{T^{-1}}(t, t_1)\right) z(t_1)$

$\Rightarrow \frac{d}{dt}\left(\Psi^{T^{-1}}(t, t_1)\right) z(t_1)$

$$= -\Psi^{T^{-1}} \frac{d}{dt}\Psi^T(t, t_1)\, \Psi^{T^{-1}}(t, t_1)\, z(t_1)$$

$= -\Psi^{T^{-1}}\, \Psi^T(t, t_1)\, (A - KC)^T\, \Psi^{T^{-1}}(t, t_1)\, z(t_1)$

$= -(A - KC)^T\, \Psi^{T^{-1}}(t, t_1)\, z(t_1)$

$$= \Psi^T(t_1, t)\, z(t_1)$$

$\Rightarrow \quad \Psi^{T^{-1}}(t, t_1)\, z(t_1)$ satisfies the

ode $\quad \dot{z} = -(A - KC)^T z \quad \square$

Then $\quad u(t) = -K^T(t)\, z(t)$

$$= -K^T \Psi^T(t_1, t)\, a$$

and $\quad b = z(t_0) = \Psi^T(t_1, t_0)\, a$

for any a, optimal linear estimate

Then $\quad a^T \hat{x}(t_1) = -\int_{t_0}^{t_1} u^T(\sigma)\, dy(\sigma) + b^T m_0$

$$= a^T \int_{t_0}^{t_1} \Psi(t_1, t)\, K(t)\, dy(t) + a^T \Psi(t_1, t_0) m_0$$

This equation is satisfied (for all a) if we define

$$\hat{x}(t_1) = \int_{t_0}^{t_1} \Psi(t_1, t)\, K(t)\, dy(t) + \Psi(t_1, t_0) m_0$$

Clearly

$$d\hat{x}(t_1) = \Psi(t_1, t_1)\, K(t_1)\, dy(t_1)$$

$$+ \int_{t_0}^{t} \frac{\partial \Psi}{\partial t_1}(t_1, t)\, K(t)\, dy(t)\, dt_1$$

$$+ \frac{\partial \Psi}{\partial t_1}(t_1, t_0)\, m_0\, dt_1$$

$$= K(t_1)\,dy(t_1)$$

$$+ \int_{t_0}^{t_1} \Big(A(t_1) - K(t_1)C(t_1)\Big)\,\Psi(t_1,t)\,K(t)\,dy(t)\,dt_1$$

$$+ \big(A(t_1) - K(t_1)C(t_1)\big)\,\Psi(t_1,t_0)\,m_0\,dt_1$$

$$= K(t_1)\,dy(t_1) + \big(A(t_1) - K(t_1)C(t_1)\big)\,\hat{x}(t_1)\,dt$$

$$= A(t_1)\,\hat{x}(t_1)\,dt + K(t_1)\big(dy(t_1) - C(t_1)\hat{x}(t_1)\,dt\big)$$

This is true for any $t_1 > t_0$

$$\Rightarrow \quad d\hat{x}(t) = A(t)\hat{x}(t)\,dt + K(t)\big(dy(t) - C\hat{x}(t)\,dt\big)$$

Clearly $\hat{x}(t_0) = m_0$.

Thus we have proved the following

<u>Theorem</u> (KB) : Optimal (MMSE) linear estimate is computed recursively by

$$d\hat{x} = A\hat{x}\,dt + K\,(dy - C\hat{x}\,dt)$$
$$\hat{x}(t_0) = m_0$$
$$K = PC^T R_2^{-1}$$
$$\dot{P} = AP + PA^T + R_1 - PC^T R_2^{-1} CP$$
$$P(t_0) = R_0 \qquad \square$$