Continuous - Time Kalman Filter

The continuous-time system:

$$\dot{\underline{x}}(t) = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{w}(t)$$

$$\underline{z}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t)$$

For small Δt,

$$\frac{\underline{x}(t+\Delta t) - \underline{x}(t)}{\Delta t} \cong \dot{\underline{x}}(t) = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{w}(t)$$

$$\underline{z}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t)$$

$$E[\underline{w}(t)\underline{w}^T(t)] = \underline{Q}(t)/\Delta t$$

$$E[\underline{v}(t)\underline{v}^T(t)] = \underline{R}(t)/\Delta t$$

or,

$$\underline{x}(t+\Delta t) - \underline{x}(t) = \Delta t\,\underline{F}(t)\underline{x}(t) + \Delta t\,\underline{G}(t)\underline{w}(t)$$

and

$$\underline{\Phi}(t+\Delta t) = \underline{I} + \Delta t\,\underline{F}(t)$$

$$\underline{\Gamma}(t) = \Delta t\,\underline{G}(t)$$

We will derive the continuous time Kalman filter by looking at the discrete time version. Assume that the continuous time state and measurements are sampled every Δt seconds, and we will examine the behavior as $\Delta t \to 0$.

From the solution to the discrete-time problem,

$$\underline{\hat{x}}(t+\Delta t|t+\Delta t) = \underline{\Phi}(t+\Delta t, t)\,\underline{\hat{x}}(t|t)$$
$$+ \underline{K}(t+\Delta t)\left[\underline{z}(t+\Delta t) - \underline{H}(t+\Delta t)\,\underline{\Phi}(t+\Delta t, t)\,\underline{\hat{x}}(t|t)\right]$$

$$\underline{\Sigma}(t+\Delta t|t+\Delta t) = \left[\underline{I} - \underline{K}(t+\Delta t)\underline{H}(t+\Delta t)\right]\left[\underline{\Phi}(t)\,\underline{\Sigma}(t|t)\,\underline{\Phi}^T(t)\right.$$
$$\left. + \underline{G}(t)\underline{Q}(t)\underline{G}^T(t)\right]$$

where $\underline{\Phi}(t) = \underline{\Phi}(t+\Delta t, t)$.

The Kalman gain is

$$\underline{K}(t+\Delta t) = \underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t)\left[\underline{H}(t+\Delta t)\,\underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t)\right.$$
$$\left. + \underline{R}(t+\Delta t)\right]^{-1}$$

where

$$\underline{\Sigma}(t+\Delta t|t) = \underline{\Phi}(t+\Delta t, t)\,\underline{\Sigma}(t|t)\,\underline{\Phi}^T(t+\Delta t, t)$$
$$+ \underline{G}(t)\underline{Q}(t)\underline{G}^T(t)$$

From the solution to the discrete-time problem,

$$\hat{\underline{x}}(t+\Delta t | t+\Delta t) = \underline{\Phi}(t+\Delta t, t)\, \hat{\underline{x}}(t|t)$$

$$+ \underline{K}(t+\Delta t)\left[\underline{z}(t+\Delta t) - \underline{H}(t+\Delta t)\, \underline{\Phi}(t+\Delta t, t)\, \hat{\underline{x}}(t|t)\right]$$

$$\Sigma(t+\Delta t | t+\Delta t) = \left[I - \underline{K}(t+\Delta t)\, \underline{H}(t+\Delta t)\right]\left[\underline{\Phi}(t)\, \underline{\Sigma}(t|t)\, \underline{\Phi}^T(t) + \underline{G}(t)\, \underline{Q}(t)\, \underline{G}^T(t)\right]$$

and

$$K(t+\Delta t) = \Sigma(t+\Delta t | t)\, \underline{H}^T(t+\Delta t)\left[H(t+\Delta t)\, \underline{\Sigma}(t+\Delta t | t)\, \underline{H}^T(t+\Delta t) + \underline{R}(t+\Delta t)\right]^{-1}$$

and

$$\underline{\Sigma}(t+\Delta t | t) = \underline{\Phi}(t+\Delta t, t)\, \underline{\Sigma}(t|t)\, \underline{\Phi}^T(t+\Delta t | t) + \underline{G}(t)\, \underline{Q}(t)\, \underline{G}^T(t)$$

