Lecture II

Information Flow in the Kalman Filter

(Joint work: Nigel Newton)

Paper: SIAM J. on Control to appear

Various other papers

Mutual Information

$$I(X;Y) = \mathbb{E}\log\left(\frac{dP_{XY}}{d(P_X \otimes P_Y)}\right).$$

$$= \mathbb{E}\, h\left(P_{X|Y}(\cdot, Y) \,\middle|\, P_X\right).$$

= Average Information Gain on X arising from the process of observing Y.

<u>Path Estimator</u>

Let

$(X_t \mid 0 \le t < \infty)$: Markov Process

(<u>Ex:</u> Gauss-Markov Process)

is to be estimated on the basis of an observations process

$(Y_t \mid 0 \le t < \infty)$ (Signal + Noise : Gaussian)

$$I\left((X_r \mid 0 \le r \le s); (Y_r \mid 0 \le r \le t)\right)$$

$0 \le s, t < \infty$

= Ave. Inf. gain on part of the path of X arising from observations on part of the path of Y

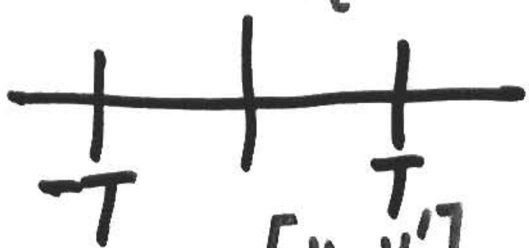
By manipulating this we are able to identify information source and dissipation rates for the filter

Time Reversal of Signal-Filter pair information supply and dissipation exchange roles in a dual filter. Time-reversed filter process becomes a dual signal and the time-reversed signal becomes the filter. for the process.

Information Flow in the Filter

$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$: Prob. Space

(1) $$X_t = \xi + \int_{-T}^{t} (B X_s + b(s))\,ds + V_t$$



$-T \leq t \leq T$

"State"

$$[V, V']_t = \int_{-T}^{t} A\,ds$$

$$m(t) = m_i + \int_{-T}^{t} (B m(s) + b(s))\,ds$$

Mean

$$P(t) = P_i + \int_{-T}^{t} ((B P(s) + P(s) B') + A)\,ds$$

Covariance

$$Y^i = \xi + \varphi$$

$N(0, M)$.

Initial Observation

Running Observation

$$Y^r_t = \int_{-T}^{t} G X_s \, ds + W_t \qquad -T \le t \le T.$$

$$G \geq 0$$

$$[W, W]_t = \int G \, ds$$

$$\equiv \quad \hat{Y}^r_t = \int_{-T}^{t} \Gamma X_s \, ds + W_t \qquad W_t \sim N(0, I)$$

$$\boxed{\Gamma' \Gamma = G.}$$

Matrix Signal to Noise Ratio

Combine Y^i and Y^r

$$Y_t = C Y^i + Y^r_t \qquad C : \text{non-singular}$$

Choose $C = I$

$$(2) \quad Y_t = Y^i + \int_{-T}^{t} G X_s \, ds + W_t \qquad \text{Signal + Noise}$$

Estimate

$$X_t \text{ given } \left(Y_s \middle| -T \le s \le t\right)$$

(Subtract out mean: to estimate zero-mean process). $\boxed{P\left(X_t \middle| Y_{-T}^{t}\right).}$

Filter

$$\hat{X}_{-T} = P_i \left(P_i + M\right) Y^{i} \quad \boxed{N\left(\hat{X}_t, Q(t)\right)}$$

$$d\hat{X}_t = \left(B - Q(t)G\right)\hat{X}_t dt + Q(t) dY_t$$

$$Q(-T) = \left(P_i^{-1} + M^{-1}\right)^{-1} \quad \boxed{\hat{X}_t = E\left(X_t \middle| \mathcal{F}_t^{Y}\right)}$$

$$\dot{Q}(t) = BQ(t) + Q(t)B' + A - Q(t)GQ(t)$$

Innovations

$$\nu_t = Y_t - \hat{X}_{-T} - \int_{-T}^{t} G\hat{X}_s ds \quad \left(\begin{array}{c}\text{New}\\ \text{Information}\end{array}\right)$$

Let

$$C(t) = I\left(X_t ; (Y_s \mid -T \le s \le t)\right)$$

= Average Inf. Gain on X_t arising from Observation path.

