

State, System-Identificationについて

電気試験所 深 尾 毅

1. 序 言

近年、多くの最適制御の問題が活潑に議論され、多くの成果があがっていることは周知の通りである。しかしながら、これらの成果を実際に応用しようとするとき、種々の制約があつて smooth に行かない。これは1つには実際的な問題をかなり抽象化したモデルで考察しているからであろうが、その様な抽象化があつてはじめて理論は進むのだからやむを得ない。問題は従つて、実際にすぐ使えないから問題にしないというのではなくて、打樹てられた理論と実際面との交流をどの様にしてはかるかにある。理論は実際面から唯1つの formulation を行つて、その中で精密な議論をするのに満足しては、この様な交流ははかれないであろう。諸種のデータから各様の formulation を試みるべきであろう。そこに新しい問題として、生のデータからモデルをつくる formulation の方法に関する議論がおこる。それに関連する、しかしまだかなりレベルの低い議論が system characterization とか system identification である。model の形式に関する問題、微分方程式であらわされるとか、有限集合の element 間の関係であらわされるとか、そういった model 形式の選択に関する問題を system characterization とよぶ。そういう形式が定つて、system の類別の議論が次に出てくる。それを system classification とよぶ。そして、実際に観測された input, output にもとづいて、今問題にしている system が、どの class に属しているのか、あるいは属しているのがもつともらしいかというのが system identification の問題である。それが快ればその class の代表 model で system があらわされる。今問題にしている system が、それらの各 class に属することが真であるという仮説を考えるなら、system identification は仮説検定の問題であると云つてもよい。

特に、system の形が、例えば実関数による変換で表現出来て、その中にふくまれる parameter が未知であるという具合に characterize される場合も多い。この場合には、仮説検定というよりは推定 (estimation) の問題と云つた方がよいであろう。

更に、system には「状態」(state) という量を導入した形で characterize されると都合がよいことが多い。ある時刻での「状態」とは、いわばそれ以前の過去の system の経歴が集約された量であつて、それを知れば、過去の膨大な断片際な information は捨

ても後で困らないといったものである。ところで、その様な都合のよい「状態量」が直接には output として測れないかも知れない。従って、「状態」が、現在、未来、過去においてどの様な値をとっているか、とりうるか、とつていたかを推定する問題がやはり重要である。これも system identification の中にふくめられるが、普通は区別して state-identification とよんでいる。

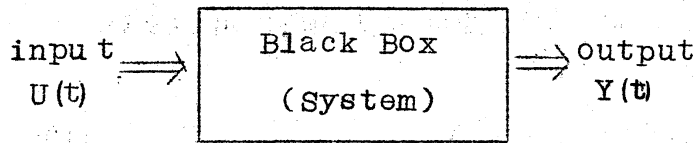
system identification などを伴う必要のある system control の問題は結局のところ、与えられる（場合によつては求めて得られる）データをどの様に確認しようかということ、つまりモデル化、定式化というものを（部分的にでも）必要としている問題と云える。従来の経験から、大体のモデル、定式化はあるが、その「近傍」でそれを多少 modify して実際のデータを説明しうる、あるいはしやすいものに行こうということである。

従って、system characterization や identification は、だんだん広げて行けば、生のデータからかなり高等な定式化を導ける様な process にもなり、理論と実際とのかけ橋として重要な意味をもつてくるであろうし、更に、程度の問題を別にすれば、人間の思考過程 (thought process) といったものにも関連するであろう。さて、生のデータが与えられ、それから推定したり、制御したりする場合、差当りうまいデータ処理の方法は確率論的あるいは統的な方法であろう。そういう意味で確率的あるいは統計的モデルをとりあげるのであつて、何もこれにこだわる必要はない。

尚、identification における重要な点の1つは、その逐次性、つまり、必要なデータを全部とりそろえてから一挙に identify するのではなく、データが逐次入ってくるに従つて段々と推定なり実験方法なりを新めて行くことである。これは大げさな云い方をすれば学習過程 (learning process) の1つである。又、別の重要な点としては、実験選択の問題がある。システムへのインプット系列（原因）のえらび方はある程度任意であり、システムからのアウトプット（結果）のまとめ方にも任意性がある。インプット系列とアウトプットのまとめ方（あるいは観測の仕方）は1つの実験を構成する。従つて identification の問題には実験の選択がともなう。しかも a priori に「最良の」実験を選択するわけにいかないから、本質的に、我々のその時々知識に応じて適当な意味での optimal な実験を選択しなければならない。

2. 有限オートマトンと制御系

第1図に示す様に、ある black box があつて、その input U と output Y とが観測出来、black box の中味は直接見るわけにはいかないというモデルはよくあらわれる。



第 1 図

ここで状態 (state) という量を導入する。(厳密には次にのべるオートマトンの定義の中に無定義元素としてふくまれる。)それは、各時刻で、システムがたどつて来た過去の経路を完全に表現出来るものでありたい。従つてある時刻で、その様な状態を指定すれば、過去にどのような input, output の系列があつてシステムが現在に至つているかを一々記述する必要がなくなる。

又、ある時間間隔はなれた時刻毎に input が加わり、それに応じて output が出るものとする。つまり同期信号源があつて、それに依存して、一斉に現象が進行する様な discrete time system を考える。それを同期システムとよぶ。

さて、

$$\text{有限インプット集合} \quad U = \{ u_1, \dots, u_p \}$$

$$\text{有限アウトプット集合} \quad Z = \{ z_1, \dots, z_q \}$$

$$\text{有限状態集合} \quad X = \{ x_1, \dots, x_n \}$$

と、次の1対の関数 f_X, f_Z をもつた同期システムを有限オートマトンとよぶ:

$$X(t+1) = f_X(X(t), U(t)) \quad (1)$$

$$Z(t) = f_Z(X(t), U(t)) \quad (2)$$

f_X, f_Z が確率的な関数である場合、確率的有限オートマトンとよぶ。

例 20円切符自動販売機

インプット集合 : { 10円玉を入れる, 50円玉を入れる }

アウトプット集合 : { 何も出さない, 切符を出す, 30円のつり銭を出す, 40円のつり銭を出す }

状態集合 : { 貯えなし, 10円貯え }

前の(1), (2)と等価ではあるが次の(1'), (2') の形式を用いた方が便利な場合も多い。

$$X(t+1) = f_X(X(t), U(t)) \quad (1')$$

$$Z(t) = f_Z(X(t)) \quad (2')$$

これはたとえばoutputが観測量である場合、すなわち(2')が観測装置で、観測したからと云ってsystemに擾乱はおきないという場合である。制御系の議論では、このような場合が多いので以下では(1'), (2')の形式を採る。

有限オートマトンでのインプット, アウトプット, 状態の集合には、普通位相は導入されていない。今、それらの有限性をはなれて、各集合にユークリッド距離による普通の実数位相を導入して、それらが連続濃度をもつとした時、(1'), (2')なる1対の実関数 f_X, f_Z をもつ同期システムを連続オートマトンあるいは単に制御系と以下では呼ぶことにする。普通、制御理論で取扱われる対象は多くはこの様な連続オートマトンあるいは制御系である。(実際には勿論これのみではなく、離散時間システムの代りに連続時間システムをとる場合——この場合(1')は微分方程式であらわされることが多い——、インプット, アウトプット, 状態の集合のうちのいくつかが離散あるいは有限濃度をもつ場合など。)

いずれにせよ、ここで強調したいことは、有限オートマトンも、制御理論の対象になる制御系も、実は広い意味では同一範疇に属するシステムであり、従って共通した問題点、考え方があることとしてこれら「両極端」の中間に、位相の導入の仕方によつて各様の「オートマトンあるいは制御系」があり、それらを総括するところに広い意味での制御理論の新しい領域があるということである。因に、有限オートマトンはいわばデジタル計算機の抽象化であり、連続オートマトンはアナログ計算機の抽象化である。