For small Δt,

$$\underline{\Sigma}(t+\Delta t | t) \cong (I + \Delta t\, F(t))\, \underline{\Sigma}(t|t)\, (I + \Delta t\, \underline{F}^T(t)) + \underline{G}(t)\, \underline{Q}(t)\, \underline{G}^T(t)\, \Delta t$$

$$= \Delta t\left[\underline{F}(t)\, \underline{\Sigma}(t|t) + \underline{\Sigma}(t|t)\, \underline{F}^T(t) + \underline{G}(t)\, \underline{Q}(t)\, \underline{G}^T(t)\right]$$

$$+ \underline{\Sigma}(t|t) + (\Delta t)^2\, \underline{F}(t)\, \underline{\Sigma}(t|t)\, \underline{F}^T(t)$$

For small Δt,

$$\underline{\Phi}(t+\Delta t, t) \cong \underline{I} + \Delta t\, \underline{F}(t)$$

From Chapter 3 of Gelb,

$$\underline{G}(t)\underline{Q}^*_d(t)\underline{G}^T(t) \longleftrightarrow \underline{G}(t)\underline{Q}_c(t)\underline{G}^T(t)\Delta t$$

and

$$\underline{R}_d(t) = \underline{R}_c(t)/\Delta t$$

where the subscripts "d" and "c" denote discrete and continuous time, respectively. The first ~~equation~~ relation is from the linear variance equation. The second is because ~~$\underline{w}_d(t)$ and $\underline{w}_c$~~ $\underline{v}_d(t)$ and $\underline{v}_c(t)$ ~~are~~ have impulsive auto-correlations. The effect of $\underline{v}_d(t)$ over the time interval $[t, t+\Delta t]$ is the cumulative effect of the cont's time white noise process $v_c(t)$ over the interval. The integral of this effect is the increments process (a random walk in cont's time) and has variance proportional to Δt — thus the need for the $1/\Delta t$ term.

Gelb explains this slightly differently; neither is entirely correct — the fundamental problem being that $\dot{\underline{x}}(t)$, and $\underline{w}(t)$ and $\underline{v}(t)$ don't physically (or mathematically) exist. The math just happens to work out correctly for <u>linear</u> systems (only).

In going to the limit as $\Delta t \to 0$, we need an equation for $\dot{\underline{\Sigma}}(t|t)$ and $\underline{K}(t)$. We have $\underline{\Sigma}(t+\Delta t|t+\Delta t)$ in terms of $\underline{\Sigma}(t+\Delta t|t)$, and $\underline{\Sigma}(t+\Delta t|t)$ in terms of $\underline{\Sigma}(t|t)$, so some algebra is required.

To condense some of the equations, subscripts (arguments) are dropped when they are t (or $t|t$).

Beginning with $\underline{\Sigma}(t+\Delta t|t+\Delta t)$, we need to write this as a function of $\underline{\Sigma}(t|t)$:

$$\underline{\Sigma}(t+\Delta t|t+\Delta t) = [\underline{I} - \underline{K}(t+\Delta t)\underline{H}(t+\Delta t)][\underline{\Phi}\underline{\Sigma}\underline{\Phi}^T + \underline{G}\underline{Q}\underline{G}^T\Delta t]$$

$$= [\underline{I} - \underline{\Sigma}(t+\Delta t|t)\underline{H}^T(t+\Delta t)[\underline{H}(t+\Delta t)\underline{\Sigma}(t+\Delta t|t)\underline{H}^T(t+\Delta t) + \underline{R}(t+\Delta t)/\Delta t]^{-1} \underline{H}(t+\Delta t)]\underline{\Sigma}(t+\Delta t|t)$$

For small Δt, it is safe to assume $\underline{H}(t+\Delta t) \cong \underline{H}(t)$ & $\underline{R}(t+\Delta t) \cong \underline{R}(t)$ and drop these arguments.