= Inf. Stored by the Filter at time t

$$= I\left(X_t ; \hat{X}_t\right) \qquad \hat{X}_t = E\left(X_t \mid \mathcal{F}_t^Y\right)$$

Suff. Stat.

$$C(-T) = \boxed{\tfrac{1}{2}\log\left|P_i + M\right| - \tfrac{1}{2}\log\left|M\right|}$$

+

$$\dot{C}(t) = \boxed{\tfrac{1}{2}\mathrm{Tr}\left(GQ(t)\right)} - \tfrac{1}{2}\mathrm{Tr}\left(A\bar{Q}^{1}(t) - \bar{P}^{1}(t)\right)$$

Information Stored
= Information Supply
− Information Dissipated

ANSATZ

Information Supply upto t

$$S(t) = \frac{1}{2}\log|P_i + M| - \frac{1}{2}\log|M| + \frac{1}{2}\int_{-T}^{t} \mathrm{Tr}\left(G Q(s)\right) ds$$

Information Dissipated upto time t
(by Fisher)

$$= \frac{1}{2}\int_{-T}^{t} \mathrm{Tr}\left(A\left(Q^{-1}(s) - P^{-1}(s)\right)\right) ds$$

$$D(t) = \frac{1}{2}\int_{-T}^{t} \mathbb{E}\,\mathrm{Tr}\left(A\,F(X_s, s)\right) ds$$

$$F(x,t) = \mathbb{E}\left|\nabla_x \log(A)(x,t,Y)\right|^2$$

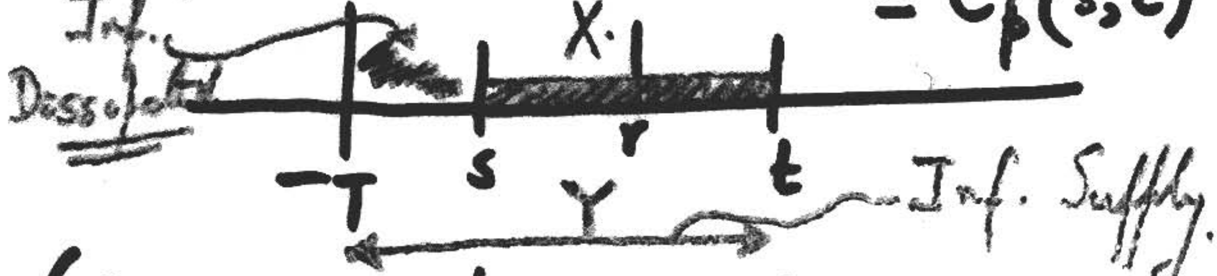
(Fisher Inf. Matrix)

Likelihood of (X_t) given $(Y_s \mid -T \le s \le t)$

$$S(t) - D(s)$$

$$= I\Big((X_r \mid s \leq r \leq t) ; (Y_r \mid -T \leq r \leq t) \Big).$$

$$= C_p(s,t)$$



(Requires looking at processes under two measures). See (17 bis)

$S(-T) =$ Inf. Gain on whole process X arising from Y^i

$S(t) - S(-T) =$ Inf. Gain arising from $Y^r \quad -T \leq r \leq t$

Average Total Information in the Observation upto t

$$O(t) = \frac{1}{2}\log|P_i + M| + \frac{1}{2}\operatorname{Tr}(P_i + M - I) - \frac{1}{2}\int \operatorname{Tr}(GP - Q)\,ds$$

Bayesian Estimator splits into Information Gain on the Estimand & Residual term concerning, for example Observation Noise

<u>Average Residual Inf.</u> for the estimator of the path $(X_r \mid s \le r \le t)$ given $(Y_s \mid -T \le s \le t)$

$$R_p(s,t) = O(t) - C_p(s,t)$$

Average Residual Information for the filter

$$R_f(t) = O(t) - C(t)$$
$$= R_p(-T, t) + D(t)$$

Historical Information about X