相互情報를 이용한 Stochastic 시스템의 관측도 판단

박 진 옥 육군사관학교 전자공학과
°황 순 식* 육군사관학교 전자공학과

Degree of Observability of the Stochastic system Using Mutual Information

1. 개요

Stochastic 시스템의 관측도 (degree of observability)는 deterministic 시스템에 있어서의 可觀側性(observability)판단과 같이 그 결과를 "예","아니요" 의 형태로 말 할수가 없다. 이와같은 관점에서 본 논문에서는 Shannon 의 相互情報(mutual information) 의 개념을 사용하여 Stochastic시스템의 관측도를 판단하는 한 방법을 소개한다.

전통적으로 사용되는 Fisher의 정보개념대신 Shannon 의 정보개념을 특별히 사용하는것은 후자에의한 방법이 전자에의한 방법보다 시스템의 과측도를 계산하는 과정에서 흔히 발생되는 몇가지 어려움을 해결하기위한 것이다. 즉stochastic 상태방정식(state equation) 과 관측방정식(observation equation)으로 주어진 한 시스템의 임의 하나이상의 상태가 어떤 이유로 관측불가한 경우 Fisher의 개념에의한 정보량을 실재 계산하기 위해서는 정보행렬(information matrix) 을 계산하여야 하는데 이런경우 정보행렬이 singular하게되어 실재계산이 불가능해 진다. 그러나 Shannon 의 정보개념에 의한 상호정보는 이런경우에도 관측불가한 상태는 단지 정보량의 성장이 없을 뿐 정보행렬 자체의 계산을 불가능하게 만들지는 않으므로 관측이 가한 상태의 정보량계산을 가능케 하여 시스템 전채의 관측도를 판단할 수 있게 한다.

여기에서 계산되는 상호정보란 측정된 관측자료 중에 포함된 상태물에 관한 정보량으로서 이 양은 그대로 주어진 시스템의 관측도라고 볼수 있으므로 본 논문에서는 이를 가지고 해당시스템의 관측도 판단의 기준으로 삼았다. 상호정보행렬의 계산에는 conditional 확률밀도함수와 unconditional 확률밀도함수가 필요한데 이들의 정확한 계산에는 stochastic 편미분방정식을 풀어 그 해를 얻어야하는 문제가 있어 이를 푸는대신 두 확률밀도의 covariance의 비로서 상호정보를 계산할수 있음을 보인다.

제2 절에서는 일반적인 Ito stochastic 시스템 방정식의 형태로 주어진 선형 혹은 비선형계의 상호정보를 계산하는 방법을 Fisher의 정보와 함께 소개하며 3 절에서는 간단한 2차원 선형시스템인 자유낙하 물채에 대한 시스템의 시뮤레이션을 통하여 상호정보가 주어진 stochastic 시스템의 관측도를 판단하는 한 기준이 됨을 보이고 Fisher의 정보량과 비교한다. 마지막 절에는 본 연구의 결론으로 되어있다.

2. Stochastic 시스템의 상호정보

일반적인 Ito stochastic 시스템의 상태 및 관측방정식이

$$dx_t = f(x_t)dt + g(x_t)dw_t \quad , \quad x_{t_0} = x_0 \qquad (1)$$
$$dy_t = h(x_t)dt + dv_t \qquad (2)$$

와 같이 주어져 있다. 여기서 $x_t \in R^n, y_t \in R^m$, $f(), g(), h()$ 는 상태 x_t 에 관한 적절한 차원의 함수행렬 또는 벡터이고 잡음process $\{w_t\}$ 및 $\{v_t\}$는 서로독립인 Wiener process 로서 variance가 각각 $Q(t), R(t)$ 이다[1]-[4]

(1),(2) 시스템의 상태 x_t의 conditional 확률밀도 $p_t(a)$ 는

$$p_t(a) = dF(x_t \le a | Y_S)/da \; , \qquad (3)$$

Y_S 는 $\{y_s \; , s \le t\}$ 에 의하여 생성된 sub-σ-algebra 이고 이에 해당하는 Fisher의 정보량 $J(x_t, y_t)$는

$$J(x_t, y_t) = -E\left[\frac{\partial^2}{\partial a \partial a^T} \ln p_t(a)\right] \tag{4}$$