Substituting for $\underline{\Sigma}(t+\Delta t|t)$ (and dropping arguments):

$$\underline{\Sigma}(t+\Delta t|t+\Delta t) = \underline{\Phi}\underline{\Sigma}\underline{\Phi}^T + \underline{G}\underline{Q}\underline{G}^T\Delta t$$
$$- [\underline{\Phi}\underline{\Sigma}\underline{\Phi}^T + \underline{G}\underline{Q}\underline{G}^T\Delta t]\underline{H}^T[\underline{H}[\underline{\Phi}\underline{\Sigma}\underline{\Phi}^T + \underline{G}\underline{Q}\underline{G}^T\Delta t]\underline{H}^T + \underline{R}]^{-1}$$
$$\times \underline{H}[\underline{\Phi}\underline{\Sigma}\underline{\Phi}^T + \underline{G}\underline{Q}\underline{G}^T]$$

Using $\underline{\Phi} \to \underline{I} + \Delta t \underline{F}$ for small Δt, and associating like terms,

$$\underline{\Phi}\underline{\Sigma}\underline{\Phi}^T + \underline{G}\underline{Q}\underline{G}^T\Delta t \cong \underline{\Sigma} + \Delta t(\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T) + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^2$$

Now,

$$\frac{\underline{\Sigma}(t+\Delta t|t+\Delta t) - \underline{\Sigma}(t|t)}{\Delta t} \cong \underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T + \Delta t\, \underline{F}\underline{\Sigma}\underline{F}^T$$

$$- \frac{1}{\Delta t}\left[\underline{\Sigma} + \Delta t[\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T] + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^T\right]\underline{H}^T$$

$$\times \left[\underline{H}\left[\underline{\Sigma} + \Delta t(\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T) + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^T\right]\underline{H}^T + R/\Delta t\right]^{-1}$$

$$\times \underline{H}\left[\underline{\Sigma} + \Delta t(\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T) + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^T\right]$$

$$= \underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T + \Delta t\, \underline{F}\underline{\Sigma}\underline{F}^T$$

$$- \left[\underline{\Sigma} + \Delta t[\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T] + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^T\right]\underline{H}^T$$

$$\times \left[\Delta t\, \underline{H}\left[\underline{\Sigma} + \Delta t(\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T) + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^T\right]\underline{H}^T + \underline{R}\right]^{-1}$$

$$\times H\left[\underline{\Sigma} + \Delta t(\underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T) + (\Delta t)^2 \underline{F}\underline{\Sigma}\underline{F}^T\right]$$

Taking $\lim_{\Delta t \to 0}$:

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

where everything on the RHS is a function of t (or $t|t$).

This is the Riccati equation for the conts time optimal linear filter:

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

Its solution gives the evolution of the error covariance versus time.

Look at the components:

$F(t)\Sigma(t|t) + \Sigma(t|t)F^T(t)$ — influence of system dynamics

$G(t)Q(t)G^T(t)$ — influence of process noise (increases error)

$-\Sigma(t|t)H^T(t)R^{-1}(t)H(t)\Sigma(t|t)$ — influence of measurements (reduces error)

Note:

The greater $Q(t)$ is, the faster predictions (estimates w/o new measurements) will diverge from the state.

The greater $R(t)$, the less measurements help.

The less stable the dynamics are, the more past errors affect the quality of future estimates.

the Kalman gain:

$$\underline{K}(t+\Delta t) = \underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t)\left[\underline{H}(t+\Delta t)\,\underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t) + \underline{R}(t+\Delta t)\right]^{-1}$$

in discrete time.

