モデルなしＬＱ学習制御 (研究室資料)

河村嘉顯　大阪電気通信大学工学部電子工学科

1　ＬＱ最適制御

1.1　簡単な制御問題

本研究室で開発したモデル無しＬＱ学習制御の理解には状態変数と現代制御理論において中心的役割を果たしてきたＬＱ最適制御の考えが不可欠である。実際、モデル無しＬＱ学習制御はこれらを用いる。このため、最初にＬＱ最適制御 [1]〜[4] について説明する。

最初に，ごく簡単な例により，問題の概略を理解しておこう。離散時間 $t = 0, 1, 2, \cdots$ において定義された１状態システム

$$x(t+1) = x(t) + u(t) \tag{1}$$

を考えよう．ここで，$x(t)$ はスカラ値の状態変数と呼ばれ，その変化は入力 $u(t)$ により制御される．ここでは，制御入力を $u(t) = gx(t)$ と与えることとし，制御の目的は状態 $x(t)$ を安定に原点 0 に近づけることとする．この制御の良し悪しをより正確に判断するために, 評価関数

$$J = \sum_{t=0}^{\infty} [\{x(t)\}^2 + 2\{u(t)\}^2] \tag{2}$$

を導入する．この値が小さいほど, 状態および入力は原点に近い値をとることになる．したがって，ここでは J を最小とするように, フィードバックゲイン g を決定しよう．

制御入力を (1) 式の状態方程式に代入すると，

$$x(t+1) = (1+g)x(t) \tag{3}$$

である．このとき，$x(0)$ を基準として，$t = 0, 1, 2, \cdots$ を順次代入すると，

$$x(t) = (1+g)^t x(0) \tag{4}$$

が得られる．すなわち、応答は項比 $(1+g)$ の等比数列である。これを (2) 式の評価関数に代入し, 等比級数の和の公式を用いて,

$$J = \sum_{t=0}^{\infty} [(1+g)^{2t} + 2g^2(1+g)^{2t}]x(0)^2$$

$$= \sum_{t=0}^{\infty}(1+2g^2)(1+g)^{2t}x(0)^2$$
$$= \frac{1+2g^2}{1-(1+g)^2}x(0)^2 \tag{5}$$

が得られる．ただし、最下式は和が存在する場合の結果である。

以上より，J は g の値によって異なった値をとる．例えば，$g=-0.2$ のとき, 等比級数の項比は 0.8 となり，(4) 式は 0 に収束する減衰級数であり，(5) 式は有限の和を持つ．一方，$g=0.2$ のとき，項比は 1.2 であり，(4) 式は発散級数であり，有限の和を持たない．正確にいえば，g が開区間 (-2,0) 上にあるときのみ，J は有限の値を持つ．この区間内では，J は g に関して滑らかな連続関数であり，g がこの区間の両端に近いほど大きい値をとり，-2 あるいは 0 の極限において $J\to\infty$ になる．これより，J はこの区間上の少なくとも 1 点において最小値をとる．また，この点において，J の勾配 dJ/dg は 0 となる．なぜならば, この点において, J の勾配が正の値を持つならば, この点より左側において J はより小さい値を持ち，また負の値を持つならば, 右側においてより小さい値を持つ．このため．J がこの点において最小であることに矛盾するからである．

ここで,(5) 式を微分すると,

$$\frac{dJ}{dg} = \frac{4g\{1-(1+g)^2\}-(1+2g^2)\{-2(1+g)\}}{\{1-(1+g)^2\}^2}x(0)^2$$
$$= \frac{(1-g)(1+2g)}{\{1-(1+g)^2\}^2}x(0)^2 \tag{6}$$

である．したがって，$dJ/dg=0$ の解として $g=-0.5$, と 1 が得られるが，区間 $(-2,0)$ 上の解は-0.5 だけである．したがって、J を最小とするゲイン g_{opt} と，これを (5) 式に代入した J の最小値 J_{min} は以下のようになる．

$$g_{opt} = -0.5 \tag{7}$$
$$J_{min} = 2x(0)^2 \tag{8}$$

なお，以上で扱った問題は時不変状態フィードバックゲイン g の最適化である．したがって，g を最適化して得られる入力 $u(t)$ があらゆる入力の中で，評価関数 J を最小にするか否かは明らかでない。以下において、この点を説明する．

1.2 LQ 問題

例題の制御では，設計者が具体的に制御の方法を指定したのではなく，望ましい制御を評価関数の形で与えることにより，具体的な制御は理論的に誘導されることが特徴といえる．このような制御法は一般に最適制御と呼ばれ

る．とくに，線形システムに対して 2 次形式の評価関数を最小にする制御入力を求める最適制御はＬＱ (linear-quadratic) 制御と呼ばれ，1960 年頃に米国の Kalman によって定式化された．すなわち，最適制御は可能なありとあらゆる制御入力の中で，評価関数をこれ以上は決して小さくできないような (この意味で最適な) 制御といえる．ただし，最適制御は初期状態 $x(0)$ が異なると，異なった制御が最適制御となる．この意味で，最適制御は初期状態 $x(0)$ の関数である。

例題の手法をそのまま高次システムに拡張し、評価関数が決してこれ以上小さくなりえない最適制御を求めることが容易とは考えられない．以下の関係は Kalman の結果を基にしている．ただし、LQ 最適制御は微分方程式で表される連続時間システムに対して記述されるのが標準的であるが [1]，本文では，コンピュータでの処理を前提として，離散時間システムについて [2], [3]，[4] 記述する．

上記の評価関数では，評価関数が状態 $x(t)$ の 2 次形式だけでなく，入力 $u(t)$ の 2 次項を含むことに注意されたい．もし，状態をできるだけ原点に近づける事のみを重視すれば，評価関数を $x(t)^2$ の項のみから構成すればよく, $u(t)^2$ の項は不必要に思われるかもしれない. 実際, 上の例では，$u(t) = -x(t)$ とするとき，初期状態にかかわらず 1 クロック ($t = 1$) において，状態は 0 となり，$x(t)^2$ の評価はは最小化される．このように，状態を有限期間内に 0 に一致させる制御は有限製定制御 (デッドビート制御）[3] と呼ばれている．しかしながら，たとえば高速に走行する自動車をほぼ一瞬で止めることは不可能に近い．すなわち，この制御はより複雑な制御問題に対して，しばしば非現実的に大きい制御入力を用いて, 強引に状態を 0 に近づける制御となる．また、制御入力が無限大に発散する場合も少なくない．さらに，入力数より出力数の多い場合などでは，どのような制御を用いても一瞬に状態を 0 にすることはできない．これに対して, 上の例のように入力の大きさも考慮して, 入力と状態のバランスをとりながらこれらの和を最小とする制御は状態を最小とする制御の最良近似解を求める問題であり，はるかに広いクラスの線形システムに対して簡単で実用的な制御となることが知られている．実際、上の例題では，$u(t) = -x(t)$ ではなく，$u(t) = -0.5x(t)$ が解となっている。

以下では，最初に有限期間の LQ 最適制御を、次に無限期間の LQ 最適制御を説明する。

1.3 ＬＱ制御の定式化

最初に LQ 最適制御について、よく知られた標準的な関係を説明する．離散時間（$t = 0, 1, 2, \cdots, L$）において定義された線形システム

$$x(t+1) = Ax(t) + Bu(t) \tag{9}$$

を考えよう．ただし $x(t)$ は状態変数 (通常はベクトル) であり，$u(t)$ は制御入力である．このシステムの状態をできるだけ 0 にする基礎的な制御問題はレギュレータ問題と呼ばれる。このシステムに対し，２次形式評価関数

$$J(\tau, L) = \sum_{t=\tau}^{L-1} \{x(t)^T Q x(t) + u(t)^T R u(t)\} + x(L)^T P_0 x(L). \tag{10}$$

を導入する．ただし，$Q = C^T C \geq 0$, $P_0 \geq 0$（半正定行列）であり，$R = D^T D > 0$（正定行列）とする．これにより、評価関数は負にはならない。いま、初期状態 $x(0)$ が与えられたとして、可能なあらゆる入力系列の中で、この評価関数をこれ以上決して小さくすることができない入力系列を考えよう．上記評価関数を最小化する制御入力を求める問題はＬＱ（linear-quadratic）問題と言われる．ここでは，一般的なシステムについてＬＱ最適制御を求めよう．

1.4　定理

［定理１－１］ ＬＱ問題の解は線形時変状態フィードバック

$$u(t) = G(t)x(t) \tag{11}$$

によって与えられる．ただし，状態フィードバックゲイン $G(t)$ は終端条件

$$P(L) = P_0 \tag{12}$$

を指定したリッカチ行列差分方程式

$$\begin{aligned} P(t) &= A^T P(t+1)A + Q \\ &- A^T P(t+1)B\{B^T P(t+1)B + R\}^{-1} B^T P(t+1)A \end{aligned} \tag{13}$$

の半正定解 $P(t)$ を用いて，

$$G(t) = -\{B^T P(t+1)B + R\}^{-1} B^T P(t+1)A \tag{14}$$

と与えられる．またこのときの $J(t, L)$ の最小値は $x(t)$ の値に関わらず

$$J_{min}(t, L) = x(t)^T P(t) x(t) \tag{15}$$

と与えられる．

（注意） 上に示すように，制御問題のリッカチ差分方程式は終端値 $P(L) = P_0$ が与えられると，逆時間方向（$t = L-1, L-2, L-3, \cdots$）に解 $P(t)$ を求めていくことができる．これに基づいて，最適ゲインが決定できるが，このことは極めて複雑と思われる最適制御が，実はやや信じがたいほどの単純な

線形の状態フィードバックにより与えられることを意味している．また、この結果は最適制御が逆時間方向に定まることを表す．実際,「制御としてどのようなものが良いか」については,「いつ制御を中止するか」が，重要な意味を持つ．とくに，ある瞬間の入力の影響が十分時間の経過した後の出力に現れる場合，制御を中止する時刻の直前に制御入力を与えても無駄であるといえる．なお、以上のゲ C ン $G(t)$ は状態 $x(t)$ と無関係に定まるため，いったんこれらゲインを求めると，システムの初期状態 $x(0)$ にかかわらず，改めて $G(t)$ を求めなくても，$G(t)x(t)$ により最適制御を実行できる．

1.5 例題

(1) 式で与えられるシステムについて，L を有限とする (10) 式を評価関数とする場合の最適制御を求めよう．この場合，$A = B = Q = 1$，$R = 2$ を代入すると，リッカチ差分方程式 (13) 式は

$$p(t) \quad = \quad p(t+1) + 1 - \frac{p(t+1)^2}{p(t+1)+2} \tag{16}$$

である．ここで，終端条件を

$$p(L) = 0 \tag{17}$$

とする．(16) 式の右辺 $p(t+1)$ にそれぞれ求まった値 $(t = L-1,\ L-2,\ \cdots)$ を代入することにより，

$$p(L) = 0,\ p(L-1) = 1,\ p(L-2) = 1.67,\ p(L-3) = 1.91,$$
$$p(L-4) = 1.98,\ p(L-5) = 1.994,\ \cdots$$

となる．したがって，最適制御の状態フィードバックゲインは

$$g(t) = -\frac{p(t+1)}{p(t+1)+2} \tag{18}$$

より，

$$g(L-1) = 0,\ g(L-2) = -0.33,\ g(L-3) = -0.46,$$
$$g(L-4) = -0.488,\ g(L-5) = -0.497,\ \cdots$$

と与えられ，有限期間の制御であるため，ゲインは時変である．

1.6 証明

数学的帰納法を用いて定理 1 - 1 を証明する．いま，ある区間 $[t+1, L]$ において，$J(t+1, L)$ を最小とする最適制御が存在し，(13)，(14) 式と (15) 式

において t を $t+1$ に置き換えた関係が成立すると仮定しよう．すなわち，同区間の $J(t+1,L)$ が正定行列 $P(t+1)$ を用いて $x(t+1)$ の値に関わらず

$$x(t+1)^T P(t+1)x(t+1) = J_{min}(t+1,L) \tag{19}$$

と表されるとしよう．この仮定のもとに，初期時刻が t の場合を考察する．

制御の問題においては，動的計画法と呼ばれる手法が一般に有効である．動的計画法では、「ある時刻 t 以後において最適な制御が実施されている場合，それ以後の任意の時刻（$t+1$，$t+2$，$\cdots$ などの勝手な時刻）を初期時刻と考えた場合にも，さきの最適制御はこの範囲内の最適制御になっている」との性質を利用する．このため、いま、$P(t+1)$ は求まっていると仮定する。