3. 最適制御問題

3-1. Markov 的多段決定過程

(A) 連続的オートマトンあるいは制御系:

$$X(t+1) = f_X(X(t), U(t), r(t)) \quad (1)$$

$$Y(t) = f_Y(X(t), U(t), W(t)) \quad (2)$$

を考える。ここで、 $r(t)$, $W(t)$ はwhite noise であるとする。 f_X, f_Y は決定論的関数で system の random behavior は $r(t)$, $W(t)$ に起因するものとする。

output $Y(t)$ は2つの部分にわかれ、

$$Y'(t) = f'_Y(X(t), W(t)) \quad (3)$$

$$Y''(t) = f''_Y(X(t), U(t)) \quad (4)$$

$Y'(t)$ は"観測量", $Y''(t)$ はsystem の"生産性"あるいは"経済性"を示す量 (cost) であるとする。

(B) system の状態が1)に従って次々と遷移し、その都度発生する output $Y''(1)$, $Y''(2), \dots, Y''(N)$ の関数 $F\{Y''(1), Y''(2), \dots, Y''(N)\}$ が次の形をとる場合、それを一応、シーケンシャル型の関数ということにする：

$$\begin{aligned} & F\{Y''(1), Y''(2), \dots, Y''(N)\} \\ &= F_1\{Y''(1), F_2\{Y''(2), F_3\{Y''(3), \dots, F_N\{Y''(N)\} \dots\}\}\} \end{aligned} \quad (5)$$

ここで $F_i(A, B)$, $i=1, \dots, N-1$, $F_N(A)$ は A, B あるいは A の任意の実関数。たとえば $F_1(A, B) = A+B$, $F_1(A, B) = A \cdot B$, $F_1(A, B) = \max(A, B)$ など。特別な場合 (例えば $F_i(A, B) = A+B$ あるいは $A \cdot B$) に

$$\begin{aligned} & F_1\{Y''(1), F_2\{Y''(2), \dots, F_N\{Y''(N)\} \dots\}\} \\ &= F_N^*\{Y''(N), F_{N-1}^*\{Y''(N-1), \dots, F_1^*\{Y''(1)\} \dots\}\} \end{aligned} \quad (6)$$

(C) system の可能なすべての状態に対して、夫々たゞ一つの制御 U を対応させた関数 $U = g(X)$ を決定関数あるいは制御関数という。これは制御の、状態 X に対するルールを示すものであり、戦略 (strategy) という言い方もする。

(D) 連続オートマトン又は制御系に対し、そのoutput系列のシーケンシャル型の関数の、
 適当な意味での平均値を最小又は最大ならしめる制御関数

$$U_{op} = g_{op} \{ \hat{X}(k) ; k \}, \quad k = 1, 2, \dots, N \quad (7)$$

を求める問題をMarkov的多段決定過程という。 $\hat{X}(k)$ は $X(k)$ の推定値である。

3-2. 確率的最適制御過程

連続オートマトンあるいは制御系の基本式並びに $r(t)$, $W(t)$ の確率的性質が完全に既知なる
 時のMarkov的多段決定過程を確率的最適制御過程という。

特別な場合としてBayes最適制御, minimax最適制御もこの中にふくまれる。

Bayes最適制御⁴⁾

State-estimationを行つてから、それにもとづいて、optimal controlを
 決める代りに、過去の観測データ及び制御にもとづくstateのconditional dis-
 tributionによる平均値をoptimalにする制御を採用する。すなわち、時刻 t では t
 より最終時刻に至る評価関数を $F_t(X(t))$ とすると、

$$\int E[F_t(X(t))] p(X(t) | Y(1), \dots, Y(t); U(1), \dots, U(t-1)) dX(t)$$

をoptimalにする U を採用する。

但し、 E は $r(t)$, $r(t+1)$, ..., $r(N)$ によるexpectationを示す。

この問題は一見(7)の範疇に属していない様に見えるが、次の如く解釈すれば、その特別な場合
 であることがわかる。すなわち、逐次推定のprocessは、若し適当に「知識の状態」を示す
 状態を定義すれば、state transition equationであらわされるsystemとな
 る。従つて、逐次推定を伴う制御系を全体として又制御系になる様にするには、「物理的」状態
 と、上述の「知識の」状態とを成分としてもつ「状態」が必要である。そして、その遷移特性は
 物理系のそれと同時に知識系のそれ、すなわち逐次推定あるいはidentification pro-
 cessとから成る様に考える。今の、Bayes最適制御では $p(X(t) | Y(1), \dots)$ が、 t 時刻
 における知識の状態である。(そしてそのestimateでもある。)

しかし、多くの場合、最適制御を決定するのに、物理的状态については、あらゆる可能な状態

を考慮するのに、知識の状態あるいは知識系は一定の決った形で取り入れられているにすぎない。物理系と知識系の併立した system の完全な記述は Bellman によつて行われた。^{8), 9)}

さて、Bayes 最適制御はかなり一般的であるが、しかし t 時刻における制御関数が

$$U_{Op} = g(Y(1), \dots, Y(t), U(1), \dots, U(t-1))$$

というきわめて次元の高い多変数関数になり、しかも t の進行と共に次元がふえていくから計算上大きな困難がありうる。

Minimax 最適制御

今、最適化とは評価関数の最小化を意味するものとしよう。 $X(t)$ がわからないので最悪の事態を考へて、 $E\{f_t(X(t))\}$ を最大 (最小問題にとつて最悪) にする $X^*(t)$ が実現したとして、その時の $E\{f_t(X^*(t))\}$ を最小にする制御 $U(t)$ を採用する：

$$\min_{U(t)} \cdot \max_{X(t)} \{ E\{f_t(X(t))\} \}$$

あるいは、いくつかの実験 e_1, \dots, e_k が平行に行われて、それらから $X(t)$ の推定値 $\hat{X}^{(1)}, \dots, \hat{X}^{(k)}$ が得られているとし

$$\min_{U(t)} \cdot \max_{\hat{X} \in \{\hat{X}^{(j)}\}} \{ E\{f_t(\hat{X}(t))\} \}$$

と云つた model も考えられる。

この2つの例は両極端である。

3-3. 適応最適制御過程

連続オートマトンあるいは制御系の基本式並びに $r(t)$, $W(t)$ の確立的性質が不完全にしか知られていない時の Markov 的多段決定過程を適応制御過程という。制御系の基本式及び $r(t)$, $W(t)$ のパラメーターが未知であるという形で議論する場合、パラメーター適応制御過程ということもある。主として議論されているのはこの範疇に属するものである。

適応制御過程では従つて、未知パラメーターの推定、あるいは system identifica-

tion と、これに前と同じく state の推定、あるいは state-identification とを行う必要がある。

ここでも各種の方式が考えられるが、確率的最適制御のところでのべた様な議論が成立つ。

たとえば Bayes 最適制御で云えば

$$\int E\{F_t(X(t))\} p(X(t), \alpha; Y(1), \dots, Y(t); U(1), \dots, U(t-1)) dx(t)$$

を optimal にする U を採用する。ここで α はシステムの未知パラメータである。

4. Wiener-Kalman Filter

システムへの外部環境、すなわち random input がどの様になっているかが未知である場合が多い。それを推定する問題、signal estimation, signal identification の問題を考える。

今、確率的性質のわかっている random process $X(t)$, $-\infty < t < \infty$ と、それと確率的に従増関係にあるもう 1 つの random process $Y(t)$ とがある。(たとえば $X(t)$ が signal で $Y(t)$ は signal plus noise.) これに対する linear filtering problem とは、 $Y(t)$ の $(-\infty, T)$ にわたる観測値 $Y(\tau)$, $-\infty < \tau < T$ が与えられたとき、