Define

$$\underline{K}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\underline{K}(t+\Delta t)$$

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t)\left[\underline{H}(t+\Delta t)\,\underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t) + \underline{R}_c(t)/\Delta t\right]^{-1}$$

$$= \lim_{\Delta t \to 0} \underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t)\left[\underline{H}(t+\Delta t)\,\underline{\Sigma}(t+\Delta t|t)\,\underline{H}^T(t+\Delta t)\,\Delta t + \underline{R}_c(t)\right]^{-1}$$

$$\boxed{\underline{K}(t) = \underline{\Sigma}(t|t)\,\underline{H}^T(t)\,\underline{R}^{-1}(t)}$$

And

$$\frac{\hat{\underline{x}}(t+\Delta t|t+\Delta t) - \hat{\underline{x}}(t|t)}{\Delta t} \cong \underline{F}(t)\,\hat{\underline{x}}(t|t) + \frac{1}{\Delta t}\underline{K}(t+\Delta t)\left[\underline{z}(t+\Delta t) - \underline{H}(t+\Delta t)\,\underline{\Phi}(t+\Delta t, t)\,\hat{\underline{x}}(t|t)\right]$$

and lim $\Delta t \to 0$:

$$\boxed{\dot{\hat{\underline{x}}}(t|t) = \underline{F}(t)\,\hat{\underline{x}}(t|t) + \underline{K}(t)\left[\underline{z}(t) - \underline{H}(t)\,\hat{\underline{x}}(t|t)\right]}$$

Summarizing, the cont's time Kalman filter is:

$$\dot{\hat{\underline{x}}}(t|t) = \underline{F}(t)\hat{\underline{x}}(t|t) + \underline{K}(t)\left[\underline{z}(t) - \underline{H}(t)\hat{\underline{x}}(t|t)\right]$$

where

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

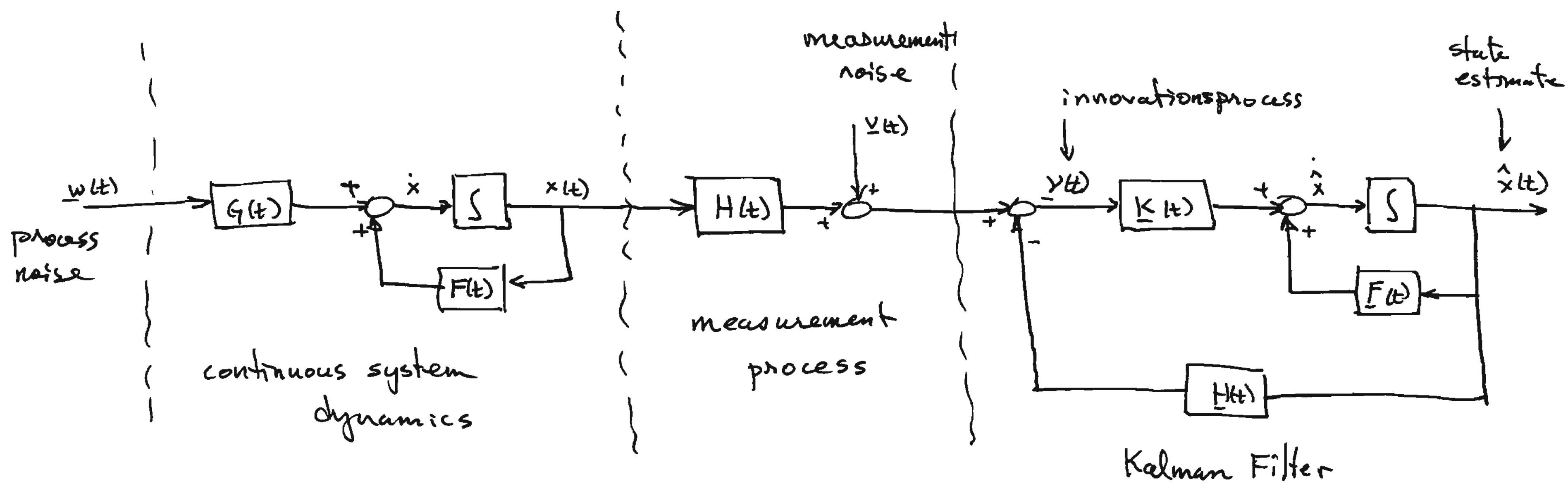
$$\underline{K}(t) = \underline{\Sigma}(t|t)\underline{H}^T(t)\underline{R}^{-1}(t)$$