この場合の評価関数 $J(t,L)$ は $[t+1,L]$ においても最適な制御となっているので，$t+1$ 以後において (10) 式と (19) 式の関係を用いると，

$$\begin{aligned} J(t,L) &= x(t)^T Q x(t) + u(t)^T R u(t) \\ &+ x(t+1)^T P(t+1) x(t+1) \end{aligned} \tag{20}$$

と与えられる．この式に (9) 式を代入すると，

$$\begin{aligned} J(t,L) &= x(t)^T Q x(t) + u(t)^T R u(t) \\ &+ \{Ax(t)+Bu(t)\}^T P(t+1)\{Ax(t)+Bu(t)\} \end{aligned} \tag{21}$$

が得られる．この場合に決定すべき制御変数は $u(t)$ だけであり，$P(t+1)$ は固定して考えることができる．このように，ある一つの時刻の制御入力を求める問題は１段決定問題と言われる．この１段決定問題において $J(t,L)$ が $u(t)$ に関する簡単な２次式であることに注意すると，$J(t,L)$ を最小にする $u(t)$ がただ１つ定まることがわかる．ここで，$u(t)$ 微小な値 $\delta u(t)$ だけ変化した場合の $J(t,L)$ の微小変化（全微分）を $\delta J(t,L)$ とおくと，

$$\begin{aligned} \delta J(t,L) &= u(t)^T R \delta u(t) + \delta u(t)^T R u(t) \\ &+ \{Ax(t)+Bu(t)\}^T P(t+1) B \delta u(t) \\ &+ \delta u(t)^T B^T P(t+1)\{Ax(t)+Bu(t)\} \\ &= 2\Big[u(t)^T R \\ &+ \{Ax(t)+Bu(t)\}^T P(t+1) B\Big]\delta u(t) \end{aligned} \tag{22}$$

となる．なお，$\delta J(t,L)$ スカラ値なので，各項を転置してもその値は変わらないこと，および行列の積の転置は個々の行列の転置において積の順序を交換したものになることを用いている．これより，$u(t)$ がこの１段決定問題の最適入力であるための必要条件は $\delta u(t)$ の正負に関わらず $\delta J(t,L)$ が０となることである．これより，最適入力は

$$u(t)^T R + \{x(t)+Bu(t)\}^T P(t+1) B \;=\; 0 \tag{23}$$

を満たすことになる．他にこの条件を満たす $u(t)$ は存在しないから，この場合の $u(t)$ が与えられた $P(t+1)$ のもとに，$J(t,L)$ を最小にする．

この両辺に $\{B^TP(\tau+1)B+R\}^{-1}$ を掛けて整理すると，

$$u(t) = -\{B^TP(t+1)B+R\}^{-1}B^TP(t+1)Ax(t) \tag{24}$$

が $J(t,L)$ 最小にする入力である．ここで，

$$G(t) = -\{B^TP(t+1)B+R\}^{-1}B^TP(t+1)A \tag{25}$$

とおくと，この場合の $u(t)$ は (14) 式のように線形状態フィードバックで与えられることがわかる．すなわち，$[t,L]$ における (14) が成立することがわかる．なお，少なくとも $R>0$ であれば，この場合の逆行列は存在する。

つぎに，(21) 式を整理して

$$\begin{aligned} J(t,L) &= x(t)^TA^TP(t+1)Ax(t) + x(t)^TQx(t) \\ &+ x(t)^TA^TP(t+1)Bu(t) + u(t)^TB^TP(t+1)Ax(t) \\ &+ u(t)^T\{B^TP(t+1)B+R\}u(t) \end{aligned} \tag{26}$$

がえられる．この中の $u(t)$ に (24) 式を代入すると，右辺の第 3 〜 5 項はそれぞれ＋あるいは－の値を持つ同じ項となり，

$$\begin{aligned} J_{min}(t,L) &= x(t)^T\Big[A^TP(t+1)A+Q \\ &- A^TP(t+1)B\{B^TP(t+1)B+R\}^{-1}B^TP(t+1)A\Big]x(t) \end{aligned} \tag{27}$$

が成立する．したがって，すでに与えられた $P(t+1)$ を用いて，新たに $P(t)$ を (13) 式のように与えると，これを上式に代入して，

$$J_{min}(t,L) = x(t)^TP(t)x(t) \tag{28}$$

となる．このとき，任意の $x(t)$ に対して $J_{min}(t,L)\geq 0$ であるから，$P(t)\geq 0$ である．

以上により，ある $t+1$ に関する最適性の仮定から，t についても同じ最適性が成立することが証明できた．したがって，$t=L$ の場合に，この関係が成立すれば，数学的帰納法により，全ての $t=L,L-1,L-2,\cdots,0$ について定理が成立する．$t=L$ の場合，(10) 式と $P(L)=P_0$ より，

$$J_{min}(L,L) = x(L)^TP(L)x(L) \tag{29}$$

が成立するから，ただちにこの仮定が成立することがわかる．

以上により，数学的帰納法による証明が完結した．一般に，離散時間最適制御の問題はある区間にわたる全ての制御入力を求める必要があり，多段決定問題と言われる．この問題は，はるかに簡単な 1 段決定問題を終端時刻か

ら逆時間方向に繰り返し解くことに帰着される．とくにＬＱ問題では，各１段決定問題の評価関数は制御入力（またはフィードバックゲイン）に関する２次式になっており，その微分は線形方程式となる．このため，最適制御は逆行列を用いて簡単に与えられる．

2　無限期間ＬＱ制御

2.1　リッカチ差分方程式の収束

ここでは，終端時刻 L が十分大きく，当分の間は最適制御を実施続ける場合を考えよう．先に述べたように，最適制御は制御を中止する時刻が重要な意味を持つ．L が十分大きい場合，終端時刻 L を ∞ とみなしたほうが簡単で，安定性に結びついた結果が得られる．

(9) 式のシステムについて，無限期間の評価関数

$$J(t,\infty)=\sum_{\tau=t}^{\infty}\{x(\tau)^TQx(\tau)+u(\tau)^TRu(\tau)\} \tag{30}$$

を考える．

［定理２－１］システムの行列が (A,B) が可制御，(C,A) が可観測，$R=D^TD>0$ と仮定する．このとき代数リッカチ方程式

$$\begin{aligned} P &= A^TPA+Q \\ &- A^TPB\{B^TPB+R\}^{-1}B^TPA \end{aligned} \tag{31}$$

は正定行列解 $P>0$ をただ一つ持つ．このとき，(13) 式のリッカチ差分方程式の解 $P(t)$ は $t\to-\infty$ の場合にこの解 P に収束する．このとき，(30) 式を最小とする最適制御は

$$G=-\{B^TPB+R\}^{-1}B^TPA \tag{32}$$

として，時不変状態フィードバック

$$u(t)=Gx(t) \tag{33}$$

で与えられる．このとき，$(A+BG)$ は安定である．

（注意）無限期間の最適制御は有限期間より簡単になり、時不変ゲインを持つことが明らかになった．したがって，いったん一つの G を決定しておくと，あとは $Gx(t)$ を計算することによりいつも最適制御となる．制御前のシステムの安定性にかかわらず，LQ 制御を行ったシステムが常に安定である

ことを定理は保障する．なお，システムが可制御でなければ，不安定なモードを制御できるとは限らず，また，可観測でなければ、フィードバックで不安定なモードを発見できるとは限らない．したがって、可制御，可観測性が無限期間の LQ 最適制御の安定性を保証する．

2.2 例題

(1) 式のシステムに対し，評価関数が (2) 式の場合を考えよう．明らかに (A, B) は可制御，(C, A) は可観測であり，$D > 0$ である．この場合の代数リッカチ方程式は

$$\begin{aligned} p &= p + 1 - \frac{p^2}{p+2} \\ g &= -\frac{p}{p+2} \end{aligned} \tag{34}$$

で与えられる．両辺に $p+2$ を掛けて整理すると，p は

$$p^2 - p - 2 = (p-2)(p+1) = 0$$

の解である．解として $p = 2$ と $p = -1$ が得られるが，正定解（スカラの場合は正数）は $p = 2$ となる．これより $g = -0.5$，$u(t) = -0.5x(t)$ が無限期間の LQ 最適制御であり，1.2 節の例題の結果と一致する．すなわち，さきの例題の解は LQ 最適制御であったといえる．このとき，$(A + Bg) = 0.5$ であり，安定である．なお，1.5 節の例題において，終端時刻を $L \to \infty$ とおくと，$g(t)$ は $0, \cdots, -0.488, -0.497, \cdots$ であり，上の -0.5 に収束する．

2.3 証明

定理 $1-1$ から同 $2-1$ を誘導する．最初に，$P(L) = 0$ の場合の解を考えよう．この場合，$t < L$ において $P(t+1)$ と $P(t)$ とを比較すると，前者の区間 $[t+1, L]$ において後者の区間 $[t, L-1]$ の最適制御（したがって一般に前者の最適ではない）を実施する場合，観測期間が減って終端コストが無くなる分だけ，評価関数は後者より小さい．前者の最適制御はさらにこれ以下であるため，$P(t+1) \leq P(t)$ がえられる．すなわち，t が小さくなるにしたがって，$P(t)$ は単調非減少である．一方，(A, B) が可制御であるとき，任意の初期状態 $x(t)$ に対して，高々 $t+n$ において $x(t+n) = 0$ とするような制御 $u(t),\ u(t+1),\ \cdots,\ u(t+n-1)$ が存在する．このときの (30) 式の評価関数を $J'(t, L)$ と置くと，任意の $L \geq t+n$ において $J'(t, L)$ は有限かつ一定である．$x(t)^T P(t) x(t)$ は最適制御の評価関数を与えるから，その値は $J'(t, L)$ より小さい．すなわち，$P(t)$ は $t \to -\infty$ のとき，単調非減少かつ有界となる．単調非減少かつ有界列は必ず収束するので，$P(t)$ は収束する．この極限を P

と置くと，明らかに $P \geq 0$ である．この極限において，$P(t)$ および $P(t+1)$ ともに P に収束するから，リッカチ差分方程式から (31) 式の代数リッカチ方程式が得られる．

つぎに，$(A+BG)$ が安定であることを示そう．いま，$(A+BG)$ が漸近安定でなく，$|\lambda| \geq 1$ を満たす固有値 λ と

$$(A+BG)x_\lambda = \lambda x_\lambda \tag{35}$$

を満たす固有ベクトル x_λ が存在すると仮定する．(31) 式に (32) 式を代入すると，

$$P = Q + G^T RG + (A+BG)^T P(A+BG) \tag{36}$$

が得られるが，さらに (35) 式を代入すると，

$$(1-|\lambda|^2)\bar{x}_\lambda^T P x_\lambda = \bar{x}_\lambda^T (Q + G^T RG) x_\lambda \tag{37}$$

が得られる．ただし，$\bar{x}_\lambda$ は x_λ の複素共役ベクトルとする．上式左辺は非正であり，右辺は非負であるため, 両辺 0 でなければならない．一方, 右辺が 0 のとき，固有ベクトル x_λ に対して評価関数は 0 となる．このことは (C, A) は可観測であることに反する．これより，$(A+BG)$ は漸近安定となる．

つぎに，(31) 式の代数方程式が 2 つの異なる正定解 P_1 と P_2 を持つと仮定しよう．このとき，

$$\begin{aligned}
&(A+BG_2)^T P_2 (A+BG_2) + G_2^T RG_2 \\
&\quad = (A+BG_1)^T P_2 (A+BG_1) + G_1^T RG_1 \\
&\quad - (G_1-G_2)^T (B^T P_2 B + R)(G_1 - G_2)
\end{aligned} \tag{38}$$

が得られる．実際，$G_2 = -(B^T P_2 B + R)^{-1} B^T P_2 A$ を用いると，上式両辺はともに $A^T P_2 A - A^T P_2 B (B^T P_2 B + R)^{-1} B^T P_2 A$ となる．これを用いると，(36) 式から

$$\begin{aligned}
P_1 - P_2 &= (A+BG_1)^T P_1 (A+BG_1) + G_1^T RG_1 \\
&\quad - (A+BG_2)^T P_2 (A+BG_2) + G_2^T RG_2 \\
&= (A+BG_1)^T (P_1 - P_2)(A+BG_1) \\
&\quad + (G_1-G_2)^T (B^T P_2 B + R)(G_1 - G_2)
\end{aligned} \tag{39}$$