(1) $Y(\tau)$, $-\infty < \tau \leq T$, の線型汎関数で、

(2) 二乗平均誤差 $E\{|X(t_1) - \hat{X}(t_1)|^2\}$ を最小にする如き、

$X(t_1)$ の estimate $\hat{X}(t_1)$ を見出す問題である。他し

$$t_1 = T$$

若し $t_1 > T$ なら prediction problem, $t_1 < T$ なら smoothing problem とよばれる。

この問題を Wiener は、random process の stationality の条件のもとで、有名な Wiener-Hopf の積分方程式を解く問題に reduce し、それを解いて上述の解としての線型回路、linear filter (Wiener filter) の設計法を示した。

これに対し、近年、Kalman は、⁽¹⁾

(1) random process $X(t)$ が、white noise を input とする、有限次元の

linear dynamic system の output である (第2図参照) :

$$X(t+1) = \Phi(t+1, t) X(t) + \Gamma(t+1, t) r(t) \quad (1)$$

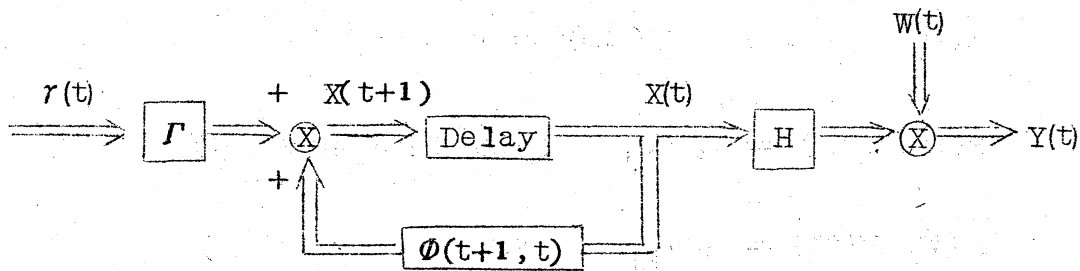
ここで $\Phi(t+1, t)$ は $X(t)$ から $X(t+1)$ への遷移行列。

(2) 観測される random process $Y(t)$ は $X(t)$ と線型に關係する :

$$Y(t) = H(t) X(t) + W(t) \quad (2)$$

但し, $W(t)$ は white noise,

という条件のもとで, random process が non-stationary である場合をふくめて, smoothing, filtering, prediction の解を完全に与え, それらの収束性についても議論した。

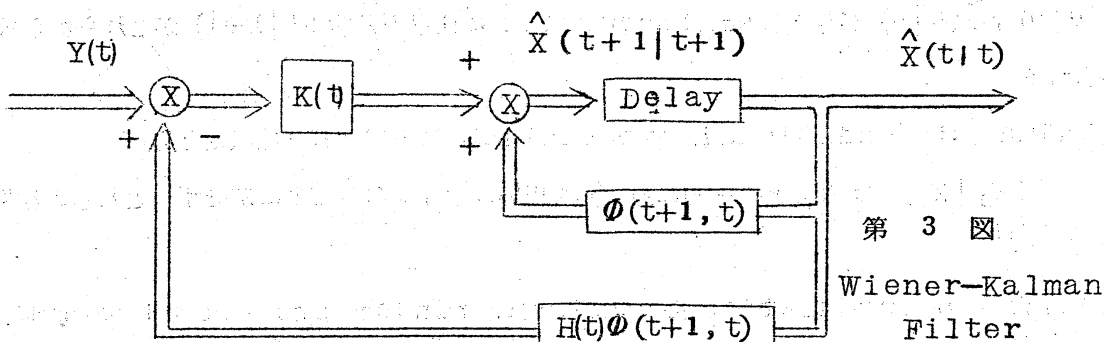


第 2 図

(1)の仮定は, random process が Markov process であるということであり,

(2)の仮定で Y の次元が X の次元より小さいとすること (すなわち H は singular) によつて, 多重 Markov 性を確保出来る。実際, 応用上出てくる random process はほとんどの範疇に属する。

彼の結果は, 例えば stationary case の filtering problem の解は⁽²⁾ (第3図参照)



第 3 図

Wiener-Kalman Filter

$$K(t) = \tilde{P}(t+1|t) H(t) (H(t) \tilde{P}(t+1|t) H(t)^t + R(t))^{-1}$$

$$\hat{X}(k+1|k+1) = \Phi \hat{X}(k|k) + \tilde{P}(k+1|k) H^t (H \tilde{P}(k+1|k) H^t + R)^{-1} \cdot (Y(k+1) - H \Phi \hat{X}(k|k)) \quad (3)$$

$$\tilde{P}(k+1|k+1) = \tilde{P}(k+1|k) - \tilde{P}(k+1|k) H^t (H \tilde{P}(k+1|k) H^t + R)^{-1} \cdot H \tilde{P}(k+1|k) \quad (4)$$

$$\tilde{P}(k+1|k) = \Phi \tilde{P}(k|k) \Phi^t + \Gamma Q \Gamma^t \quad (5)$$

但し,

$$E(r(k)) = 0, \quad E(W(k)) = 0,$$

$$E(r(k) r(j)^t) = Q \delta_{kj} \quad (r \text{ の white noise 性})$$

$$E(W(k) W(j)^t) = R \delta_{kj} \quad (W \text{ の } " \quad)$$

$$E(r(k) W(j)^t) = 0$$

\tilde{P} は error $X - \hat{X}$ の covariance matrix. $A(k+1|k+1)$ とあるのは $k+1$ 時刻迄の観測量が得られた時の $k+1$ 時刻の A の意味。

(3) は optimal estimation (filter) equation, (4) は covariance equation ともよばれる。(5) は extrapolated covariance $\tilde{P}(k+1|k)$ を与える式である。これらは Wiener-Hopf の方程式に対応するもので、a priori な情報 $\hat{X}(0|0), \tilde{P}(0|0)$ が与えられれば recursive に任意の $\hat{X}(k+1|k+1)$ がこれらから求められる。

Kalman は Conditional expectation に関する直交性 (たとえば $X_1 - E(X_1|X_2)$ は X_2 と確率的に独立) を用い、 r, W の分布型に関せず、これらの結果を求めた。⁽¹⁾

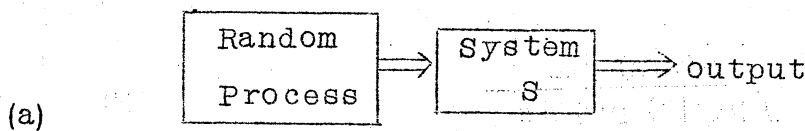
Kalman の formulation の大きな特徴は、random process を 1 つの制御系 (連続オートマトン) としてとらえ、従って、process の「内部構造」をはつきり指定した点

であろう。従つて、filtering等の問題は、state-estimationあるいはstate-identificationの問題にreduceする。

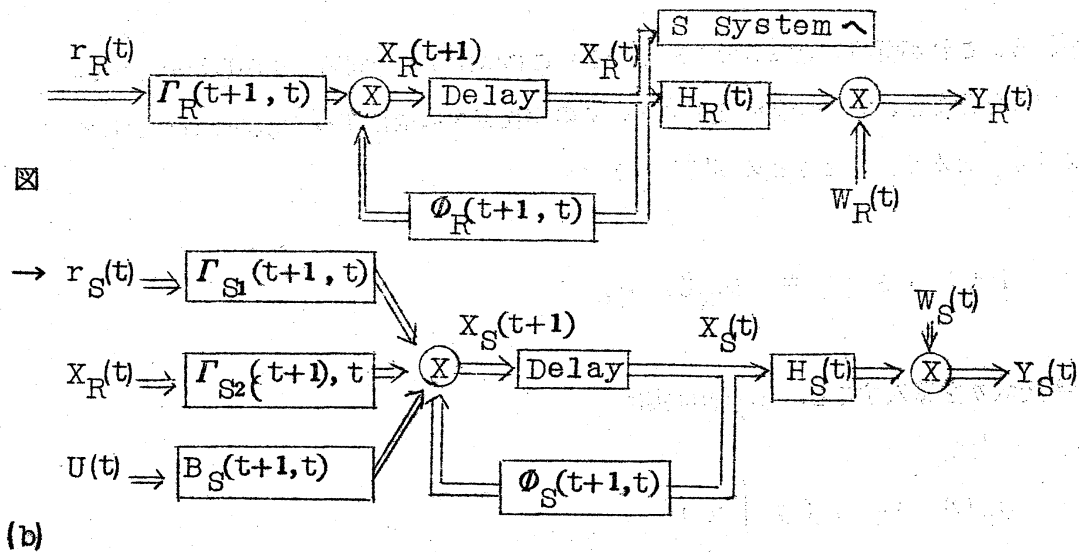
従来の理論、たとえばWienerの理論を実際に用いるには、生のデータYから、stationaryなcovariance functionあるいはpower-spectral densityを先づ求める必要があるが、Kalmanの場合には、それは生のoutput data Yから内部構造をestimateする問題、つまりsystem identificationの問題になる。