Often, slightly suboptimal initial performance is acceptable in favor of a simpler implementation. For a LTI system with asymptotically stable $\underline{F}$ and observable (or at least detectable) pair $(\underline{F}, \underline{H})$, the steady state solution to the Riccati equation exists and is positive semi-definite. In this case, it is the $\underline{\Sigma} \geq 0$ that solves the <u>Algebraic Riccati Equation</u>:

$$\underline{0} = \underline{F}\underline{\Sigma} + \underline{\Sigma}\underline{F}^T + \underline{G}\underline{Q}\underline{G}^T - \underline{\Sigma}\underline{H}^T\underline{R}^{-1}\underline{H}\underline{\Sigma}$$

and $\underline{K} = \underline{\Sigma}\underline{H}^T\underline{R}^{-1}$

Block Diagram of System & Filter



Exercise: the linearized dynamics of a helicopter are given by the LTI system

$$\dot{\underline{x}} = \underline{A}\,\underline{x} + \underline{B}\,\underline{u} + \underline{w}$$

$$\underline{y} = \underline{C}\,\underline{x} + \underline{v}$$

where

$$\underline{A} = \begin{bmatrix} 0. & 1. & 0. & 0. & 0. \\ 0. & -0.415 & -0.0111 & 0. & -0.0111 \\ 9.8 & -1.43 & -0.0198 & 0. & -0.0198 \\ 0. & 0. & 1. & 0. & 0. \\ 0. & 0. & 0. & 0. & -0.2 \end{bmatrix}$$

$$\underline{B} = \begin{bmatrix} 0. \\ 6.27 \\ 9.8 \\ 0. \\ 0. \end{bmatrix} \qquad \underline{C} = \begin{bmatrix} 1. & 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 1. & 0. \end{bmatrix}$$

$$\underline{x}(0) = \begin{bmatrix} 0.02 & 0. & 0. & 0. & 0. \end{bmatrix}^T$$

$$E[\underline{w}] = \underline{0}, \quad E[\underline{v}] = \underline{0}.$$

$$E[\underline{v}\,\underline{v}^T] = \text{diag}(1.52 \times 10^{-5}, 0.272)$$

$$E[\underline{w}\,\underline{w}^T] = \text{diag}(0., 0., 0., 0., 19.6)$$

Design a controller of the ~~form~~ $\underline{u}(t) = \underline{G}\,\underline{x}(t)$ where $\underline{G}$ is an optimal linear quadratic regulator (LQR) gain. (Look up the documentation for LQR () in Matlab's control system toolbox, and apply it with $J = \int_0^\infty \underline{x}^T(t)\,\underline{Q}\,\underline{x}(t) + \underline{u}^T(t)\underline{u}(t)\,dt$ and $\underline{Q} \geq 0$.) Choose $\underline{Q}$ so the simulation of the initial transient settles out in a few seconds (5-20s).

Next, find the optimal steady-state Kalman filter, and replace the state feedback controller above with $\underline{u}(t) = \underline{G}\,\hat{\underline{x}}(t|t)$. Model and simulate the closed-loop system in Simulink. Compare the two results.