이고 Shannon 의 정보개념에의한 프로세스 x_t 와 y_t간의 상호정보 $I(x_t, y_t)$는

$$I(x_t, y_t) = E\left[\ln \frac{p_t(a)}{q_t(a)}\right] \tag{5}$$

이다. E 는 기대치연산자이고 (5)에서 unconditional확률밀도함수 $q_t(a)=dF(x_t \le a)/da$ 이다 [3],[5] .

(4),(5) 에서 확률밀도 $q_t(a)$ 와 $p_t(a)$ 는 각각 Kolmogorov의 forward equation과 Kushner equation의 해에서 이들을 구하기 위해서는 복잡한 편미방을 풀어야 한다. 더구나 이들의 해는 간단한 선형 시스템의 경우외에는 해석적으로 풀리지 않으므로 이들을 풀지않고 J(.) 나 I(.) 를 구하는 개략적 방법이 모색 되었다[2],[3]. 대부분의 경우 확률밀도함수를 구하는 대신 밀도함수의 처음몇차 모우먼트를 구하는것이 통예이다.

시스템 noise w_t가 없는 경우의 Fisher의 정보 (4)의 실제계산공식은

$$J(.) = \phi^T(t_0, t)P_0^{-1}\phi(t_0, t) + \int_{t_0}^{t}\phi(s,t)h_x^T R^{-1} h_x \phi(s,t)ds \tag{6}$$

의 형태로 주어진다. 여기서 $\phi(.)$는 함수 $f(.)$의 x_t에 관한 1차선형와 부분에 대한 전이행렬이고 h_x는 $h(x_t)$ 의 선형부분이다.

상호정보 (5)의 개략적 계산을 위하여 엔트로피의 개념을 도입한다 [7]. 프로세스 x_t의 확률밀도함수를 $q_t(x_t)$라면 이의 엔트로피는

$$H(x_t) = -E[\ln q_t(x_t)] \tag{7}$$

이고 conditional 확률밀도함수 $p_t(x_t)$에 대한 conditional entropy $H(x_t|y_t)$ 는

$$H(x_t|y_t) = -E[\ln p_t(x_t)] \tag{8}$$

이므로 (5)의 상호정보는 (7),(8) 을 대입하여

$$I(x_t, y_t) = H(x_t) - H(x_t|y_t) \tag{9}$$

가 된다. 실제 marginal 엔트로피 $H(x_t)$ 와 조건부 엔트로피 $H(x_t|y_t)$ 의 계산을 위하여 entropy-covariance관계식을 이용하면 [7],[8]

$$H(x_t) = \frac{n}{2} \ln A + \frac{1}{2} \ln(\det Q_{Tt}) \tag{10}$$

$$H(x_t|y_t) = \frac{n}{2} \ln A + \frac{1}{2} \ln(\det P_{Tt}) \tag{11}$$

을 이용하면

$$I(x_t, y_t) = H(x_t) - H(x_t|y_t) = \frac{1}{2} \ln\left(\frac{\det Q_{Tt}}{\det P_{Tt}}\right), \quad Q_{T0} = P_{T0} \tag{12}$$

가 된다. 관계식 (10),(11)은 밀도함수 q_t와 p_t 의 mean 과 covariance (Q_{Tt} ,P_{Tt}) 가 존재 하기만하면 성립하는 식이며 A 는 밀도함수의 종류에따라 결정되는 상수이다.

식(12)의 초기조건을 $Q_{T0} = P_{T0}$ 로 놓음으로서 시스템의 초기정보는 항상

$$I(x_0, y_0) = \frac{1}{2} \ln(1) = 0 \tag{13}$$

으로 normalize 된다. 개별상태, 예를들면 총 n 개의 상태중 i-번째 상태의 상호정보

$$I_i(x_i, y_t) = \frac{1}{2}\ln\left(\frac{Q_{Tii}}{P_{Tii}}\right), \quad i=1,2,\dots n \tag{14}$$

가 된다. Q_{Tii} ,P_{Tii} 는 Q_{Tt} 와 P_{Tt} 의 주대각선의 i-번째 요소이다.