が得られる．したがって，$(A+BG_1)$ は漸近安定であるから，

$$\begin{aligned}
P_1 - P_2 &= \sum_{k=0}^{\infty} (A+BG_1)^{Tk} (G_1-G_2)^T \\
&\quad (B^T P_2 B + R)(G_1 - G_2)(A+BG_1)^k
\end{aligned} \tag{40}$$

であり，$P_1 - P_2 \geq 0$ となる．以上の関係は添え字 1 と 2 を入れ換えてもよいから, 結局 $P_1 = P_2$ となり，正定解は唯一となる．

最後に，終端値を任意の $P(L) \geq 0$ とする場合について，リッカチ差分方程式 (13) 式の解が代数方程式の解 $P > 0$ に収束 することを示そう．$P(L) = 0$ の解を $P(t)^*$ として，一般の解と区別すると，これが P に収束することはすでに示した．そこで，(39) 式と同様にして，

$$\begin{aligned} P(t) - P(t)^* &= \{A + BG(t)\}^T \{P(t) - P(t)^*\} \{A + BG(t)\} \\ &+ \{(G(t)^* - G(t)\}^T \{B^T P(t+1) B + R\} \{G(t)^* - G(t)\} \end{aligned} \quad (41)$$

が得られる．ここで，$P(L)^* \leq P(L)$ が得られる．つぎに，$P(\tau)^* \leq P(\tau)$ $(\tau = L, L-1, \cdots, t+1)$ が成立すると，この式からただちに t においても同じ関係が成立することがわかる．よって，全ての $t \leq L$ についてこの関係が成立する．ここで，$G(t)$ を時不変最適ゲイン G とした場合の評価関数を $J(t, L) = x(t)^T \hat{P}(t) x(t)$ と置くと，$\hat{P}(t)$ は有限期間では最適でないので，$P(t) \leq \hat{P}(t)$ が成立する．一方，定義より，

$$\lim_{t \to -\infty} \hat{P}(t) \quad = \quad \sum_{k=0}^{\infty} (A + BG)^{Tk} (Q + G^T R G)(A + BG)^k \qquad (42)$$

となる．この値はリッカチ代数方程式の解であるから．$P = \lim_{t \to -\infty} P(t)^* \leq \lim_{t \to -\infty} P(t) \leq \lim_{t \to \infty} \hat{P}(t) = P$ が成立するから，$P(t)$ は代数リッカチ方程式の正定解 P に収束する．

2.4 可制御・可観測条件の拡張

時不変最適制御に関する定理２－１の結果は可制御・可観測条件のもとで証明した．この結果を拡張しよう．すなわち，(A, B) が可制御でなくても，不可制御モードが全て安定モード（固有値の絶対値が１以下）であるとき，(A, B) は可安定と呼ばれる．一方，(C, A) が可観測でなくとも，全ての不可観測モードが安定モードであるとき，(C, A) は可検出と呼ばれる．

［定理２－２］ システムの行列が (A, B) が可安定，(C, A) が可検出，$R = D^T D > 0$ と仮定する．このとき代数リッカチ方程式

$$\begin{aligned} P &= A^T P A + Q \\ &- A^T P B \{B^T P B + R\}^{-1} B^T P A \end{aligned} \quad (43)$$

は半正定行列解 $P > 0$ をただ一つ持つ．このとき，リッカチ差分方程式の解 $P(t)$ は $t \to -\infty$ の場合に，上の代数方程式の解に収束する．またこのとき，(30) 式を最小とする最適制御は

$$G = -\{B^T P B + R\}^{-1} B^T P A \qquad (44)$$

として

$$u(t) = Gx(t) \tag{45}$$

と与えられる．このとき，$(A + BG)$ は安定である．

制御の対象となるほとんどの線形システムは可安定かつ可検出である．定理によれば，システムの一部にフィードバック制御が行えない部分が存在しても，もともとその部分が安定であれば，最適制御は安定なシステムを与える．

3　応答の直交化に基づくＬＱ最適制御の学習

3.1　モデル無しＬＱ学習制御開発の経緯

以上の準備の下に、モデル無しＬＱ学習制御について説明する。現代制御理論に基づく制御系設計では，同定等によるモデリング（方程式を求める操作）とモデルに基づくコントローラの設計とは基本的に独立に扱われてきた．したがって、未知システムのＬＱ最適制御を求めるには、前もってシステムの方程式を同定等により求め、これに基づいてＬＱ最適制御を 2.1 節のように理論的な解析により誘導してきた。たとえば、1.1 節の例題においても、$A = B = 1$ が未知であれば、(6) 式あるいはリッカチ方程式を用いた最適制御の誘導は不可能である。これとは別に、モデルの情報が不完全な場合に直接に制御系を導く適応制御理論の研究も盛んであったが [1],[5]、従来の適応制御では LQ 最適レギュレータはほとんど扱われていない。

これに対して，河村らは 1980 年代に同定と制御系設計を一体化したＬＱレギュレータ構成法を提案した [6]，[8]，[9]。その基礎はＬＱ最適制御の応答信号がある直交関係を満たすことを発見したことによる [10]。これより，特性が未知のシステムであっても，その応答信号を直交化することにより，ＬＱレギュレータの逐次的な構成が可能となった。

実システムでは，静止摩擦など、その動作が完全に線形システムで記述される場合は少ない．この場合、システム同定等によって得られるモデル（方程式）は実システムに対する近似モデルとなる．したがって，モデルを基礎とする従来の LQ 制御が必ずしも望ましい性能を持たない場合がある．モデリングは現実の制御系設計にとって，きわめて面倒な問題である．本方式は閉ループ応答を利用するが，独立した同定のプロセスを持たない．基礎となる応答の直交関係はある程度の非線形をもつシステムに対しても最適条件として有効と思われる。この意味で、本方式はモデリングに伴う困難の軽減に有効である可能性を持つ。

以上の基礎的なアルゴリズムは内積の計算期間が長いため，システム雑音等の影響を受けやすいものであった．これに対し，動的計画法に基づく１段最適化の考えを用い、内積の短期間での再帰形式の計算と統計処理を用いたアルゴリズム [11]，[12]，[13] を導出し，この困難を回避した．実際、実倒立振り子の安定化を達成する実用的な手法が示された [14]，[13].

近年，古田 [15],[16]，Hjalmarsson[17], Chan[18], 藤崎，池田 [19]，らも応答信号をもとにＬＱレギュレータを求める手法を与えている。ただし，これらはそれぞれ考え方が基本的に異なり，応答の直交化に基づく本方式とはまったく違ったものといえる [9]．以下の各章では、河村により提案されたモデル無しＬＱ学習制御について解説する。

3.2 応答信号の内積

制御における信号の直交関係については、初期の Narendra の結果などが知られているが [20],[21]、その後この方面での発展は見られなかった。以下では、河村らの誘導した応答信号のみによりＬＱレギュレータを特徴づける直交条件について説明する．なお、ベクトルの内積はよく知られているが、以下の内積は信号同士の間の内積である．

離散時間 $t = 0, 1, 2, \cdots$ で定義された $x(t) \in R^n$ を状態，$u(t) \in R^m$ を入力とする線形時不変システム

$$x(t+1) = Ax(t) + Bu(t) \tag{46}$$

を S とする．このシステムの被制御出力を

$$z(t)^T = (\{Cx(t)\}^T, \{Du(t)\})^T) \tag{47}$$

とする．直達項 $Du(t)$ の存在に注意されたい．

(10) 式で表された評価関数に対応して、応答の内積を定義しよう。有限期間 $[0, L]$ 上の出力 $z(t)$ の全体 $(z(0)^T, z(1)^T, \cdots, z(L)^T)^T$ を $\boldsymbol{z}$ と表そう．異なる２つの出力 $\boldsymbol{z}_1$ と $\boldsymbol{z}_2$ の内積を

$$\begin{aligned} < \boldsymbol{z}_1, \boldsymbol{z}_2 > &= \sum_{t=0}^{L} z_1(t)^T z_2(t) \\ &= \sum_{t=0}^{L-1} \{x_1(t)^T Q x_2(t) + u_1(t)^T R u_2(t)\} \\ &\quad + x_1(L)^T P_0 x_2(L) \end{aligned} \tag{48}$$

と定義する．この内積は $\boldsymbol{z}_1 = \boldsymbol{z}_2$ の場合に LQ 評価関数に一致する．本文では，$\boldsymbol{z}_1 = \boldsymbol{z}_2 = \boldsymbol{z}$ とした場合の $\lim_{L\to\infty} < \boldsymbol{z}, \boldsymbol{z} >$ を最小とする無限期間ＬＱ最適レギュレータの構成を扱う．

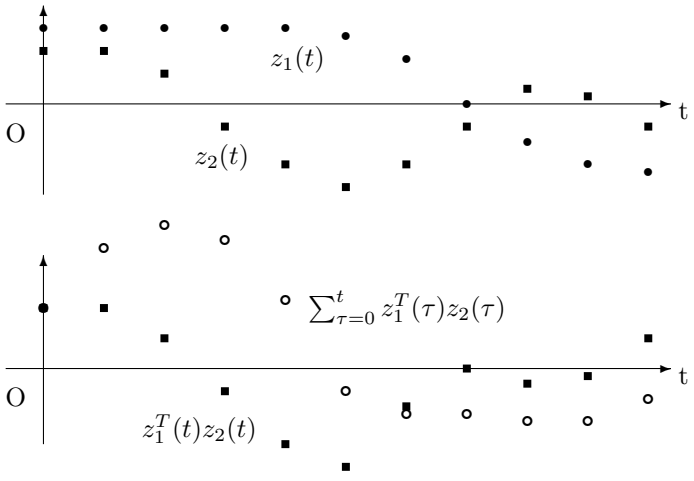


Fig.1 Inner Product of responses

以上に定義した内積の厳密な理解には離散時間ヒルベルト空間の議論を必要とするが [22]，簡単のため，$\boldsymbol{z}$ をスカラ信号として，Fig.1 に説明する．上図に $z_1(\tau)$ と $z_2(\tau)$ の一例を示す。下図の黒色角点はこれらの積 $z_1(t) \times z_2(t)$ を示す。下図の中抜き丸点は初期時刻からその時刻 t までの $z_1(\tau) \times z_2(\tau)$ の総和である。したがって、$L = 10$ では、右端（$t = 10$）の中抜き丸点が内積である。もし、$L \to \infty$（右側極限）において、中抜き丸点が収束するとき、この収束値が $\lim_{L\to\infty} < \boldsymbol{z}_1, \boldsymbol{z}_2 >$ である。

3.3 直交条件

無限期間の最適レギュレータは時不変ゲインによる状態フィードバックにより構成できることがよく知られている．そこで，補助的な追加入力 $v(t)$ を伴ったつぎの時不変状態フィードバックを考える．

$$u(t) = Gx(t) + v(t) \tag{49}$$

この閉ループシステムを改めて $\bar{S}$ と表す．$v(t)$ は $\bar{S}$ の入力となる．

$\bar{S}$ に対し，初期状態を $x^a(0)$，追加入力を $v^a(t) \equiv 0$ とする初期値応答を $\boldsymbol{z}^a$ と表そう．初期時刻の単位インパルスを $\delta(t)$ として，同様に，$x^b(0) = 0$, $v^b(t) \equiv v^b(0)\delta(t)$ とするインパルス応答を と表す．簡単のため，本文では以下の標準条件を仮定する．

条件１ (A, B) は可制御，(C, A) は可観測，$C^T D = 0$，$D^T D > 0$.

これら閉ループシステムの応答を用いて，つぎの直交関係が得られる [8],[10].