~~Optine~~

Optional: Discretize the system and noise with $\Delta t = 0.2$ second, and repeat the process above with discrete-time versions of LQR and Kalman filter. Compare the results. What happens for different Δt?

Linear System Forced By Correlated Noise.

$$\dot{\underline{x}} = \underline{F}\,\underline{x} + \underline{G}\,\underline{w}$$

$$\underline{y} = \underline{H}\,\underline{x} + \underline{v}$$

Suppose

$$\dot{\underline{w}} = \underline{M}\,\underline{w} + \underline{w}_1 \quad , \quad \underline{w}_1 \sim N(\underline{0}, \underline{Q})$$

$$\dot{\underline{v}} = \underline{N}\,\underline{v} + \underline{v}_1 \quad , \quad \underline{v}_1 \sim N(\underline{0}, \underline{R})$$

We can convert this estimation problem to that of the standard optimal filter (KBF) by state <u>vector augmentation</u>:

Let

$$\underline{x}' = \begin{bmatrix} \underline{x} \\ \underline{w} \\ \underline{v} \end{bmatrix}$$

then

$$\dot{\underline{x}}' = \begin{bmatrix} \dot{\underline{x}} \\ \dot{\underline{w}} \\ \dot{\underline{v}} \end{bmatrix} = \begin{bmatrix} \underline{F} & \underline{G} & \underline{0} \\ \underline{0} & \underline{M} & \underline{0} \\ \underline{0} & \underline{0} & \underline{N} \end{bmatrix} \begin{bmatrix} \underline{x} \\ \underline{w} \\ \underline{v} \end{bmatrix} + \begin{bmatrix} \underline{0} \\ \underline{w}_1 \\ \underline{v}_1 \end{bmatrix}$$

$$\underline{y} = [\underline{H} \;\; \underline{I} \;\; \underline{0}]\,\underline{x}' = [\underline{H} \;\; \underline{I} \;\; \underline{0}] \begin{bmatrix} \underline{x} \\ \underline{w} \\ \underline{v} \end{bmatrix} = \underline{H}\,\underline{x} + \underline{w}$$

Application: Wind. (see Engr. Dyn Systems Exhibit)

current speed & direction of wind is correlated with past. It can be modeled as

$$\dot{\underline{w}} = \underline{M}\,\underline{w} + \underline{w}_1 \quad \text{where } \underline{w}_1 \text{ is Gaussian white noise } (\sim N(\underline{0}, \underline{Q}))$$

- exponentially correlated r.p.

if $\underline{\Sigma}(t) = E\{\underline{w}_t \underline{w}_t^T\}$, then $\underline{\Sigma}(t)$ satisfies

$$\dot{\underline{\Sigma}}(t) = \underline{M}\,\underline{\Sigma} + \underline{\Sigma}\,\underline{M}^T + \underline{Q}$$

Now, $E\{\underline{w}_t \underline{w}_{t_1}^T\} = E\left\{\left(\underline{\Phi}(t,t_1)\underline{w}_{t_1} + \int_{t_1}^{t} \underline{\Phi}(t,\tau)\underline{w}_1(\tau)\,d\tau\right)\underline{w}_{t_1}^T\right\}$

$$= \underline{\Phi}(t,t_1)\,E\{\underline{w}_{t_1}\underline{w}_{t_1}^T\} + E\left\{\int_{t_1}^{t} \underline{\Phi}(t,\tau)\underline{w}_1(\tau)\,d\tau\,\underline{w}_{t_1}^T\right\} \quad (\to 0)$$

$$= e^{M(t-t_1)}\,\underline{\Sigma}(t_1) \qquad \text{for } t \geq t_1$$

exponential correlation.