만일 covariance Q_{Tt}와 P_{Tt} 가 각각 mean

$$\bar{x}_t = E[x_t|X_s, 0 \le s \le t], \quad \hat{x}_t = E[x_t|Y_s, 0 \le s \le t]$$

covariance Q_t,P_t 를 갖는다면 (14) 는

$$I(\hat{x}_t, y_t) = H(\bar{x}_t) - H(\hat{x}_t|y_t) = \frac{1}{2} \ln\left(\frac{\det Q_t}{\det P_t}\right), \quad Q_0 = P_0 \tag{15}$$

가 되어 (15) 는 실제 시스템의 관측도를 판단하는 기준으로 사용할수 있는 공식이 된다. $\bar{x}_t$,$\hat{x}_t$, 또는 Q_t,P_t 의 계산에는 이미 널리 알려져 있는 여러 filtering algorithm 이 쓰일수 있다. 예를들면 선형부분만을 고려할때 Q_t 와 P_t 는

$$\dot{Q}_t = f_x(\bar{x}_t)Q_t + Q_t f_x(\bar{x}_t) + g(t)Q(t)g^T(t),$$

$$f_x(.) = \left.\frac{\partial f}{\partial x_t}\right|_{x_t=\bar{x}_t} \tag{16}$$

$$\dot{P}_t = f_x P_t + P_t f_x^T + g(t)Q(t)g^T(t) - P_t h_x^T R^{-1} h_x P_t$$

$$, h_x(.) = \left.\frac{\partial h}{\partial x_t}\right|_{x_t=\hat{x}_t}, \quad P_{t0} = P_0 = Q_0 \tag{17}$$

이다. (16),(17) 에서 g 는 오직 t 의 함수로

만 가정하였다. (16),(17) 대신 정확도를 높이기위한 어떠한 고차 algorithm도 쓰일수있다.

Fisher의 정의에따른 정보량(4) 또는 그의 실제적용공식 (6)이 정보량의 계산에 보편적으로 널리 쓰이나, 시스템의 관측도 판단을 위한 상호정보의 계산에는 Shannon 의 정의에 따른 정보량 (5)또는 실제 적용을위한 공식 (12) 및(15)가 다음과 같은 점에서 더 효과적으로 쓰일수 있다.

1). Fisher정보량 계산식 (6)은 모든상태의 deterministic 한 부분이 observable하고 영이아닌 초기정보를 가지는 경우에만 적용할수있다. 그렇지 않으면 정보행렬 (6) 은 singular하게되어 정보량계산은 할 수 없게된다. 그러나 상호정보량(12), 또는 (15)는 설사 어떤상태가 deterministically 관측가하지 않더라도 초기정보량을 항상 갖게 놓음으로서 계산자체가 불가해지는 경우는 없게 된다. 단지 관측가하지 않은 상태에 해당하는 정보량의 성장이 시간에따라 매우 느리거나 또는 성장이 없는것으로 나타난다.

2). 관측이 불가한 상태를 관측가한 상태와 구별하는데 있어 Fisher의 정보에서는 경험이나 시행착오등의 방법으로 행하나 상호정보에 있어서는 해당상태의 정보량의 증가를 보아 즉각적으로 알수있다.

3). 이론적으로는 두 방법이 모두 시스템및 관측방정식의 잡음을 포함할수 있으나 실제 문제의 적용에 있어 Fisher의 정보 (6) 은 언제나 시스템잡음이 없는 경우에만 적용이 가능하나 상호정보(15)에 있어서는 두 잡음을 모두 고려한다.

4). 비선형계에 대한 Fisher의 정보량은 언제나 선형화한 부분에 대하여만 계산이 가능하나 상호정보량 (15)는 어떤 고차의 filter algorithm 이든지 제한이 없다.