［定理３－１］条件１を仮定する．このとき，時不変状態フィードバック $u(t) = Gx(t)$ が無限期間のＬＱ最適制御であるための必要十分条件はすべての $x^a(0) \in R^n$ と $v^b(0) \in R^m$ について

$$\lim_{L\to\infty} < \boldsymbol{z}^a, \boldsymbol{z}^b > \ = 0$$

が成立することである．

上の関係は入力が最適制御である場合に限って，Fig.1 下図の中抜き円点が $t \to \infty$ のときに 0 に収束することを表す．この関係を直感的に示せば、Fig.2 のようになる。いま、初期値応答は O を基点とする矢印で表されると仮定しよう。ただし、その頂点は入力により異なるが、常に直線 BC 上にあるとする。いま、適当な入力を加えた場合の初期値応答を矢印 OA で表し、その評価関数の大きさを矢印の長さ ||OA|| で表すと仮定する。このとき、長さ ||OA|| が最小となる（すなわち評価が最良となる）応答は同図のように、直線 BC と直交する OA^* である。実際、もし最適応答が直線 BC と直交しない場合（たとえば、矢印 OA の場合）、BC と直交する別の直線がより短

い（よい評価を持つ）ことになり、最適性に反する。このとき、BC 上の線分 AA^* はこれらの入力差に対する応答と言える。

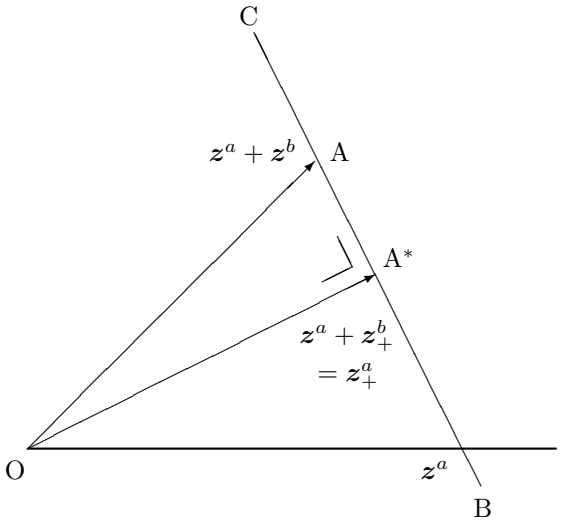


Fig.2 Orthogonality of responses

以上の議論において、初期値応答の入力は必ずしも $u(t) \equiv 0$ である必要はない。すなわち、入力を持つ応答 OA すなわち（$z^a + z^b$）を改めて初期値応答 z_+^a と表そう。この場合の入力にさらに追加入力を加えた場合の応答を $z_+^a + z_+^b$ と表そう。追加入力は z_+^a に摂動を与えていると言える。この追加入力として、インパルスを考えれば十分であることがわかっている。簡単のため、以下では表記 + を省略して表記する。このとき、定理 3-1 の関係を最適応答 z^a と摂動（インパルス応答）z^b との直交性と見ることができる（Fig.2 参照）．これはヒルベルト空間の議論であるが，以上の関係は z^a をより小さくする別の最適な入力の存在を意味する．なお、システムが不安定で応答が発散する場合、もちろんこの直交性は成立しない。

ＬＱ最適制御と最適推定のためのカルマンフィルタには多くの双対関係があることが知られている [7]。上記の直交条件は最適推定の状態推定誤差とイノベーションの直交性と双対関係にある [8]．このように、直交条件ではリッカチ方程式を用いた定式化と異なり、システム方程式を用いずに最適制御を定式化することができ、モデルなしの制御システムの構成に役立つ。

3.4 直交性についての例題

さきの 1.1 節に示した簡単な例題について、応答の内積がどのようになるか見てみよう。ただし、入力を

$$u(t) = gx(t) + v(t) \tag{50}$$

のように与えるとする。この場合、状態方程式は

$$x(t+1) = (1+g)x(t) + v(t) \tag{51}$$

で与えられる。したがって、初期値応答は次式で与えられる。

$$\begin{aligned} x^a(t) &= (1+g)^t x^a(0), \\ u^a(t) &= g(1+g)^t x^a(0) \end{aligned}$$

一方、インパルス応答は次のようになる。

$$\begin{aligned} x^b(0) &= 0, \qquad x^b(t) = (1+g)^{t-1} v^b(0) \cdot (t = 1, 2, 3, ...), \\ u^b(0) &= v^b(0), \qquad u^b(t) = g(1+g)^{t-1} v(0) \cdot (t = 1, 2, 3, ...) \end{aligned}$$

これより、(2) 式の評価関数に対応する応答の内積は応答が収束する場合、すなわち $-2 <$ g < 0 の範囲で、

$$\begin{aligned} \lim_{L\to\infty} \qquad < \boldsymbol{z}^a, \boldsymbol{z}^b > \; &= \Bigg[\sum_{t=1}^{\infty} (1+g)^{2t-1} \\ &\quad + 2\Big\{ g + \sum_{t=1}^{\infty} g^2 (1+g)^{2t-1} \Big\} \Bigg] x^a(0) v^b(0) \\ &= \Bigg[2g + (1+2g^2) \Big\{ \sum_{t=1}^{\infty} (1+g)^{2t-1} \Big\} \Bigg] x^a(0) v^b(0) \end{aligned} \tag{52}$$

である。第 2 右辺の { } で囲まれた部分は無限等比級数である。その初項は $(1+g)$ であり、項比は $(1+g)^2$ である。

一般に、初項が s_0 で項比が γ の無限等比級数の和は $|\gamma| < 1$ のとき、

$$S_\infty = \sum_{t=0}^{\infty} s_0 \gamma^{\,i} = \frac{s_0}{1-\gamma} \tag{53}$$

で与えられる。これを用いて計算すると、

$$\begin{aligned} \lim_{L\to\infty} \qquad < \boldsymbol{z}^a, \boldsymbol{z}^b > \; &= \Big\{ 2g + (1+2g^2) \frac{1+g}{1-(1+g)^2} \Big\} x^a(0) v^b(0) \\ &= [2g\{1-(1+g)^2\} + (1+2g^2)(1+g)] \frac{1}{1-(1+g)^2} x^a(0) v^b(0) \\ &= (-2g^2 + g + 1) \frac{1}{1-(1+g)^2} x^a(0) v^b(0) \\ &= (2g+1)(1-g) \frac{1}{1-(1+g)^2} x^a(0) v^b(0) \end{aligned} \tag{54}$$

これより、右辺が 0 となるのは $g = -1/2$ と 1 である。ただし、$g = 1$ の場合、級数は発散するので、上の関係は成立せず、内積は 0 とならない。内積が 0 となるのは $g = -0.5$ の場合だけである。実際、図 3 に代表的なフィー

ドバックゲイン g に対する $<\boldsymbol{z}, \boldsymbol{z}>$ の値を示す。この図からも初期値応答とインパルス応答の内積が 0 となるのは 2.2 節に述べた最適ゲイン $g = -1/2$ の場合に限ることが確認できる。

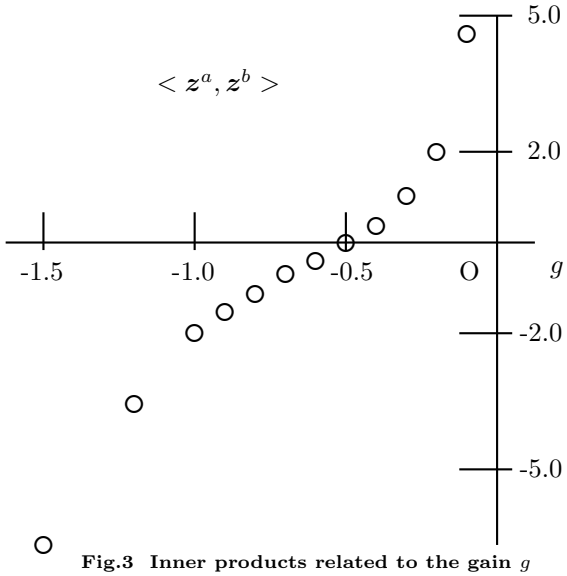


Fig.3 Inner products related to the gain g

3.5 評価関数の感度

さきの定理 3.1 の結果によると，上に述べた応答を直交化すれば，最適レギュレータを構成できるはずである．たとえば、適当な初期状態 $x^a(0)$ に対する初期値応答 $\boldsymbol{z}^a$ と適当なインパルス $v^b(0)\delta(t)$ に対するインパルス応答 $\boldsymbol{z}^b$ をそれぞれ 1 個測定して、その内積 $\lim_{L\to\infty} < \boldsymbol{z}^a, \boldsymbol{z}^b >$ を計算しよう。もし、その値が 0 でなければ、これら応答は直交していない。したがって、少なくとも $x^a(0)$ に対するゲイン G の値は適切でないと想定される。これより、G の値を変更する必要がある。ただし、このままでは G をどのように変更すれば、さきの内積が 0 に近ずくか不明である。また、G を変更すると、いま加えた $x^a(0)$ 以外の状態に対する入力も変化する。すなわち，他の状態に対する直交性を劣化させるかもしれない。

これら疑問に答えるためには、上に述べた内積だけでは不十分であり、多数の応答の観測を用いた勾配法を考える必要がある。以下では、勾配法に基づく直交化を議論する．内積の双線形性より，つぎの関係が成立する．

$$\begin{aligned} < \boldsymbol{z}_1^a + \boldsymbol{z}_1^b, \boldsymbol{z}_2^a + \boldsymbol{z}_2^b > &= < \boldsymbol{z}_1^a, \boldsymbol{z}_2^a > + < \boldsymbol{z}_1^a, \boldsymbol{z}_2^b > + < \boldsymbol{z}_1^b, \boldsymbol{z}_2^a > \\ &\quad + < \boldsymbol{z}_1^b, \boldsymbol{z}_2^b > \\ &= x_1^a(0)^T \Gamma^{aa} x_2^a(0) + x_1^a(0)^T \Gamma^{ab} v_2^b(0) + v_1^b(0)^T \Gamma^{ba} x_2^a(0) \\ &\quad + v_1^b(0)^T \Gamma^{bb} v_2^b(0) \end{aligned} \tag{55}$$

応答 $\boldsymbol{z}_i^b$ を $\boldsymbol{z}_i^a$ に対する摂動と考えると，右辺の第１，第２, ３および第４項はそれぞれ２次評価関数の摂動に関する０次，１次および２次感度を表す.

ＬＱ評価関数の感度は応答信号の内積と直接結びつく．以下では対称半正定行列

$$\Gamma = \begin{pmatrix} \Gamma^{aa} & \Gamma^{ab} \\ \Gamma^{ba} & \Gamma^{bb} \end{pmatrix} \tag{56}$$

を基本感度行列と呼ぶことにする．この行列は (55) 式の応答の内積から初期値項 $x_i^a(0)$ と $v_i^b(0)$ に依存する部分を取り外したものである。

3.6　直交化のためのゲイン修正

以上の議論から，最適制御を得るためには，上に述べた応答信号を直交化すればよいことがわかる．さきの信号の内積を用いると、応答信号のみからこの直交化が実行できる．最も基礎的な逐次直交化アルゴリズムとして，河村らはニュートン法に基づいて

$$\Delta G_k = -(\hat{\Gamma}_k^{bb})^{-1}(\hat{\Gamma}_k^{ab})^T \tag{57}$$

を提案した [6],[8]．ただし，$k = 0, 1, 2, \cdots$ はゲインの修正回数（学習回数），$\Delta G_k = G_{k+1} - G_k$ はゲイン G_k の修正量，$\hat{\Gamma}_k^{ab}$，$\hat{\Gamma}_k^{bb}$ は時不変ゲインと観測長さをそれぞれ G_k，L_k に固定した場合の各感度行列の推定値とする．

4　直接直交化学習制御

4.1　直接直交化アルゴリズム

方程式が未知のシステムに対して、上記の直交化を行うためには，応答信号の内積から感度行列を求めなければならない．すなわち，データの内積（55）式から信号 $x(t)$ や $v(t)$ を取り外す必要がある．これを実行する比較的素朴な計算を以下に示す [6],[8]。

初期状態を $x_{ki}(0)$ （$i = 1, 2, \cdots, N_k$）とするシステムに追加入力 $v_{ki}(t) = v_{ki}(0)\delta(t)$ を加えた場合の応答をそれぞれ $z_{ki}(t)$ として，(55）式より，この場合の基本感度行列を応答信号の内積を用から以下のように容易に決定 (推定) できる．

$$\begin{pmatrix} \hat{\Gamma}_k^{aa} & \hat{\Gamma}_k^{ab} \\ \hat{\Gamma}_k^{ba} & \hat{\Gamma}_k^{bb} \end{pmatrix}$$

$$= \Sigma_k^{-1}\left[\sum_{i=1}^{N_k}\sum_{j=1}^{N_k}\xi_{ki}(0) < \boldsymbol{z}_{ki},\ \boldsymbol{z}_{kj}\ > \xi_{kj}(0)^T\right]\Sigma_k^{-1}, \quad (58)$$

$$\Sigma_k = \sum_{i=1}^{N_k}\xi_{ki}(0)\xi_{ki}(0)^T, \quad (59)$$

$$\xi_{ki}(0) = \begin{pmatrix} x(0)_{ki} \\ v(0)_{ki} \end{pmatrix}$$

ただし，$N_k \geq n + m$ であり，L_k は $k \to \infty$ の場合に十分長い期間とする．

実際、感度行列の定義より、応答の内積と感度行列には、

$$< \boldsymbol{z}_{ki},\ \boldsymbol{z}_{kj}\ >= \xi_{ki}(0)^T\hat{\Gamma}_k\xi_{kj}(0) \quad (60)$$