この様に、random processを適当なsystem(制御系)のoutputとみなす場合は、我々技術者にとっては非常透明かつ扱いやすく、更に、このprocessがinputされるsystem(制御系)と全く同一形式であるので一貫して扱えることになる。例えば第4図の様に、random process $X_R(t)$ がlinear system Sにadditiveに加わる時は全系は、

$$\begin{bmatrix} X_S(t+1) \\ X_R(t+1) \end{bmatrix} = \begin{bmatrix} \Phi_S(t+1, t) & \Gamma_{S2}(t+1, t) \\ 0 & \Phi_R(t+1, t) \end{bmatrix} \begin{bmatrix} X_S(t) \\ X_R(t) \end{bmatrix} + \begin{bmatrix} B_S(t+1, t) \\ 0 \end{bmatrix} U_S(t) + \begin{bmatrix} \Gamma_{S1}(t+1, t) & 0 \\ 0 & \Gamma_R(t+1, t) \end{bmatrix} \begin{bmatrix} r_S(t) \\ r_R(t) \end{bmatrix}$$



第 4 図



この様に、一般には system S への signal とか noise とかも又 1 つの system と見なせるので、それらを混然一体とした total system の, state-identification あるいは system identification によつて、本来我々が関心をもっている system S の内外における諸種の推定時間を総合的に扱えることになる。従つて以下では system S 内の parameter の推定も、S 外の parameter の推定も区別せず 議論する。先に、r か W を white noise として system を考えたのは、この様な総合的な system を考えてのことであつた。

5. Bayesian Approach (Learning)

例えば、2 確率変数 ξ, η に対して conditional probability は

$$P(x|y) = \frac{P(x, y)}{P(y)}$$

と定義され、従つて

$$P(x|y) = \frac{P(x, y)}{P(y)} = \frac{P(y|x) P(x)}{\sum_x P(y|x) P(x)}$$

となる。density function の定義される場合は、これから

$$p(x|y) = \frac{p(y|x) p(x)}{\int p(y|x) p(x) dx} \quad (1)$$

が容易に導かれる。これが所謂 Bayes の定理である。これは ξ, η の同時分布から出発しているが、逆に、 $p(x)$ を分布としてもつ確率変換 ξ と、 ξ の実現値 x に依存する $p_x(y)$ を分布としてもつ確率変数群 η_x とから、 ξ と η との同時分布が

$$p(x, y) = p_x(y) p(x)$$

となる如き確率変数 η を導入すると、 $p_x(y)$ は実は

$$p_x(y) = p(y|x)$$

になることが示される。従つて、 ξ の a priori 分布 $p_0(x)$ から出発して、形式的に $p_x(y) = p(y|x)$ 等として Bayes の定理をくり返して用いて、次々と新たな確率変数を導入し、形式的に用いた Bayes の定理が正しい意味をもつ同時分布に従つてその各種条件付分布に到達しうる。

従つて一般に、 ξ の a priori 分布を $p_0(x)$ とすると観測値の組 $(y(1), y(2), \dots, y(k)) \triangleq \mathcal{Y}(k)$ を得てその後の x の条件付き分布 $p(x|\mathcal{Y}(k))$ と、更にもう 1 つ観測値がふえた組 $(y(1), y(2), \dots, y(k), y(k+1)) \triangleq \mathcal{Y}(k+1) = (\mathcal{Y}(k), y(k+1))$ を得て後の x の条件付き分布 $p(x|\mathcal{Y}(k), y(k+1))$ との間には

$$p(x|\mathcal{Y}(k), y(k+1)) = \frac{p(y(k+1)|x)p(x|\mathcal{Y}(k))}{\int p(y(k+1)|x)p(x|\mathcal{Y}(k))dx} \quad (2)$$

$$= \frac{\prod_{i=1}^{k+1} p(y(i)|x) p_0(x)}{\int \dots \int \prod_{i=1}^{k+1} p(y(i)|x) p_0(x) dx} \quad (3)$$

以上の process は、観測データが得られて後の、 x の a posteriori 分布を決定するだけで、一義的な推定値を決めるには、この他に適当な評価を導入して "best" なものを選ばねばならない必要がある。^{*}) その様な評価基準としてよく用られるものは

1. most probable estimate \hat{x}_1

(分布の peak を与える x の値)

2. Conditional mean \hat{x}_2

$$\int_{-\infty}^{\infty} \|x - \hat{x}_2\|^2 p(x|\mathcal{Y}(N)) dx$$

3. minimax estimate \hat{x}_3

$$\min_{\hat{x}_3} \cdot \max_x |x - \hat{x}_3|$$

これら $\hat{x}_1, \hat{x}_2, \hat{x}_3$ は分布 $p(x|\mathcal{Y}(N))$ が Gaussian である場合には同じ値になることが知られている。

^{*}) (尤も、Bayes 最適制御では一義的な推定値を決める必要もなく、a posteriori 分布を求めさえすればよい。)

さて、上述の(1), (2)は常に一定の η の条件付き分布を求めるという、static な場合であるが、

$$X(t+1) = f_X(X(t), U(t)) \quad (4)$$

$$Y(t) = f_Y(X(t)) \quad (5)$$

なる制御系の state $X(t_1)$ を推定する所謂 state-identification の問題 では $X(t)$ が dynamic に次々と変って行くために $p(X(k)|Y(k))$ から $p(X(k+1)|Y(k+1))$ を求める必要がある。

そこで $X(k)$ から $X(k+1)$ への transition の関係(4)を用いて

$$p(X(k+1)|Y(k+1)) = \frac{\int p(Y(k+1)|X(k+1)) p(X(k+1)|X(k)) p(X(k)|Y(k)) dx(k)}{\iint " " dx(k) dx(k+1)} \quad (6)$$

ここで、 $Y(k) \triangleq (Y(1), \dots, Y(k))$ 。

あるいは state transition 関係には(4)に見る如く input が関係しているので、もう少し詳しく書けば、 $U(k) \triangleq (U(1), \dots, U(k))$ として

$$p(X(k+1)|Y(k+1); U(k)) = \frac{\int p(Y(k+1)|X(k+1)) p(X(k+1)|X(k); U(k)) p(X(k)|Y(k); U(k-1)) dx(k)}{\iint " " dx(k) dx(k+1)} \quad (6')$$

従つて、Bayesian approach による state-estimation あるいは identification は次の4段階より構成される：

- (1) $p(Y(k+1)|X(k+1))$ を(5)式より求める。
- (2) $p(X(k+1)|x(k); U(k))$ を(4)式より求める。
- (3) $p(X(k+1)|Y(k+1); U(k))$ を(6)あるいは(6')式より求める。
- (4) $p(X(k+1)|Y(k+1))$ なる今布から、評価 1 あるいは 2 あるいは 3 を用いて 'best' estimate \hat{X} を求める。