본 논문에서는 위와같은 잇점을가진 상호정보에 의한 방법으로 시스템의 관측도 판단을 위하여 정보량 (15) 를 사용하여 판단의 기준으로 삼았으며 다음절에서는 간단한 시스템을 예로들어 시뮤레이션한 결과를 보이므로서 상호정보량이 관측도의 높고 낮음을 판단하는 한 기준이 됨을 보이고자한다.

3. 2 차원 선형시스템의 예

일정한 중력가속도 g 를 받아 자유낙하 하는 물체의 운동방정식은 다음과같이 표시된다

$$\ddot{z}(t) = -g\ , \quad t\geq 0 \tag{18}$$

$x_1 = z$(위치), $x_2 = \dot{z}$(속도) 로 놓아 얻은 상태방정식은 간단한 2차원 선형계이며

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) \quad + \begin{bmatrix} 0 \\ -g \end{bmatrix} \tag{19}$$

이고 잡음을 고려한 관측방정식은

$$y_t = H\, x(t) \ + \ v_t \tag{20}$$

이 되며 간단한 rank test 로 위치 x_1을 관측할 때는 시스템의 deterministic 한 부분이 가관측적이지만 속도 x_2를 관측할 때에는 가관측성이 아님을 알수있다.

이 시스템에 대한 Fisher의 정보 (6)은 초기정보가 없을경우

$$J(.) = \int_{t0}^{t} \phi^T(s,t) H^T R^{-1} H\, \phi(s,t)ds, \tag{21}$$

이 되며 이의 decrete 한 형태는

$$J(k,1) = \sum_{i=1}^{k} \Phi_i^T H_i^T R_i^{-1} H_i \Phi_i\ , \tag{22}$$

가 되어 x_1을 관측하여 observable한 경우에는 정보행렬이 $J(k,1) = \begin{bmatrix} k & k \\ k & 2k \end{bmatrix}$ 가되어 $k \geq 2$ 부터는 J 가 nonsingular 하지만 x_2를 관측하여 unobservsble한 경우에는 $J(k,1) = \begin{bmatrix} 0 & 0 \\ 0 & k \end{bmatrix}$ 가 되어 J 는 항상 singular 하여 정보량을 계산할수 없음을 알수있다.

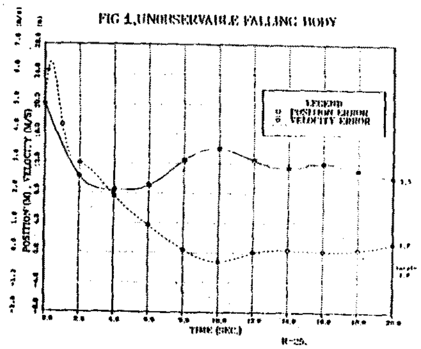
한편 이 시스템의 관측도 판단을 위하여 식(15)에 의한 상호정보를 계산한결과를 표1에 나타내었다. Covariance P_t, Q_t를 계산하기 위하여 사용한 filter는 Kalman-Bucy filter 이고 잡음은 가우시안 잡음을 사용하였다. 비교를위하여 선형스케일과 log 스케일을 동시에 나타내었다.

보는바와 같이 x_1 상태를 관측하여 deterministic시스템이 observable한 경우에는 전체 시스템의 상호정보 또는 관측도(O_T) 가 최종시간 20 초에 5.7까지 증가하고 위치및 속도에 관한정보(O_p및O_v) 도 4.9와1.8 까지 각각 증가하여 전 시스템이 매우강하게 observable하나 x_2(속도) 를 관측하여 unobservable한 경우에는 O_T가 2.8, O_p=2.3,O_v=1.8 까지밖에 정보가 성장하지못한다. 관측된변수인 속도변수만이 관측도 1.8로서 前者와같고 나머지는 훨씬 미치지못함을 알수있다. 이와같은 관측도의차이 즉 상호정보량의 차이로 인한 filtering error 를