の関係が成り立つ。多数のデータの内積を用いた (58) 式の右辺にこの関係を代入すると、右辺の [　] の内部は $\xi_{ki}(0)\xi_{ki}(0)^T\hat{\Gamma}_k\xi_{kj}(0)\xi_{kj}(0)^T$ の i と j に関する和となる。このとき、(59）式右辺が $n+m$ 個以上の線形独立な $\xi_{ki}(0)$ を持つと、逆行列 $(\Sigma_k)^{-1}$ が存在する．すなわち、(58) 式の右辺全体は $\hat{\Gamma}_k$ に左から $(\Sigma_k)^{-1}\Sigma_k$ を、右からその転置行列を掛けたものに等しい。逆行列 Σ_k が存在する場合に $(\Sigma_k)^{-1}\Sigma_k$ は単位行列に等しいから、直ちに感度行列と内積の関係である (58) 式が得られる。(58) 式の右辺はモデル（方程式）が未知の場合にも、応答のデータから容易に計算可能である。以下では、数値計算の部分は別として、感度行列を用いて議論を進める。

4.2　アルゴリズムの収束性

いま，$(A + BG_k)$ が安定で，L_k が十分大きいと仮定し，(57), (58) 式をシステム行列を用いて書き直して整理すると，リッカチ代数方程式の数値解法

としてよく知られたニュートン・ラプソンアルゴリズムが得られる [23]．後者では，$(A + BG_0)$ が安定である場合に，G_k はＬＱ最適ゲインに収束する．したがって，前者も上記仮定のもとに同じ収束性を持つ．さらに前者は有限長 L_k の観測期間に対して定義されるため，初期ゲインが安定化ゲインでない場合にも有効であり，シミュレーションでは優れた収束性を示す [6],[8]．

この方式の欠点として、内積の計算期間が長いため、システム雑音 (とくに直流雑音）など実システムの不確実動作の影響を受けやすい。このため、現段階では、実システムに有効な手法となっていない。

5　再帰形式アルゴリズム

5.1　再帰型の直交条件

先に述べた直交化法は応答の直交化の観点から最も基礎的なものといえるが，逆行列 Σ_k が存在するためには、$n+m$ 個以上の線形独立な $\bar{x}_{ki}(0)$ が必要である。すなわち、1回のゲイン修正に少なくとも $n+m$ 個以上の十分の長さの応答観測を必要とする．この手法の精度を上げるためには各観測器間 L_k を十分長くすることが必要であるが，信号に含まれる不要な雑音成分の影響は増大する．一方、無限期間の LQ 最適制御問題は逐次的な1段最適化問題に変換できることはよく知られている [1]〜[4]．上に述べた直交化の欠点の克服を目指す一つの方法として，直交化のための感度行列（内積）の計算において，1段最適化の考えを適用した再帰形式のアルゴリズムが提案された [12],[13]，これを用いて，倒立振り子などの実機テストで実用的な結果を得ている．

先に定義した信号の内積 (48) 式を1段最適化問題に適用する．このとき，$L=1$ とする上記の内積は

$$\begin{aligned} < \boldsymbol{z}_1, \boldsymbol{z}_2 > |_{L=1} \quad &= \quad x_1(0)^T C^T C P x_2(0) + u_1(0)^T D^T D u_2(0) \\ &\quad + x_1(1)^T P x_2(1) \end{aligned} \tag{61}$$

となる．ただし，内積の計算期間は短くなるが，これを用いるめには，ゲインだけでなく，行列 P も未知数となることに注意されたい．この場合の直交条件は以下のように修正される [13]。なお、第1式が P に関する条件である。

［定理５－１］条件1を仮定する．このとき，椙 s 変状態フィードバック $u(t)=Gx(t)$ が無限期間のＬＱ最適制御であるための必要十分条件はある半正定行列 P が存在し，すべての $x^a(0) \in R^n$ と $v^b(0) \in R^m$ について

$$\begin{aligned} x^a(0)^T P x^a(0) \ &= \ < \boldsymbol{z}^a, \boldsymbol{z}^a > |_{L=1} \\ < \boldsymbol{z}^a, \boldsymbol{z}^b > |_{L=1} \ &= 0 \end{aligned}$$

が成立することである．

5.2　評価関数のテーラー展開

よく知られたＬＱ最適制御器（ＬＱレギュレータ）の設計では，前もってシステム方程式を求めておく（具体的には行列 A，B を求めておく）ことが必要である．しかしながら機械系などでは，静止摩擦などの非線形性その他

の影響により，現実のシステムの動作は方程式のそれとの間にかなりの誤差が存在する場合が少なくない．このため，理論上は適切な動作を行うはずの最適制御が必ずしも現実によい制御を与えるとは限らない．以下では，このような難点を改善する方法として筆者らの開発した再帰形式のモデルなしＬＱ最適制御系設計法 [12],[13] を説明する

ここでは無限期間のＬＱ最適制御を考えよう．これを解くための (21) 式の１段最適化問題を感度と勾配法の観点から再考する．最適化の第 k 段階（$k = 0, 1, 2, \cdots$）において，状態 $x(\tau + 1)$ の値に関わらず

$$J(t+1, L) = x(t+1)^T P_k x(t+1) \tag{62}$$

を満たす P_k と状態フィードバックゲイン G_k が仮に与えられているとしよう．この１段最適化問題の最適ゲイン G_{k+1} を $G_k + \Delta G_k$ と置いて，(21) 式に (9) 式と

$$u(t) = G_k x(t) + \Delta G_k x(t) \tag{63}$$

を代入すると，

$$\begin{aligned} J(t, L) &= x(t)^T \Big[\{Q + (G_k + \Delta G_k)^T R (G_k + \Delta G_k) \\ &+ \{A + B(G_k + \Delta G_k)\}^T P_k \{A + B(G_k + \Delta G_k)\} \Big] x(t) \end{aligned} \tag{64}$$

となる．上式は ΔG_k に関する２次式であるから，これを

$$J(t, L) = x(t)^T (\Gamma_k^{aa} + \Gamma_k^{ab} \Delta G_k + (\Delta G_k)^T \Gamma_k^{ba} + (\Delta G_k)^T \Gamma_k^{bb} \Delta G_k) x(t) \tag{65}$$

と書くことができる．また，新たな P_{k+1} を

$$P_{k+1} = \Gamma_k^{aa} + \Gamma_k^{ab} \Delta G_k + (\Delta G_k)^T \Gamma_k^{ba} + (\Delta G_k)^T \Gamma_k^{bb} \Delta G_k \tag{66}$$

と置くと，(62) 式と同様に

$$J(t, L) = x(t)^T P_{k+1} x(t) \tag{67}$$

が成立する．これらはそれぞれ G の変化 ΔG_k に対する $J(t, L)$ あるいは P_{k+1} の０次，１次および２次の変化率（感度）を表す．以下では，(66) 式を各要素の定義とする $(m + n)$ 行 $(m + n)$ 列の行列

$$\Gamma_k = \begin{pmatrix} \Gamma_k^{aa} & \Gamma_k^{ab} \\ \Gamma_k^{ba} & \Gamma_k^{bb} \end{pmatrix} \tag{68}$$

を基本感度行列と呼ぶ．ただし、Γ は対称行列と仮定する．

以上の感度行列をシステム行列 A と B を用いて表すと，

$$\begin{aligned}\Gamma_k^{aa} &= Q + G_k^T R G_k + (A + BG_k)^T P_k (A + BG_k), \\ \Gamma_k^{ab} &= \Gamma_k^{baT} = G_k^T R + (A + BG_k)^T P_k B, \\ \Gamma_k^{bb} &= R + B^T P_k B \end{aligned} \tag{69}$$

の関係が成立する．(68) 式の右辺にこれらを代入すると，

$$\begin{aligned}\Gamma_k &= \begin{pmatrix} Q + G_k^T R G_k & G_k^T R \\ RG_k & R \end{pmatrix} \\ &\quad + \begin{pmatrix} A^T + G_k^T B^T \\ B^T \end{pmatrix} P_k \begin{pmatrix} A + BG_k & B \end{pmatrix} \end{aligned} \tag{70}$$

となる．

5.3 感度に基づく１段最適化問題の解

（65）式〜（67）式を見ると，P_k が与えられている場合，全ての $x(t)$ に対して $J(t, L)$ を最小とする ΔG_k が P_{k+1} を最小化する．このとき，P_{k+1} は ΔG_k に関して２次式であるため，容易に最適解を求めることができる．すなわち，(24）式の $u(\tau)$ を求めたのと同様の議論により，(66) 式は

$$\begin{aligned}P_{k+1} &= \Gamma_k^{aa} - \Gamma_k^{ab} (\Gamma_k^{bb})^{-1} \Gamma_k^{ba} \\ &\quad + \{\Gamma_k^{ab} (\Gamma_k^{bb})^{-1} + \Delta G_k^T\} \Gamma_k^{bb} \{(\Gamma_k^{bb})^{-1} (\Gamma_k^{ab}) + \Delta G_k\} \end{aligned} \tag{71}$$

に変形される．この右辺第 3 項は非負であるから，1 段最適化問題に対して P_{k+1} を最小にする ΔG_k はこの項を 0 にする ΔG_k である．これより，いわゆるニュートン法として

$$\Delta G_k = -(\hat{\Gamma}_k^{bb})^{-1} (\hat{\Gamma}_k^{ab})^T \tag{72}$$

と与えられる．さらに，このとき

$$P_{k+1} = \hat{\Gamma}_k^{aa} - \hat{\Gamma}_k^{ab} (\hat{\Gamma}_k^{bb})^{-1} \hat{\Gamma}_k^{ba} \tag{73}$$

として，$x(t)$ に関わらず

$$J(t, L) = x(t)^T P_{k+1} x(t) \tag{74}$$

が成立する．また新たなゲインは $G_{k+1} = G_k + \Delta G_k$ として

$$G_{k+1} = G_k - (\hat{\Gamma}_k^{bb})^{-1} (\hat{\Gamma}_k^{ab})^T \tag{75}$$

と与えられる [6],[8].