Correlated Process & Measurement noise:

$$E\{\underline{w}(t)\underline{v}^T(\tau)\} = \underline{C}(t)\delta(t-\tau)$$

$$\dot{\underline{x}} = \underline{F}\underline{x} + \underline{G}\underline{w}$$

$$\underline{z} = \underline{H}\underline{x} + \underline{v}$$

Let $\underline{D} = \underline{G}\underline{C}\underline{R}^{-1}$

and

$$\dot{\underline{x}} = \underline{F}\underline{x} + \underline{G}\underline{w} + \underline{D}(\underline{z} - \underline{H}\underline{x} - \underline{v})$$

$$= (\underline{F} - \underline{D}\underline{H})\underline{x} + \underline{G}\underline{w} - \underline{D}\underline{v}$$

$$\underline{z} = \underline{H}\underline{x} + \underline{v}$$

$\underline{G}\underline{w} - \underline{D}\underline{v}$ is process noise

$$E\{(\underline{G}\underline{w} - \underline{D}\underline{v})(\underline{G}\underline{w} - \underline{D}\underline{v})^T\} = \underline{G}\underline{Q}\underline{G}^T + \underline{D}\underline{R}\underline{D}^T - \underline{D}\underline{C}\underline{G}^T - \underline{G}\underline{C}\underline{D}^T$$

$$= \underline{G}\underline{Q}\underline{G}^T + \underline{G}\underline{C}\underline{R}^{-1}\underline{R}\,\underline{R}^{-1}\underline{C}^T\underline{G}^T - \underline{G}\underline{C}\underline{R}^{-1}\underline{C}^T\underline{G}^T - \underline{G}\underline{C}\underline{R}^{-1}\underline{C}^T\underline{G}^T$$

$$= \underline{G}\underline{Q}\underline{G}^T - \underline{G}\underline{C}\underline{R}^{-1}\underline{C}\,\underline{G}^T$$

$$E\{(\underline{G}\underline{w} - \underline{D}\underline{v})\} = \underline{G}m_w - \underline{D}m_v$$

$$E\{(\underline{G}\underline{w} - \underline{D}\underline{v})\underline{v}^T\} = \underline{G}\underline{C} - \underline{G}\underline{C}\underline{R}^{-1}\underline{R} = \underline{G}\underline{C} - \underline{G}\underline{C} = \underline{0}$$

∴ new process (equiv.)

$$\underline{\dot{x}} = (\underline{F} - \underline{D}\,\underline{H})\underline{x} + \underline{\gamma}$$

$$\underline{z} = \underline{H}\underline{x} + \underline{v}$$

$$E\{\underline{\gamma}\underline{\gamma}^T\} = \underline{G}(\underline{Q} - \underline{C}\underline{R}^{-1}\underline{C})\underline{G}^T$$

$$E\{\underline{\gamma}\} = \underline{G}\underline{m}_w - \underline{D}\underline{m}_v$$

$$E\{\underline{v}\underline{v}^T\} = \underline{R}$$

$$E\{\underline{v}\} = \underline{m}_v$$

Non-zero mean noise:

$$E\{v\} = \underline{m}_v$$

re-define as $v' = v - m_v$

$$\underline{\dot{x}} = \underline{F}\underline{x} + \underline{G}\underline{w}$$

$$\underline{z} = \underline{H}\underline{x} + \underline{m}_v + \underline{v}'$$

or

$$E\{\underline{w}\} = \underline{m}_w \qquad \underline{w}' = \underline{w} - \underline{m}_w$$

$$\underline{\dot{x}} = \underline{F}\underline{x} + \underline{G}\underline{m}_w + \underline{G}\underline{w}'$$

if both are non-zero mean,

$$\underline{\dot{x}} = \underline{F}\underline{x} + \underline{G}\underline{m}_w + \underline{G}\underline{w}^T$$

$$\underline{z} = \underline{H}\underline{x} + \underline{m}_v + \underline{v}'$$

$$\underline{\dot{\hat{x}}} = (F - KH)\underline{\hat{x}} + \underline{K}(\underline{z} - \underline{H}\underline{x} - \underline{m}_v) + \underline{G}\underline{m}_w$$