Table 1. Observability of the falling-body

t (sec) \ obs.	Observable system						Unobservable system					
	Linear scale			Log scale			Linear scale			Log scale		
	O_T	O_P	O_V	O_T	O_P	O_V	O_T	O_P	O_V	O_T	O_P	O_V
0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0
2	3.3	2.8	1.1	1.2	1.0	0.1	1.8	1.2	1.8	0.6	0.2	0.6
4	7.5	5.1	1.7	2.0	1.6	0.5	2.5	1.6	2.4	0.9	0.5	0.9
6	15.3	9.8	2.5	2.7	2.3	0.9	3.3	2.3	3.0	1.2	0.8	1.1
8	28.0	17.7	3.3	3.3	2.9	1.2	4.3	3.7	3.5	1.5	1.2	1.3
10	45.0	28.2	3.9	3.8	3.3	1.4	5.5	4.1	4.0	1.7	1.4	1.4
12	73.0	43.0	4.4	4.3	3.8	1.5	7.1	5.2	4.4	1.9	1.6	1.5
14	109.0	61.0	4.8	4.7	4.1	1.6	9.0	6.3	4.8	2.2	1.8	1.6
16	158.0	87.0	5.2	5.1	4.4	1.7	11.3	7.6	5.2	2.4	2.0	1.7
18	223.	107.	5.6	5.4	4.6	1.7	14.0	9.0	5.6	2.6	2.2	1.7
20	307.	137.	6.0	5.7	4.9	1.8	17.7	13.4	6.0	2.8	2.3	1.8

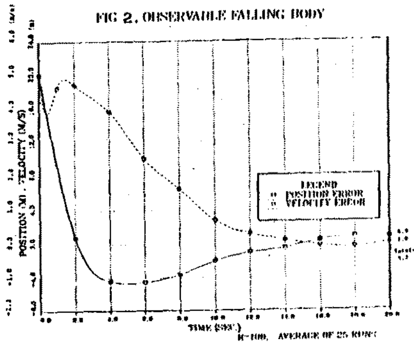
가관측성인 경우와 가관측성이 아닌 경우에 대하여 비교한것이 그림1 과2 이다. 같은 초기 오차인 위치오차20m , 속도오차5 m/s 에 대한것 인데 가관측성인 경우는 빠른속도로 두 오차가 모두 영으로 수렴하나 unobservable 한 경우는 그렇지않음을 볼수있다. 물론 이는 위치 변수의 약한 관측도 때문이다.



4. 결 론

Stochastic시스템의 관측도를 판단하는 한 방법으로서 Shannon 의 정보개념에의한 상호정보를 사용하여 관측도의 상대적인 높고 낮음을 파악하는 방법을 제시하였다. 이 방법이 전통적으로 사용되어온 Fisher의 정보개념에 의한 방법보다 실제의 계산에있어 몇가지 점에서 이점이 있음을 보았다.

간단한 2차원시스템의 시뮬레이션 예를 통하여 제시한 방법이 시스템의 관측도를 판단하는 다른 한 기준으로서 효과적으로 사용 될 수 있음을 보였다.



참고문헌

[1] C.B.Chang,"Two lower bound on the covariance for nonlinear estimation problems", IEEE Tr.on Auto.cont.,AC-26,1981.
[2]P.S.Maybeck,"Stochastic models,estimation and control(vol.2)",Academic Press,NY,1982.
[3]A.H.Jazwinsky,"Stochastic process and filtering theory",Academic Press,NY,1970.
[4]M.Aoki,"Optimization of stochastic systems" ,Academic Press,NY,1967.
[5]C.E.Shannon,W.Weaver,"The mathematical theory of communication",The Univ.of Ill. press,Urbana,1949.
[6]R.S.Liptser,A.N.Shiryayev,"Statistics of ramdom process II",Springer-Verlag,NY,1978.
[7]R.W.Hamming,"Coding and information theory" ,Prentice-Hall,NJ,1980.
[8]P.Kalata,M.Priemer,"Linear prediction, filtering andsmoothing",Info.Sci.Vol.17,pp 1-14,1979.