このようにして，感度行列を用いて，１段最適化問題を逐次的に解くことができる．実際，(73) 式と (75) 式に (72) 式と (69) 式を代入すると，

$$\begin{aligned} P_{k+1} &= A^T P_k A + Q - A^T P_k B(R + B^T P_k B)^{-1} B^T P_k A, \\ G_{k+1} &= -(R + B^T P_k B)^{-1} B^T P_k A \end{aligned} \tag{76}$$

となり，リッカチ差分方程式および時変最適ゲインの方程式に一致することがわかる．したがって，厳密に以上のゲイン修正を実行できれば，初期値 $P_0 \geq 0$ と G_0 にかかわらず，G_k は無限期間の LQ 最適ゲインに収束する．

6　再帰形式アルゴリズムのデータ処理

6.1　応答信号に基づく感度行列の算出

以上の議論より，感度行列 Γ_k 求めることにより，ＬＱ問題を解くことができる．以下では，この感度行列を (70) 式ではなく，応答信号の内積から求めよう．最適化の第 k 段階では，区間 $T_k = \{t_k, t_k + 1, t_k + 2, \cdots, t_{k+1} - 1\}$ において一定ゲインのフィードバック制御

$$u(t) = G_k x(t) + v(t) \tag{77}$$

を実行して 楹嚼Ｍ号のデータを得るものとする．ただし，$v(t)$ は適当な信号（例えば白色雑音）とする．これより，

$$\begin{aligned} x(t+1) &= (A + BG_k)x(t) + Bv(t) \\ &= (A + BG_k \quad B)\xi(t) \end{aligned} \tag{78}$$

が得られる．ただし

$$\xi(t) = \begin{pmatrix} x(t) \\ v(t) \end{pmatrix} \tag{79}$$

である．また，適当な行列 C と D を用いて $Q = C^T C$，$R = D^T D$ と与えると，

$$\begin{aligned} z(t) &= \begin{pmatrix} Cx(t) \\ Du(t) \end{pmatrix} \\ &= \begin{pmatrix} C & 0 \\ DG_k & D \end{pmatrix} \xi(t) \end{aligned} \tag{80}$$

として，信号の内積を異なる時刻 $\sigma, \tau \in T_k$ の信号を用いて

$$x(\tau)^T C^T C x(\sigma) + u(\tau)^T D^T D u(\sigma) = z(\tau)^T z(\sigma) \tag{81}$$

と表す．一方，この関係と (70) 式より，

$$\begin{aligned}\xi(\tau)^T\hat{\Gamma}_k\xi(\sigma) &= \xi(\tau)^T\Big\{\begin{pmatrix} C^T & G_k^TD^T \\ 0 & D^T \end{pmatrix}\begin{pmatrix} C & 0 \\ DG_k & D \end{pmatrix} \\ &+ \begin{pmatrix} A^T+G_k^TB^T \\ B^T \end{pmatrix}P_k\,(\,A+BG_k \quad B\,)\Big\}\xi(\sigma) \qquad (82)\end{aligned}$$

が得られる．これに (78) 式と (80) 式を代入すると，

$$\xi(\tau)^T\hat{\Gamma}_k\xi(\sigma) = z(\tau)^Tz(\sigma)+x(\tau+1)^TP_kx(\sigma+1) \qquad (83)$$

となる．すなわち，定理 5-1 を参照すれば，右辺はそれぞれ τ と σ を初期時刻とする応答と応答の内積 $<\boldsymbol{z}^a,\boldsymbol{z}^b>|_{L=1}$ といえる。

このままでは，初期値応答とインパルス応答の内積を信号の大きさに依存せずに表す Γ_k を陽に求めることができないので，以下の処理を行う。両辺に左より $\xi(\tau)$ をかけ，右より $\xi(\sigma)^T$ を掛けた後，T に属する全ての τ と σ についてこれを加算すると，

$$\begin{aligned}&\sum_{\tau\in T_k}\xi(\tau)\xi(\tau)^T\hat{\Gamma}_k\sum_{\sigma\in T_k}\xi(\sigma)\xi(\sigma)^T \\ &\quad=\sum_{\tau\in T_k}\sum_{\sigma\in T_k}\xi(\tau)\{z(\tau)^Tz(\sigma) \\ &\qquad+x(\tau+1)^TP_kx(\sigma+1)\}\xi(\sigma)^T \qquad (84)\end{aligned}$$

となる． T に含まれる $x(\tau)$ が少なくとも状態数＋入力数以上の線形独立成分を持つならば，逆行列

$$\Xi_k=\Big\{\sum_{\tau\in T_k}\xi(\tau)\xi(\tau)^T+\delta I\Big\}^{-1} \qquad (85)$$

が存在する．とくに，δ を十分小さい正数とすると，通常ではこの項は無視可能であるが，$\xi(\tau)$ の個数に関わらず Ξ_k の存在は保証される．(84) 式の左右から Ξ_K を掛けることにより，基本感度行列は次式のように求められる [13].

6.2 基本感度行列の推定値

$$\begin{aligned}\hat{\Gamma}_k &= \Xi_k\Big[\Big\{\sum_{\tau\in T_k}\sum_{\sigma\in T_k}\xi(\tau)<\boldsymbol{z}_\tau,\boldsymbol{z}_\sigma>\xi(\sigma)^T\Big\}+\delta^2\hat{\Gamma}_k^o\Big]\Xi_k \\ &= \Xi_k\Big[\Big\{\sum_{\tau\in T_k}\sum_{\sigma\in T_k}\xi(\tau)\{z(\tau)^Tz(\sigma) \\ &\quad+x(\tau+1)^TP_kx(\sigma+1)\}\xi(\sigma)^T\Big\}+\delta^2\hat{\Gamma}_k^o\Big]\Xi_k \qquad (86)\end{aligned}$$

ただし，$\boldsymbol{z}_\tau$ と $\boldsymbol{z}_\sigma$ はそれぞれ、$t=\tau$ および $t=\sigma$ を初期時刻と見做した応答を表す。Ξ_k は (59) 式の Σ_k^{-1} に相当し，

$$\hat{\Gamma}_k^o \quad = \quad \begin{pmatrix} P_k & 0 \\ 0 & \hat{\Gamma}_{k-1}^{bb} \end{pmatrix}. \tag{87}$$

なお，(86) 式の δ^2 項は $\hat{\Gamma}_k^{bb}$ が必ず正である（0 にならない）ことを保証する．また，$\hat{\Gamma}_k^o$ の初期値 $\hat{\Gamma}_0^o$ において，$\hat{\Gamma}_{-1}^o = D^T D > 0$ とする．

実際、後に示すように応答の内積から感度行列 $\hat{\Gamma}_k$ を算出する部分が本アルゴリズムにおいて、もっとも多量の計算を必要とする。

最初に，適当な初期値 $P_0 \geq 0$ とフィードバックゲイン G_0 を与える．適切な値が不明の場合，例えば $P_0 = 0$，$G_0 = 0$ とできる．T_0 において G_0 を用いた (77) 式の制御を行った場合の応答信号データーを用いて，(86) 式より T_0 の終了時に感度行列 Γ_0 を求めることができる．さらに，(72) 式と (73) 式を用いて，新たな P_1 と G_1 を求める．ゲイン G_k だけでなく，同時に P_k を求める点が直接直交化と異なるが，以上の操作を $k = 0, 1, 2, \cdots$ について繰り返すことにより，時不変ＬＱ最適制御を与える P^* と G^* （リッカチ代数方程式の解）を逐次的に求めることができる．

なお，入力数が２以上の場合，$v(t)$ の各要素は互いに独立な成分を持つ必要がある．たとえば，$v(t)$ として n 個の各成分が独立な白色雑音を用いることによって，この条件は満たされる．

6.3 標準 LQ 学習制御アルゴリズム (不確実データへの対策)

システムの動作が完全に (9) 式の通りだとすると，ニュートン法を用いた (72)，(73) 式のゲイン修正は高速に最適ゲインに収束し，有効である．しかしながら，現実のシステムの動作は多くの場合，完全に (9) 式では表されず，不確実な動作が含まれる．すなわち，各時刻毎の応答信号データはあまり信頼できないのが実状である．このような場合，ニュートン法の利用ではゲインが収束せず発散する危険があり，実際にもあまりよい結果は得られない．この場合，適当な統計処理を含む勾配法が有効である．ここでは，(72)，(73) 式を一般化し，統計処理を導入したアルゴリズムを与える [13]。現在、本研究室で実験等に用いられているアルゴリズムのほとんどはこの形式を土台としている。

$$\begin{aligned} \Delta G_k &= -(\hat{\Gamma}_k^{bb} + H_k)^{-1}(\hat{\Gamma}_k^{ab})^T, & (88) \\ G_{k+1} &= G_k + \Delta G_k \\ P_{k+1} &= \Gamma_k^{aa} + \Gamma_k^{ab}\Delta G_k + (\Delta G_k)^T\Gamma_k^{ba} + (\Delta G_k)^T\Gamma_k^{bb}\Delta G_k & (89) \\ &= \hat{\Gamma}_k^{aa} - \hat{\Gamma}_k^{ab}(\hat{\Gamma}_k^{bb} + H_k)^{-1}(\hat{\Gamma}_k^{bb} + 2H_k)(\hat{\Gamma}_k^{bb} + H_k)^{-1}\hat{\Gamma}_k^{ba}. \end{aligned}$$

ただし，半正定行列 H_k は

$$\begin{aligned} H_{k+1} &= \lambda_{1k}\, H_k + \lambda_{2k}\, \hat{\Gamma}_k^{bb}, \\ 0 &\leq \lambda_{1k} \leq 1, \\ 0 &\leq \lambda_{2k} \leq 1. \end{aligned} \tag{90}$$

により，再帰的に決定される．ただし，初期の H_k は

$$H_0 \geq 0 \tag{91}$$

を満たす半正定対称行列である．

6.4 ゲイン G_k の最適ゲインへの収束性

以上の半正定行列 H_k は大きければ大きいほど，一般にゲイン修正 ΔG_k をニュートン法より小さくし，現実のゲイン変化が統計処理に従う．とくに $\lambda_{1k} \to 1$ の場合，$H_k \to \infty$ であり，システムが安定であるとき，ゲイン修正量は順次減少する．ただし，ΔG_k があまりにも急激に減少すると，G_k が最適値に到達しない可能性がある．H_k に関する条件は G_k が最適値まで到達するための条件である．比較的大きい外乱を含むシステムでは $\lambda_{1k} \to 1$ とすることが望ましい。

類似の半正定行列の導入は Landau の適応アルゴリズム [24] と共通するものである．ただし Lanndau は現実の信号と目的信号との誤差方程式にこのような半正定行列を代入したのに対し，ここでは，リッカチ差分方程式に対してこの正定行列を用いる．なお，(89) 式は (66) 式に (88) を代入して容易に得ることができる．これによって，従来は目的どおりの応答が求められる (モデルマッチングが可能である）適応制御問題にのみ利用できた考えを最良近似解を求める LQ 最適制御問題に拡張した。実際、以上のアルゴリズムが $k \to \infty$ の場合に無限期間の LQ 最適ゲインに収束することは，やや難解ではあるが、理論的に厳密な証明を行っている [13],[26]。

6.5 Practical calculation on the sensitivity matrix

We explain calculation of the sensitivity matrix that leads to practical use of this algorithm[13]. The computer program uses these equations.

Suppose that $m = 1$ and that $t \in T_k$. Define some parameters as

$$\Psi(t) = \sum_{\tau \leq t-1 \cap \tau \in T_k} z(\tau)\xi(\tau)^T, \tag{92}$$

$$\Upsilon(t) = \sum_{\tau \leq t-1 \cap \tau \in T_k} P_k x(\tau+1)\xi(\tau)^T, \tag{93}$$

$$\Xi(t) = \Big\{ \sum_{\tau \le t-1 \cap \tau \in T_k} \xi(\tau)\xi(\tau)^T + \delta I \Big\}^{-1}, \tag{94}$$

$$\begin{aligned}\Theta(t) &= \Big\{ \sum_{\tau \le t-1 \cap \tau \in T_k} \xi(\tau)\{z(\tau)^T z(\sigma) \\ &\quad + x(\tau+1)^T P_k x(\sigma+1)\}\xi(\sigma)^T \Big\} + \delta^2 \Gamma_k^o. \end{aligned} \tag{95}$$

We rewrite (95) as folloes.

$$\begin{aligned}\Theta(t+1) &= \Theta(t) + \xi(t)z(t)^T \Big\{ \sum_{\sigma \le t-1 \cap \sigma \in T_k} z(\sigma)\xi(\sigma)^T \Big\} \\ &\quad + \xi(t)x(t+1)^T P_k \Big\{ \sum_{\sigma \le t-1 \cap \sigma \in T_k} x(\sigma+1)\xi(\sigma)^T \Big\} \\ &\quad + \Big\{ \sum_{\tau \le t-1 \cap \tau \in T_k} \xi(\tau)z(\tau)^T \Big\} P_k z(t)\xi(t)^T \\ &\quad + \Big\{ \sum_{\tau \le t-1 \cap \tau \in T_k} \xi(\tau)x(\tau+1)^T \Big\} P_k x(t+1)\xi(t)^T \\ &\quad + \xi(t)\{z(t)^T z(t) + x(t+1)^T P_k x(t+1)\}\xi(t)^T \\ &= \Theta(t) + \xi(t)z(t)^T \Psi(t) + \xi(t)x(t+1)^T \Upsilon(t) \\ &\quad + \Psi(t)^T z(t)\xi(t)^T + \Upsilon(t)^T P_k x(t+1)\xi(t)^T \\ &\quad + \xi(t)\{z(t)^T z(t) + x(t+1)^T P_k x(t+1)\}\xi(t)^T \end{aligned} \tag{96}$$

Define $\Delta\Theta$ as the remaining right-hand-side except $\Theta(t)$. Note that $\Theta(t) = \Xi(t)^{-1}\Gamma(t)\Xi(t)^{-1}$ and $\Xi(t)^{-1} = \Xi(t+1)^{-1} - \xi(t)\xi(t)^T$. Then we have

$$\begin{aligned}\Theta(t+1) &= \Theta(t) + \Delta\Theta(t) \\ &= \Xi(t+1)^{-1}\hat{\Gamma}(t)\Xi(t+1)^{-1} \\ &\quad - \xi(t)\xi(t)^T\hat{\Gamma}(t)\Xi(t+1)^{-1} - \Xi(t+1)^{-1}\hat{\Gamma}(t)\xi(t)\xi(t)^T \\ &\quad + \xi(t)\xi(t)^T\hat{\Gamma}(t)\xi(t)\xi(t)^T + \Delta\Theta(t) \end{aligned} \tag{97}$$

where the definition of $\Theta(t)$, namely (95) implies

$$\begin{aligned}\Delta\Theta(t) &= +\xi(t)\{z(t)^T\Psi(t) + x(t+1)^T\Upsilon(t)\} \\ &\quad + \{\Psi(t)^T z(t) + \Upsilon(t)^T x(t+1)\}\xi(t)^T. \\ &\quad + \xi(t)\{z(t)^T z(t) + x(t+1)^T P_k x(t+1)\}\xi(t)^T. \end{aligned} \tag{98}$$

From the difinitions, we have the following results.