5-A. Linear Gaussian System の場合

第4節でのべた linear system で, stationary な場合を考えよう:

$$X(t+1) = \Phi X(t) + \Gamma r(t) \quad (1)$$

$$Y(t) = HX(t) + W(t) \quad (2)$$

$r(t), W(t)$ は夫々 white Gaussian noise である。Gaussian であるから, 評価 1, 2, 3 いずれも同じ推定値を与え, linear であるため極めて容易に計算出来, その結果は全く Wiener-Kalman filter のそれに一致する。Wiener-Kalman filter の場合, $r(t), W(t)$ は White だが必ずしも Gaussian とは云わなかつた。しかし, 実は linear filtering problem の解と, $r(t), W(t)$ を Gauss 分布をもつ確率変数とした時の filtering problem の解とは一致をすること, 及び mean 及び covariance からは一義的に Gauss 分布が決ること, そして mean 及び covariance についてのみ議論をしていることから, Gauss 性を前提にした議論であると云つてよい。従つて, linear system に関しては Kalman の方法も Bayesian approach も等価であると云える。Bayesian approach はしかし次に示す様に non-linear system の場合にも ("optimal" とは限らないにせよ) 適用出来るし, discrete hypotheses の問題にも勿論用いられるので, 有力な統一的な方法であると考えられる。

5-B. Non-linear Gaussian System の例⁽³⁾

$$X(k+1) = f(X(k), k) + \Gamma(k) r(k)$$

$$Y(k) = h(X(k), k) + W(k)$$

r, w は平均値 0, Gauss 分布をもつ。更に

$$E\{r(j) r(k)^t\} = Q(k) \delta_{jk} \quad (\text{whiteness})$$

$$E\{W(j) W(k)^t\} = R(k) \delta_{jk} \quad (\text{whiteness})$$

$$E\{r(j) W^t\} = 0$$

このような system において、次の様な filtering problem を考える：

「 $Y(0), \dots, Y(n)$ なる観測結果が与えられて、 $\{X(0), \dots, X(n)\}$ の estimate $\{\hat{X}(0), \dots, \hat{X}(n)\}$ を求めること。」

$$\begin{aligned} p(X(0), \dots, X(n) | Y(0), \dots, Y(n)) \\ &= \frac{p(Y(0), \dots, Y(n) | X(0), \dots, X(n)) p(X(0), \dots, X(n))}{p(Y(0), \dots, Y(n))} \\ &= \frac{\prod_{k=0}^n p_W(Y(k) - h(X(k), k)) p_0(X(0)) \prod_{k=1}^n p(X(k) | X(k-1))}{p(Y(0), \dots, Y(n))} \end{aligned}$$

$p_0(X(0))$ は $X(0)$ に対する a priori distribution で mean vector m , covariance matrix $P(0)$ をもつ Gauss 分布に従うものとする。 $p(X(k) | X(k-1))$ は $r(k-1)$ の Gauss 性から mean vector $f(X(k-1), k-1)$, Covariance matrix $F(k-1) Q(k-1) F(k-1)^t$ をもつ Gauss 分布である。従つて

$$\begin{aligned} p(X(0), \dots, X(n) | Y(0), \dots, Y(n)) \\ &= c(Y(0), \dots, Y(n)) \\ &\quad \cdot \exp\left\{-\frac{1}{2} \sum_{k=0}^n \|Y(k) - h(X(k), k)\|_{R^{-1}(k)}^2 - \frac{1}{2} \|X(0) - m\|_{P^{-1}(0)}^2\right. \\ &\quad \left. - \frac{1}{2} \sum_{k=0}^{n-1} \|X(k+1) - f(X(k), k)\|_{[GQG^t]^{-1}}^2\right\} \end{aligned}$$

但し $\|V\|_A^2 \triangleq V^t A V$ (2次形式)

そこで先にのべた Bayesian approach における、unique estimate を定めるための評価法の「1: most probable estimate」を採用すれば

$$J_n = \frac{1}{2} \|X(0) - m\|_{P^{-1}(0)}^2$$

$$\begin{aligned}
& + \frac{1}{2} \sum_{k=0}^n \| Y(k) - h(X(k), k) \|_{R^{-1}(k)}^2 \\
& + \frac{1}{2} \sum_{k=0}^{n+m-1} \| X(k+1) - f(X(k), k) \|_{[GQG]^t}^2
\end{aligned}$$

を最小にする $\{ X(0), \dots, X(n) \}$ が most probable estimate である。

prediction problem ;

「 $Y(0), \dots, Y(n)$ なる観測結果が与えられて、 $\{ X(0), \dots, X(n), \dots, X(n+m) \}$ の estimate $\{ \hat{X}(0), \dots, \hat{X}(n+m) \}$ を求めること。」

$$p(X(0), \dots, X(n+m) \mid Y(0), \dots, Y(n))$$

$$= C'(Y(0), \dots, Y(n))$$

$$\begin{aligned}
& \cdot \exp \left\{ -\frac{1}{2} \sum_{k=0}^n \| Y(k) - h(X(k), k) \|_{R^{-1}(k)}^2 - \frac{1}{2} \| X(0) - m \|_{P^{-1}(0)}^2 \right. \\
& \left. - \frac{1}{2} \sum_{k=0}^{n+m-1} \| X(k+1) - f(X(k), k) \|_{[GQG]^t}^2 \right\}
\end{aligned}$$

従つて, most probable estimate は

$$\begin{aligned}
J_{n,m} & = \frac{1}{2} \| X(0) - m \|_{P^{-1}(0)}^2 \\
& + \frac{1}{2} \sum_{k=0}^n \| Y(k) - h(X(k), k) \|_{R^{-1}(k)}^2 \\
& + \frac{1}{2} \sum_{k=0}^{n+m-1} \| X(k+1) - f(X(k), k) \|_{[GQG]^t}^2
\end{aligned}$$

を最小にする $\{ X(0), \dots, X(n+m) \}$ である。

ところで

$$J_{n,m} = J_n + \frac{1}{2} \sum_{k=n}^{n+m-1} \| X(k+1) - f(X(k), k) \|_{[GQG]^t}^2$$

であるから、 $J_{n,m}$ を minimize することは明かに J_n を先づ minimize し、しかる后第 2 項を 0 にすることに他ならない。即ち

$$X(k+1) = f(X(k), k) \quad k = n, n+1, \dots, n+m-1$$

いかえれば、filtering problem を解き、后はそれよりの noise を取去つた 単なる extrapolation に他ならない。

Min. $\{J_n\}$ の Dynamic Programming による解

$[GQG^t]^{-1}$ が存在するものとする。便宜上、1 step 先迄の prediction problem ;

$$\text{Min} \{ 2 J_{n,1} \}$$

を考える。その様にしても先にのべた様に、filtering problem の解には影響はない。

$$F_0(X(0)) \triangleq \| X(0) - m \|_P^{-2}(0)$$

$$F_{k+1}(X(k+1)) \triangleq \text{Min}_{X(0), \dots, X(k)} \{ 2 J_{k,1} \}, \quad k = 0, 1, \dots, n$$

を定義すると、

$$F_{k+1}(X(k+1)) = \text{Min}_{X(k)} \left\{ F_k(X(k)) + \| X(k+1) - f(X(k), k) \|_{[GQG^t]^{-1}}^2 + \| Y(k) - h(X(k), k) \|_{R^{-1}(k)}^2 \right\}$$

($k = 0, 1, \dots, n$)

なる recurrence equation が成立つ。k=0 から始めて順次、most probable estimate $\hat{X}(k)$ をこれから求めることが出来る。但し、 $X(n+1)$ の estimate は $F_{n+1}(X(n+1))$ を minimize する $X(n+1)$ 、 $X(n)$ の estimate (我々の求める filtering problem の解) は $F_n(X(n)) + \| Y(n) - h(X(n), n) \|_{R^{-1}(n)}^2$ を minimize する $X(n)$ である。