Calculation of the sensitivity matrix

$$\Psi(t+1) = \Psi(t) + z(t)\xi(t)^T, \tag{99}$$

$$\Upsilon(t+1) = \Upsilon(t) + P_k x(t+1)\xi(t)^T, \tag{100}$$

$$\Xi(t+1) = \Xi(t) - \{\Xi(t)\xi(t)\}\{\xi(t)^T\Xi(t)\xi(t) + 1\}^{-1}\{\xi(t)^T\Xi(t)\} \tag{101}$$

The equation (101) is the matrix inversion lemma[7]. By multiplying (97) with $\Xi(t+1)$ from both sides, we have

$$\begin{aligned}\hat{\Gamma}(t+1) &= \hat{\Gamma}(t) \\ &\quad +\Xi(t+1)\xi(t)\Big[\big\{z(t)^T\Psi(t)+x(t+1)^T\Upsilon(t)\big\}\Xi(t+1) \\ &\quad -\xi(t)^T\hat{\Gamma}(t)\Big]+\Big[\Xi(t+1)\big\{\Psi(t)^Tz(t) \\ &\quad +\Upsilon(t)^Tx(t+1)\big\}-\hat{\Gamma}(t)\xi(t)\Big]\xi(t)^T\Xi(t+1) \\ &\quad +\Xi(t+1)\xi(t)\big\{z(t)^Tz(t)+x(t+1)^TP_kx(t+1) \\ &\quad +\xi(t)^T\hat{\Gamma}(t)\xi(t)\big\}\xi(t)^T\Xi(t+1). \qquad (102)\end{aligned}$$

Equations (99), (100), (101) and (102) compose practical calculation at each time on T_k. The initial values of each block are

$$\Psi(t_k) = 0, \qquad (103)$$

$$\Upsilon(t_k) = 0, \qquad (104)$$

$$\Xi(t_k) = \frac{1}{\delta}I, \qquad (105)$$

$$\hat{\Gamma}(t_k) = \hat{\Gamma}_k^o. \qquad (106)$$

The terminal value of $\hat{\Gamma}(t)$ on T_k gives the fundamental sensitivity matrix as

$$\hat{\Gamma}_k = \hat{\Gamma}(t_{Fk}+1) \qquad (107)$$

where t_{Fk} is the terminal time of T_k. These equations have only products of vectors with other matrices or vectors, and they have no matrix inversion and no products of matrices with other matrices. Therefore the order of calculation is $O((n+1)^2)$. Then the gain update is performed by (88)- (90) at the end of each T_k.

6.6 例題

さきに，2.2 節に示した 1 次システムに関する例題

$$\begin{aligned}\mathrm{x}\ (t+1) &= x(t)+u(t)、 \\ J &= \sum_{t=0}^{\infty}\{x(t)^2+2u(t)^2\}\end{aligned}$$

について，モデルが未知であると仮定して，再帰型 LQ 学習制御を実行した場合のシミュレーション結果を以下に示す。この手法の標準的な学習では、フィードバックゲインを算出した値に変更しながら実行するデータ観測と，

観測データに基づく新たなゲイン算出とが同時に繰り返し行なわれる．各ゲインの修正に 2 サンプルのデータを用いているので，$t = 2k$ $(k = 1, 2, 3, \cdots)$ において順次得られるゲイン G_k を用いて、$t = 2k$ と $2k + 1$ において、入力 $u(t) = G_k x(t) + v(t)$ がシステムに加えられ，観測が継続される．ただし，$v(t)$ はパソコンにより生成される乱数である．

雑音 $v(t)$ とこれにより駆動される状態 $x(t)$ のデータを Fig.4 に示す．ただし，横軸は時刻 t（0 から 30）である。また、順次算出されたゲイン G_k と行列（この場合はスカラ）P_k を Fig. 5 に示す。このグラフの横軸は学習回数 k（0 から 15）である。

学習アルゴリズムに用いたパラメタは以下の通りである。

各ブロック長：2,
追加雑音 $v(t)$：[-1,1] の範囲の一様雑音
初期状態 $x(0)$：=0,
初期ゲイン K_0=0,
P_0=0,
H_0=0,
$\lambda_1(t)$0.9, $\lambda_2(t)$0.2,
δ=0.00001,

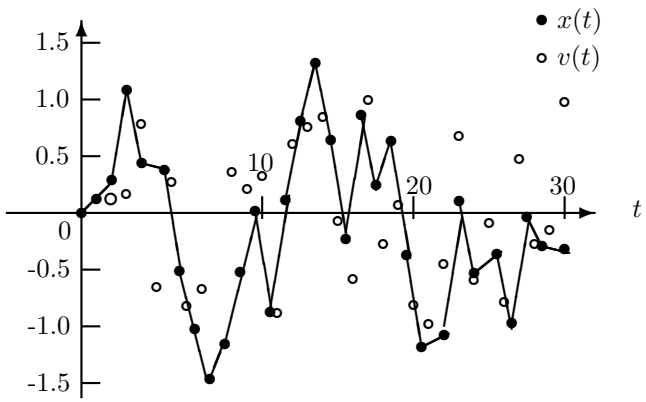


Fig. 4 Change of the added noise $v(t)$ and the state $x(t)$

,,,,

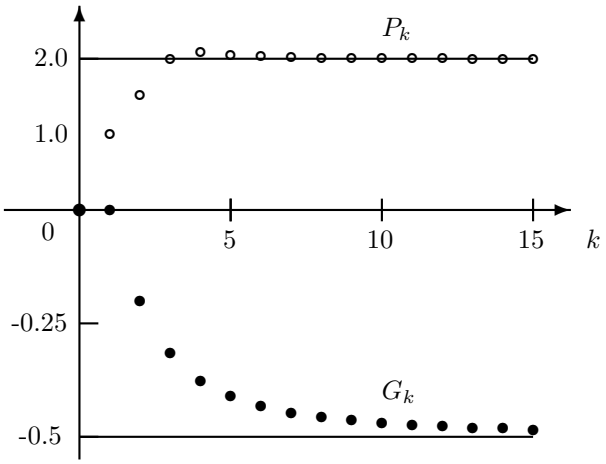


Fig. 5　Change of G_k and P_k in learning

状態 $x(t)$ の変化を見ると、駆動雑音 $v(t)$ に対する応答は有界である。実際、Fig. 5 に示される各時刻 t のゲイン $G(t)$ は $t \geqq 2$ において安定化ゲインである。特に、t が大きい場合のゲイン G_k はほぼ一定であり，得られたゲインはフィードバックシステムを安定化しているといえる．また、ゲインの算出にシステムモデルの A と B を用いていないにも関わらず，G_k と P_k は 2.2 節に示した最適値 $G^* = -0.5$ と P^*=2.0 に収束していることがわかる。ただし、$k > 0$ において $H_k > 0$ であるため、リッカチ差分方程式の収束状況とは似ているが、違いがある。

7　学習パラメタにの選定に関して

7.1　行列 Q と R

行列 Q の要素の一部を大きくする場合、その要素に対応する状態を強力に小さくする最適化がなされる。逆にある要素を小さくすると、その要素に対応する状態をあまり小さくすることができない場合が多い。一方、R を大きくすると、入力重視であるため一般に制御力は小さくなる。逆に大きくしすぎると、制御は強力ではあるが、ぎくしゃくした動作になる場合が多く、特に実機では正常な動作が行えなくなる。

7.2　入力追加雑音の振幅

入力に追加する雑音は小さすぎると実機の入力にほとんど影響が出ず、最適化は行えない。ただし、制御系は常にこの雑音の影響を受け続ける。たと

えば倒立振り子においてこの雑音が大きすぎる場合、たとえ適切なゲインが求められたとしても、振り子が雑音により強制的に倒れるため、最適化が実行できなくなる。

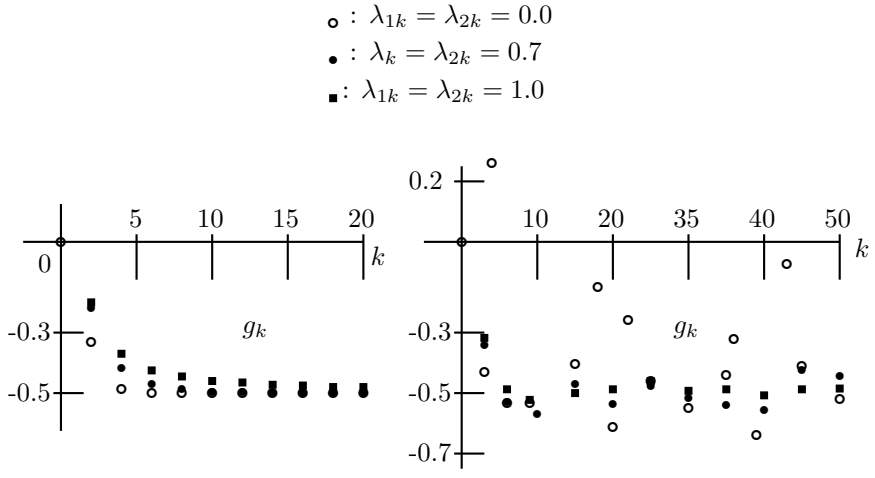
7.3 パラメタ λ_{1k}, λ_{2k} と H_k

実機のデータを用いた場合、必然的に理想的な方程式と多少は異なる不確実なデータが得られる。このようなデータの不確実さがゲイン G_k の推定値に誤差を与える。これを避けるため、アルゴリズムでは H_k が用いられている。正定行列 Γ_k^{bb} に半正定行列 H_k を追加した $(\Gamma_k^{bb}+H_k)$ の逆行列に比例してゲイン修正量 ΔG_k が決まるため、H_k は ΔG_k を小さくする役割がある。

一例として、1.1 節の例題に白色のシステム雑音 $w(t)$ を追加した

$$x(t+1)x(t)+u(t)+w(t) \tag{108}$$

を考えよう。ただし、$w(t)$ は測定できないシステム雑音とする。いま、$w(t)\equiv 0$ として、観測データに誤差を伴わないシミュレーションデータを用いた場合には、図 6 (a) に示すように λ_{1k}、λ_{2k} が小さく、したがって H_k が小さいほどゲイン G_k の収束が速い。ただし、評価関数は 1.1 説と同じであり、最適ゲインは−0.5 である。また、λ 以外の学習パラメタは同一である。なお、グラフのデータは適当に間引いて表現されている。



(a) Gain change without system noise (b) Gain change with system noise

Fig.6 Difference of Gain change related to λ

一方、$w(t)$ を平均値が 0 で分散が 0.1 の白色雑音とした場合のゲイン G_k の変化を同図 (b) に示す（ただし、横軸のスケールが (a) と異なる）。この場合、λ_{1k} と λ_{2k} 大きいほど、G_k は最適値-0.5 に近い値をとる。実際、λ_{1k} と λ_{2k} が小さいと、G_k は雑音の影響を大きく受け、正しい値に収束せずに、不