Dynamic programming による解は、最新の $F_n(X(n))$ さえ保存しておけば、更に step が進行する時には $F_n(X(n))$ から考慮すればよく、それ以前の $X(k)$ は考える必要がないという利点をもっている。($F_n(X(n))$ が「状態」の役目をする！)

6. System Identification

System identification として、ここでは system の未知パラメータの推定問題を考える。その parameter が時間的に変化しない場合 (time invariant) と、変化する場合 (time variant) とがあるが、前者の方が当然推定が容易であると同時に、しばしばより本質的である。というのは、system のより本質的な性質は、時間にかかわらず成立つといつた形であられる筈で、法則自身に時間がふくまれていても、それにふくまれている parameter は time invariant な形をとるであろうから。そういうより長期間にわたって成立つであろう様な本質的な性質の identification はむしろ学習 (learning) という名に値するものであろうが、一方、もつと身近な各種自動制御系では time variant な parameter の推定がしばしば問題になる。すなわち、system の一部の parameter が運転に際して予め設計の段階では予測のつかない具合に、そして時間と共に変動する場合は実際問題に多い。そういう場合には state identification と同じく、「完全には」 identify 出来ず、興味ある問題を生ずるが、ここではこれらにはふれず、time invariant な parameter のみを対象にする。

7. Bayesian System Identification^{5),6),7)}

7-1. Linear Convertible System

$$X(t+1) = \Phi X(t) + BU(t) + r(t) \quad (1)$$

$$Y(t) = \Psi X(t) + W(t) \quad (2)$$

$r(t)$, $W(t)$ は white Gaussian 分布をもつとする。 Ψ^{-1} が存在する system は convertible であるという。今、 Φ と、 r の平均値 \bar{r} とが未知であるとする。

(1), (2) から

$$Y'(t+1) = \Phi X(t) + BU(t) + r'(t) \quad (3)$$

を得る。但し、

$$Y(t+1) \triangleq \Psi^{-1} Y(t+1)$$

$$r'(t) \triangleq r(t) + \Psi^{-1} w(t+1)$$

$Y(t)$ は(3)から、 $\Phi X(t-1) + BU(t-1) + r'(t-1)$ によつてあらわされる母集団から発生したサンプルと考えられるが、その代りにそれが

$$\Phi X^{\hat{e}}(t-1) + BU(t-1) + r'(t-1)$$

によつてあらわされる母集団からのサンプルであると近似的に考える。すなわち、 $p(Y(t) | X(t-1))$ を $p(Y(t) | X^{\hat{e}}(t-1))$ で近似する。ここで $X^{\hat{e}}(t-1)$ は $X(t-1)$ の estimate である。

$$\phi^* = [\phi_1, \bar{r}'_1; \phi_2, \bar{r}'_2; \dots; \phi_N, \bar{r}'_N]^t$$

の a priori distribution を Gaussian として、mean vector を \bar{m}_0 、covariance matrix を $C(0)^{-1}$ とする。但し、

$$\phi_i = [\phi_{i1}, \phi_{i2}, \dots, \phi_{iN}] \dots$$

そうすると Bayesian learning (Bayes 定理の逐次使用) によつて次の recurrence relation をうる；

$$C(t) = C(t-1) + D(t-1) \tag{4}$$

$$\bar{m}(t) = C(t)^{-1} \{ C(t-1) \bar{m}(t-1) + A^{\hat{e}}(t-1) H(t-1) \} \tag{5}$$

$$= C(t)^{-1} \{ C(t-1) \bar{m}(t-1) + A^{\hat{e}}(t-1) T A(t-1)^t Z(t-1) \} \dots \tag{6}$$

superscript 't' は転置を示す。

$$\Delta^e(k) = \begin{bmatrix} \delta^e(k) & & & \\ & \ddots & & \\ & & \delta^e(k) & \\ & & & \delta^e(k) \end{bmatrix}, \quad \Delta(k) = \begin{bmatrix} \delta(k) & & & \\ & \ddots & & \\ & & \delta(k) & \\ & & & \delta(k) \end{bmatrix}$$

$$\delta^e(k) = [X_1^e(k), \dots, X_N^e(k), 1]^t$$

$$\delta(k) = [X_1(k), \dots, X_N(k), 1]^t$$

$$D(k) = \Delta^e(k) \cdot T \cdot \Delta^e(k)^t$$

T : r' の covariance matrix の inverse.

$$H(k) = [H_1(k), \dots, H_N(k)]^t$$

$$H_1(k) = \sum_{j=1}^N T_{ij} (y'_j(k+1) - \sum_p B_{jp} U_p(k))$$

$$Z(k) = [\phi_1, r'_1(k); \phi_2, r'_2; \dots; \phi_N, r'_N(k)]^t$$

このidentification process (4), (5), (6)には X(k)の estimate $X^e(k)$ がふくまれているので, 実はsystem identificationと同時にstate-identification (filtering) を行わねばならない。しかし, system の parameter が完全にはわかっていないから, 先にのべたstate-identification そのものにはならない。この様なsystem についての不完全知識下でのstate-identification は, たとえば,

1. estimated system parameter ($\bar{m}(t)$) を用いてstate-identification (Wiener-Kalman filter を用いて) を行う。

或いは

2. estimated state (既知system parameter の条件下での) を, system parameter の a posteriori distribution で平均する。

によって行う。

結局, 我々のsystem identification process ではstate-identifi-

cation が常に先行する。

この process は ϕ^* の a priori distribution か, $r'(t)$ の distribution を Gaussian として導いたが, 実はその様な制限なしに次の様な十分条件のもとで確率収束することが示される。^{5),6)}

「 $\phi_{ij}, \bar{r}',$ a priori $\alpha(0), \bar{m}(0)$ は有界とする。

次の条件を満す無限時間系列,

$$0 = t_0 < t_1 < t_2 < \dots$$

を確率 1 でえらべるとする:

(A・1) $N+1$ 個の 1 次独立な δ^e -vectors が t_{j-1} から $t_j - 1$ の区間にふくまれている。($j=0, 1, \dots$)

(A・2) $t_j - t_{j-1}$ は有界である。

$$(A \cdot 3) \quad \lambda_{\min} \left[\sum_{j=1}^n \left(\sum_{p=t_{j-1}}^{t_j-1} D(p) \right) \right] \geq M_1 \cdot n^q$$

である様な $M_1 > 0, q > 1$ が存在する。

但し, $\lambda_{\min}[A]$ は positive definite matrix A の最小固有値。

(A・4) $E(\|\delta^e(k)\|^2)$ が任意の k に対し有界,

(A・5) $E(R(k)^2)$ が任意の k に対し有界,

但し,

$$R(k)^2 = \sum_{i=1}^N \left\{ (r_i'(k) - \bar{r}_i') + \sum_{j=1}^N \phi_{ij} (X_j(k) - X_j^e(k)) \right\}^2 .$$

しからは

$$\text{Pr.} \left\{ \lim_{n \rightarrow \infty} \| \alpha(n)^{-1} \| = 0 \right\} = 1 ,$$

$$\text{Pr.} \left\{ \lim_{n \rightarrow \infty} \| \bar{m}(n) - \phi^* \| = 0 \right\} = 1 . \quad \square$$

(A・1) 乃至 (A・5) の条件は "制御可能な" system では, 通常, きわめて自然に成立

つものである。

この証明には stochastic approximation の考えが関連する。

7-2. Approximations

(I) (3)式は

$$y'_i(t+1) = \sum_{j=1}^N \phi_{ij} X_j(t) + B_i U(t) + r'_i(t) \quad i = 1, \dots, N$$

であるが、これからは $\{\phi_{i1}, \phi_{i2}, \dots, \phi_{iN}, \bar{r}'_i\}$ のみを identify するという具合に、各行毎に別々に分割して identification を行うことにすれば、7-1にくらべて計算量ははるかにへらせる。

$\{\phi_{i1}, \dots, \phi_{iN}, \bar{r}'_i\}^t$ の a priori distribution を Gaussian として、その mean vector, covariance matrix を夫々 $\bar{m}_i(0), C_i(0)^{-1}$ とすれば

$$C_i(t) = C_i(t-1) - D_i(t-1)$$

$$\begin{aligned} \bar{m}_i(t) &= C_i(t)^{-1} \left\{ C_i(t-1) \bar{m}_i(t-1) + \frac{1}{\sigma_i'^2} (y'_i(t) - \sum B_{ij} U_j(t-1)) \delta^e(t-1) \right\} \\ &= C_i(t)^{-1} \left\{ C_i(t-1) \bar{m}_i(t-1) + \frac{1}{\sigma_i'^2} \delta^e(t-1) \delta^e(t-1)^t Z_i(t-1) \right\} \\ &\quad (i = 1, \dots, N) \end{aligned}$$

$$\text{但し, } \delta(k) = [X_1(k), \dots, X_N(k), 1]^t$$

$$\delta^e(k) = [X_1^e(k), \dots, X_N^e(k), 1]^t$$

$$D_i(k) = \frac{1}{\sigma_i'^2} \delta^e(k) \delta^e(k)^t$$

$$\sigma_i'^2 = r'_i \text{ の variance}$$

この方式の収束性も、7-1に全く analogous で、ただ (A・5) が形式上

(A・5') $E(R_i(k)^2)$ が任意の k に対し有界、

但し,

$$R_1(k) = |(r_1'(k) - \bar{r}_1') + \sum_{j=1}^N \phi_{ij}(X_j(k) - X_j^0(k))|$$

(i = 1, \dots, N)

となるだけである。

(四) (I)は横わりのdecompositionであるが、たてわりのdecompositionを行つても同様に収束することが見られる。これによつて計算がより簡素化する。このことはsystemの規模が大きくなると特に重要になる。

この様な最も細いdecomposition 迄行く途中に、もつと大まかな多くのdecompositionの形式が可能である。そして、適当なdecompositionを状況に応じadaptiveに選択する方式などが大規模systemのidentificationでは重要となろう。

7-3. Non-linear System

7.1でのべたregression formのidentificationは、勿論、non-linear systemにも適用可能である。ここでは観測機構(XとYとの関係)もnon-linearである場合を考える。

$$X(t+1) = \Phi X(t) + F(X(t)) + BU(t) + r(t) \quad (1)$$

$$Y(t) = \Psi X(t) + G(X(t)) + W(t) \quad (2)$$

ここで

$$F(X(t)) = A \cdot V(t)$$

$$G(X(t)) = S \cdot V(t)$$

$A = [a_{ij}]$, $S = [s_{ij}]$ は $N \times M$ -matrix で

$$V = [\varphi_1(X(t)), \varphi_2(X(t)), \dots, \varphi_M(X(t))]^t$$

$\varphi_1, \varphi_2, \dots, \varphi_M$ は $X = [X_1, \dots, X_N]^t$ の既知関数で,

$$1, X_1, X_2, \dots, X_N, \varphi_1(X), \varphi_2(X), \dots, \varphi_M(X)$$

は 1 次独立であるとする。

$$Y(t+1) \triangleq \Psi^{-1} Y(t+1)$$

$$r'(t) \triangleq r(t) + \Psi^{-1} W(t+1)$$

とおいて, (1), (2)より

$$Y(t+1) = \Phi X(t) + A \cdot V(t) + (BU(t) + \Psi^{-1} SV(t+1)) + r'(t) \quad (3)$$

$Y'(t+1)$ が

$$\Phi X^e(t) + A \cdot V^e(t) + (BU(t) + \Psi^{-1} SV^e(t+1)) + r'(t)$$

で特徴づけられる母集団から発生するものと近似的に考えて, 7-1あるいはその近似の 7-2 と形式上全く同様な identification process が得られる。但し,

$$\delta^e(k) = [X_1^e(k), \dots, X_N^e(k), \varphi_1^e(X(k)), \dots, \varphi_M^e(X(k)), 1]^t$$

等のちがいがある。

収束のための十分条件も単なる 7-1 の拡張によつて得られる。

7-4. Non-convertible System

これは観測機構 (X と Y との関係) が一般に

$$Y(t) = g(X(t)) + W(t)$$

とあらわせるが, g^{-1} が存在しない場合である。non-linear の観測機構といつても 7-3 で記した形,

$$Y(t) = \Psi X(t) + S \cdot V(t) + W(t)$$

の様に linear term $\Psi X(t)$ をもつ場合に限れば, 各 $X_i(t)$ ($i=1, \dots, N$) が観測量 $y_j(t)$, ($j=1, \dots, K$) といずれかと相関がありさえすれば, 我々の system identification は可能である。

簡単のため linear の観測機構 :

$$Y(t) = \Psi X(t) + W(t)$$

で考えよう。たとえば $y_j(t)$ と $X_i(t)$ とが相関がある場合, すなわち Ψ の (j, i) 要素 ψ_{ji} が non-zero である場合,

$$\frac{1}{\psi_{ji}} y_j(t) = x_i(t) + \sum_{k \neq i} \frac{\psi_{jk}}{\psi_{ji}} X_k(t) + \frac{1}{\psi_{ji}} W_j(t)$$

従つて

$$\frac{1}{\psi_{ji}} y_j(t) \triangleq y'_i(t)$$

と定義すれば

$$y'_i(t+1) = \sum_{j=1}^N \phi_{ij} X_j(t) + \sum B_{ij} U_j(t) + \sum_{k \neq i} \frac{\psi_{jk}}{\psi_{ji}} X_k(t) + r'_i(t)$$

をうる。但し

$$r'_i(t) \triangleq \frac{1}{\psi_{ji}} W_j(t+1) + r_i(t)$$

従つて, 例の如く $y'_i(t+1)$ が

$$\sum_{j=1}^N \phi_{ij} X_j^e(t) + \sum B_{ij} U_j(t) + \sum_{k \neq i} \frac{\psi_{jk}}{\psi_{ji}} X_k^e(t) + r'_i(t)$$

から発生するものと近似的に考えて, 先の議論と analogous に話は進む。

さて, 再び linear invertible system にもどろう。この場合には, 実は, 今迄の様な state-identification とは独立に system identification が行える:

$$X(t+1) = \Phi X(t) + BU(t) + \Gamma r(t) \quad (1)$$

$$Y(t) = \Psi X(t) + W(t) \quad (2)$$

(2)から

$$X(t+1) = \Psi^{-1} Y(t+1) - \Psi^{-1} W(t+1)$$

$$X(t) = \Psi^{-1} Y(t) - \Psi^{-1} W(t)$$

を(1)に入れ、

$$\Psi^{-1} Y(t) \triangleq Y'(t) \quad (3)$$

$$\Psi^{-1} Y(t+1) \triangleq Y'(t+1)$$

と定義すれば、 X の transition equation (1)の代りに Y' の transition equation をうる：

$$Y'(t+1) = \Phi Y'(t) + BU(t) + r'(t) \quad (4)$$

但し

$$r'(t) = \Gamma r(t) + \Psi^{-1} W(t+1) - \Psi^{-1} W(t) \quad (5)$$

r' には unknown Φ をふくんでいるが、これをその estimate $\hat{\Phi}$ (たとえば Φ の a posteriori 分布の平均値) で近似すれば、今迄の議論と全く同様に正しい収束をする identification process をうる。更に、実は $\Phi \Psi^{-1}$ が有界でありさえすれば正しい収束が確保される。(stochastic approximation) ただ、今迄のと概念上異なる点は、(4)という(1)と "等価な" モデルに話を切りかえた所である。その新しいモデルでは state は $Y'(t)$ である。

一般に system identification の立場に 2通りのものがある。その 1つは、(1)、(2)という system の基本式が与えられた時 (モデルが与えられた時)、常にそのモデルに固執して、それにふくまれる parameter を推定する立場である。

もう一つは、我々は system の input, output しか観測出来ず、従つて input, output の関係を正しく表現出来るモデルならどれをとつても差支えないという立場である。

第1の立場の方が第2のそれにくらべて融通性に欠ける様に見えるが、第2の立場の方が常にすぐれているとも云えない。第1の立場は、かなり system の物理的内容、意味がつかめている場合で、従つて「状態」は直接観測出来なくても物理的に十分意味のある量でなければならない。そしてそのようなモデル(1)は観測系(2)をあるクラスの中で変えても(実験方法を変えても)変らないようなものとする。

これに対して第2の立場では、system の物理的内容、意味についての知識が低いからあるいは考えうる考えうる自由度が大きく、「状態」は形式的(あるいは数学的)な量と考える。この立場に関連して出てくる重要な概念は勿論、system の等価性 (equivalence) である。つまり、同じ input-output の関係をもつ system への分類である。そしてそれから1つの equivalence class の中で最も「単純な」system は何かという問題が重要になる。(有限オートマトンではそれを「minimal」automoter という。)

その様な議論が可能ではあるが、一方、第2の立場は特殊な観測系(2)に完全に依存するという狭さをもっている。観測系が変れば、等価 system のクラスわけは変つてしまう。そういう意味では、第2の立場は「単能的な」、一面的な identification にしかならないとも云えよう。

今迄はむしろ第1の立場での議論であつたが、今度議論しようというのは第2の立場からである。

さて、(2)における Ψ の inverse が存在しない場合、すなわち linear non-conver-
tible system を考える。例えば極端に、 Ψ が 1 行 N 列の matrix であるとしよう。即ち、system は N 次元だが、観測は 1 時には唯 1 つしか得られない場合である。この時、新しい「state」 $Y'(t)$ を導入して(4)に似た、(1), (2) に等価な Y' の transition equation をどの様にしてつくれるか。

1 つの方法は、 X が N 次元で Y が 1 次元だから、 $X(t)$ を $[Y(t), Y(t+1), \dots, Y(t+N-1)]^t$ でおきかえるようにすることであろう。

$$X(t+n) = \phi^n X(t) + \sum_{k=0}^{n-1} \phi^{n-k-1} [BU(t+k) + \Gamma r(t+k)]$$

が(1)より成立つから

$$Y(t) = \Psi X(t) + W(t)$$

$$Y(t+1) = \Psi \Phi X(t) + \Psi [BU(t) + \Gamma r(t)] + W(t+1)$$

$$Y(t+N-1) = \Psi \Phi^{N-1} X(t) + \Psi \sum_{k=0}^{N-2} \Phi^{N-k-2} [BU(t+k) + \Gamma r(t+k)] + W(t+N-1)$$

$$\begin{pmatrix} Y(t) \\ Y(t+1) \\ \vdots \\ Y(t+N-1) \end{pmatrix} \triangleq Y'(t), \quad \begin{pmatrix} \Psi \\ \Psi \Phi \\ \vdots \\ \Psi \Phi^{N-1} \end{pmatrix} \triangleq T,$$

$$\begin{pmatrix} W(t) \\ \Psi [BU(t) + \Gamma r(t)] + W(t+1) \\ \vdots \\ \Psi \sum_{k=0}^{N-2} \Phi^{N-k-2} [BU(t+k) + \Gamma r(t+k)] + W(t+N-1) \end{pmatrix} \triangleq A(t)$$

とにおいて

$$Y'(t) = TX(t) + A(t)$$

若し、 T^{-1} が存在すれば、 $X(t)$ は $[Y(t), \dots, Y(t+N-1)]^T$ で等価的におきかえられる。

この T^{-1} が存在することをsystemは可観測 (Observable) であるという言い方をす
る。(Kalman) これは特別な場合で、一般には $K \times N$ matrix Ψ を観測機構がもつと
き

$$\begin{pmatrix} \Psi \\ \Psi \Phi \\ \vdots \\ \Psi \Phi^m \end{pmatrix}$$

のrankがNとなる様なmが存在する時、systemはObservableであるという。(つま

り、 m 時刻にわたる K m ケの観測データが一般に N 次元空間を張る時)

さて、 T^{-1} が存在すれば

$$X(t) = T^{-1} Y(t) - T^{-1} A(t)$$

これを X -transition equation に入れば Y -transition equation をうる:

$$Y(t+1) = \Phi^* Y(t) + T B U(t) + T F r(t) + A(t+1) - \Phi^* A(t) .$$

$$\Phi^* = T \Phi T^{-1}$$

$$= \begin{bmatrix} 0 & & & & \\ \vdots & & & & \\ 0 & & \Pi & & \\ \varphi_1 & \dots & & \dots & \varphi_N \end{bmatrix} , \quad \Pi \text{ は } (N-1) \text{ 次元単位matrix .}$$

特に zero input ($U=0, r=0$) の時には

$$Y(t+N) = \sum_{i=1}^N \varphi_i Y(t+i-1) + W(t+N) - \sum_{i=1}^N \varphi_i W(t+i-1)$$

となつて、higher order の微分方程式で表わされる system と同様、unknown parameter は N^2 ケの代りに N ケとなる。

7-5 諸問題点

ここでは内容は省略するが system identification に関連して、研究されつつあるかあるいは将来の重要な課題などを思いつくままに列挙すれば次の様なものがある。

- (1) Stochastic approximation の応用
- (2) Identification の convergence "speed" を上げる様な control .
- (3) Identification の convergence を早める様な観測系 (実験) の逐次選択 (逐次実験計画法) .
- (4) Adaptive Identification (Ex. Adaptive Filter) .
- (5) 多重 Identification .

- (6) Identificationにおけるdecompositionの問題。
- (7) Identifiabilityの一般的な議論。
- (8) Systemへのcontrolの効果のidentification.
- (9) Non-stationary (time-variant) systemのidentification.
- (10) Random parameterのidentification.

文 献

- (1) R.E. Kalman: New Methods in Wiener Filtering Theory, Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability, pp.270-388, 1963, John Wiley.
- (2) R.C.K. Lee: Optimal Estimation, Identification, and Control, 1964, M.I.T. Press.
- (3) H. Cox: On the Estimation of State Variables and Parameters for Noisy Dynamic Systems. IEEE Trans. on Automatic Control, pp.5-12 Jan. 1964.
- (4) M. Aoki: Optimal Bayesian and Minimax Controls of a Class of Stochastic and Adaptive Dynamic Systems, Preprint of IFAC Tokyo Symposium, II-3, August, 1965.
- (5) T. Fukao: System Identification by Bayesian Learning Process, Preprint of IFAC Tokyo Symposium, IV-1, August 1965.
- (6) T. Fukao: System Identification by Bayesian Learning Process (I), 電気試験所彙報, vol.29, No.5, 1965年5月.
- (7) T. Fukao: System Identification by Bayesian Learning Process (II), (III). 電気試験所彙報近刊.
- (8) R. Bellman: Adaptive Control Process-A Guided Tour, Princeton U.P. 1961).
- (9) 深尾: 適応制御過程の諸問題について(I) 電気試験所彙報, vol.26, No.1,

1964年1月.

- (10) M. Aoki: Dynamic Programming Approach to a Final value Control System with a Random Variable Having Unknown Distribution Function, IRE Trans. Automatic Control, pp.270-282, 1960.
- (11) Y.C. Ho, R.C.K. Lee: Identification of Linear Dynamic Systems, Information and Control, vol.8, No.1, February, 1965. pp.93-110.