確定な変動を繰り返す。実機では、λ_{1k} を 0.9 以上にする場会いが多い。また、大きいシステム雑音が存在する場合に G_k を収束させるため、 $\lambda_{1k} \equiv 1$ とする場合も存在する。なお、システム雑音が存在していない場合に、ゲイン G_k の収束が証明されているのは λ_{1k} と λ_{2k} が 0 以上かつ 1 以下の場合である。これらが 1 を超える場合には ΔG_k があまりに小さくなり、G_k が正しい値に至らずに途中で止まってしまうことが起こりえる。状況に応じた適当な数値を用いることが望まれる。

7.4 各ブロックのサンプル数

各ブロックのデータの個数を与えるサンプル数は理論上は（状態数＋入力数）以上でなければ、正しい最適化は行えない。一般的に言えば、各ブロックあたりのサンプル数を多くしたほうが、各ブロックの Δ Gk および P_k の修正量は正確の求められる。このため、k を横軸にとったゲイン修正は良い結果を与える。ただし、ゲイン 1 回の修正に時間がかかるため（k あたりのサンプル数が多いため)、ゲインの収束に多くの時間を必要とする。

7.5 逆行列補題の初期値

逆行列補題では、$(\sum \xi(t)\xi(t)T)$ が正則でなく逆行列を持たない場合に（たとえば、各ブロックでのデータ数が少ない場合)、逆行列が求められない。これを避けるため、δ を正数、I を単位行列として、$(\sum \xi(t)\xi(t)T)$ に δI を追加している。これにより、つねに逆行列 $\Xi(t)$ が常に存在し、 この問題を取り除くことができる。ただし、δI なしでも $\Xi(t)$ が正則な場合に、δI の追加の影響を小さくするため、δ としてかなり小さな数を用いている。ただし δ をあまり小さくしすぎると、計算機が有限桁の処理のためにかえって大きな誤差を生じる。

以上に示すように、漸近的に LQ 最適システムが得られており、十分大きい t に対して、システムの動作は $v(t)$ で駆動した LQ 最適フィードバックシステムをのとほぼ同じであることがわかる。ここでは、$G_0 = 0$、$P_0 = 0$ の場合の結果を示したが、これら初期値が異なる場合、G_k と B_k の値は異なるが、同様に G^* ＝－ 0.5 と P^*=2.0 に収束する。

8 出力フィードバックのモデル無しＬＱ学習制御

8.1 拡大状態モデル

システムは可観測として、状態の一部しか直接に測定できない場合の現代制御理論の標準的な手法はオブザーバを用いて状態の推定を行い、真の状態

ではないが、推定状態のフィードバックを行うものである。ただし、標準的なオブザーバの構成にはシステムの方程式のかなりの知識が必要であり、システムが未知の場合には容易に構成できない。この出力フィードバック問題に対して，筆者らはシステムの構造がほとんど未知な場合に、モデル無し LQ 学習制御を実行する方法を提案した [8]。以下にはこの手法について説明する。

測定可能な状態からなる出力を

$$y(t) = Cx(t) \tag{109}$$

とおく。ただし (C, A) は可観測とする。可観測性に着目すると、現在の出力は過去の出力と入力により、適当な正数を $\bar{n}$ として、

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + ... + a_{\bar{n}} y(t-\bar{n}) \tag{110}$$

$$+b_1 u(t-1) + b_2 u(t-2) + ... + b_{\bar{n}} u(t-\bar{n}) \tag{111}$$

と表すことができる。実際、1 入力 1 出力システムについて、この関係は $\bar{n} = n$ としてよく知られている。ここで、測定可能な信号を用いて、

$$X(t) = (y(t)^T\ y(t-1)^T\ ...\ y(t-\bar{n}+1)^T \tag{112}$$

$$u(t-1)^T\ ...\ u(t-\bar{n}+1)^T)^T \tag{113}$$

とおき、$X((t)$ を新たな状態とみなした拡大状態方程式

$$X(t+1) = \bar{A}X(t) + \bar{B}u(t), \tag{114}$$

$$y(t) = \bar{C}X(t), \tag{115}$$

$$\bar{A} = \begin{pmatrix} a_1\ a_2\ ...\ a_{\bar{n}}\ b_2\ ...\ b_{\bar{n}} \\ \mathbf{I}_{\bar{n}-1}\ 0\ \ \mathbf{0} \\ 0\ \ ...\ \ 0\ 0\ ...\ 0 \\ \mathbf{0}\ \ 0\ \ \mathbf{I}_{\bar{n}-2}\ 0 \end{pmatrix}, \tag{116}$$

$$\bar{B} = (b_1\ \ 0\ \ ...\ \ 0\ \ I\ \ 0\ ...\ 0)^T, \tag{117}$$

$$\bar{C} = (I\ \ \ 0\ \ ...\ \ 0\ \ \ 0\ \ \ ...\ \ \ 0), \tag{118}$$

$$\bar{D} = D \tag{119}$$

を考える。この式の 1 行目が (111) 式の t を $t+1$ に置き換えたものである。ただし、$\mathbf{I}_{\bar{n}-1}$ は $\bar{n}-1$ 次のブロック対角行列である。このシステムを $\bar{S}$ とすと、次の関係が成立する [8]。

［定理８－１］ 条件 1 が成立し、$\bar{n}$ を (111) 式の表現が可能な整数とする。このとき、システム S と $\bar{S}$ は次の関係を満たす。
(1) S と $\bar{S}$ は同一の入出力 (u, y) 関係を持つ。
(2) $t \to \infty$ のとき、$X(t) \to 0$ から $x(t) \to 0$ が得られる。

(3) $(\bar{A}, \bar{B})$ は可制御であり、$(\bar{C}, \bar{A})$ は可検出であり、$\bar{D}^T\bar{D} > 0$ である。

以上の関係から、$\bar{S}$ を拡大した状態方程式とみなし，状態 $x(t)$ を拡大状態 $X(t)$ に置き換えてモデルなし学習制御に使用することができる。この場合、拡大状態フィードバック $u(t) = \bar{G}_{k1}X_1(t) + \bar{G}_{k2}X_2(t) + ... + \bar{G}_{k,2n^{-}1}X_{2\bar{n}-1}(t)$ のゲイン $\bar{G}_k$ が順次最適化される。ただし、$\bar{S}$ の状態数は S の状態数より大きくなる場合が多い。なお上記とは多少異なる拡大方程式の誘導法も存在する [8]。

8.2 例題

簡単なシステムとして，$x_2(t)$ を測定できない

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+1) \end{pmatrix} = \begin{pmatrix} 1.2 & 0.5 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + \begin{pmatrix} 0.2 \\ 0 \end{pmatrix} u(t), \qquad (120)$$
$$y(t) = 1.5x_1(t)$$

を考える (ただし、$D = 1$)．このとき、上式第 2 式の t を $t+1$ に置き換えた式に第 1 式の 1 行目を代入し、さらにその中の $x_1(t)$ と $x(t-1)$ に第 2 式と、t を $t+1$ に置き換えた第 2 式を再度代入すると

$$y(t+1) = 1.2y(t) + 0.5y(t-1) + 0.3u(t) \qquad (121)$$

が得られる．したがって，$\bar{n} = 2$ とした

$$X(t) = \begin{pmatrix} y(t) \\ y(t-1) \\ u(t-1) \end{pmatrix}, \qquad (122)$$
$$\bar{A} = \begin{pmatrix} 1.2 & 0.5 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\bar{B} = (0.3 \quad 0 \quad 1)^T,$$
$$\bar{C} = (\ 1 \quad 0 \quad 0),$$
$$\bar{D} = 1$$

は (120) 式と同じ入出力関係を持つ可安定かつ可検出な拡大モデル $\bar{S}$ となる。

たとえば、$u(t)$ が $t = 0$ の単位パルスである場合の各変数を本来の状態方程式を用いて計算すると、

$$\begin{array}{rcccccc} t = & 0, & 1, & 2, & 3, & 4, & 5, \\ u(t) = & 1, & 0, & 0, & 0, & 0, & 0, \\ x(t) = & \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \begin{pmatrix} 0.2 \\ 0 \end{pmatrix}, & \begin{pmatrix} 0.24 \\ 0.2 \end{pmatrix}, & \begin{pmatrix} 0.388 \\ 0.24 \end{pmatrix}, & \begin{pmatrix} 0.5856 \\ 0.388 \end{pmatrix}, & \begin{pmatrix} 0.89672 \\ 0.5856 \end{pmatrix} \\ y(t) = & 0, & 0.3, & 0.36, & 0.582, & 0.8784, & 1.34508, \end{array}$$

である。一方、拡大モデルを用いて計算すると、同一の入力 $u(t)$ に対して

$$
\begin{array}{rcl}
t & = & 0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \\
X(t) & = & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.3 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.36 \\ 0.3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.582 \\ 0.36 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.8784 \\ 0.582 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.34508 \\ 0.8784 \\ 0 \end{pmatrix}
\end{array}
$$

が得られる。したがって、出力 $y(t) = CX(t) = x_1(t)$ は本来の状態意方程式の場合の出力 $y(t)$ と同一であることが確認できる。

なおこの例題では、拡大モデルの弟 3 状態 $u(t-1)$ は他の状態と出力の算出に役立っていないため、除外しても同様の結果が得られる。

参考文献

[1] 計測自動制御学会: 自動制御ハンドブック，基礎編，オーム (1983).

[2] 木村: デジタル信号処理と制御，昭晃堂 (1982).

[3] 土屋、江上: 新版、現代制御工学，産業図書 (1991).

[4] 藤井: 新世代工学シリーズ，制御理論，オーム (2002).

[5] 市川、金井、鈴木、田村：適応制御，昭晃堂 (1984).

[6] 河村 嘉顯: 入出力データから最適レギュレータを構成する基礎的アルゴリズム, 計測自動制御学会論文集, 24-11,1216/1218(1988).

[7] 片山徹：応用カルマンフィルタ, 朝倉 (1983).

[8] Kawamura Y.: Direct construction of LQ regulator base on orthogonalization of signals: dynamical output feedback. *Systems & Control Letters*,**34**, 1-9(1998).

[9] 河村 嘉顯：入出力データに基づく LQ 最適制御系設計、システム/制御/情報 44 169/176(2000)

[10] 河村 嘉顯: 離散時間最適制御と最適推定について直交条件の双対性，計測自動制御学会論文集,
24-12, 1260/1267(1988).

[11] 坂田厚志，河村嘉顯他：実データを用いた線形フィードバックシステムの最適化，昭和 60 年度電気関係学会関西支部連合大会 G2-7(1985)

[12] 河村 嘉顯: ＬＱ最適レギュレータ問題のための高速学習法の考察", 計測自動制御学会論文集,
29-7, 767/775(1993).

[13] Kawamura Y, Nakano M and Yamamoto H.: Model-free Recursive LQ Controller Design (Learning LQ Control), *International Journal of Adaptive Control and Signal Processing* 18, 551-570(2004)

[14] 中野 将宏 , 河村 嘉顯: 応答信号に基づく倒立振り子の高速学習ＬＱ制御，計測自動制御学会第 26 回制御理論シンポジウム，361/365(1997).

[15] Furuta K and Wongsaisuwan M.: Closed-form solutions to discrete-time LQ optimal control and disturbance attenuation, *Systems & Control Letters* 20,427/437 (1993).

[16] Furuta K, Wongsaisuwan M.: Closed-form solution to discrete-time LQ optimal control and disturbance attenuation. *Syst. Control Lett.* 1993; **20**: 427-437.

[17] Hjalmarsson H, Gunnarsson S, Gevers M.: Data-based tuning of a robust regulator for a flexible transmission system. *European Journal of control* 1995; **1**: 148-156.

[18] Chan J. T.: Data-based synthesis of multivariable LQ regulator. *Automatica* 1996; **32**: 403-407.

[19] Fujisaki Y, Duan Y, Ikeda M, Fukuda M.: A system representation and optimal control in input-output data space. *Trans. of SICE* (Japanese) 1999; **34**: 1845-1853.

[20] Narendra K.S, and McBride L.E.: Multiparameter self-optimization system using correlation techniques, *IEEE Trans. Automat. Control* AC-9, 31/38(1964)

[21] Narendra K.S. and Streeter D.N.: An adaptive proceduru for controlling undefined linear processes, *IEEE Trans. Automat. Control*, AC-9 545/548(1964)

[22] コルモゴロフ，フォーミン：関数解析の基礎，岩波 (1962).

[23] Hewer G.A.: An iterative technique for the computation of the steady state gains for the discrete optimal regulator, *IEEE Trans. Automat. Control*, AC 16 382/384(1971)

[24] Landau I.D. and Lozano R.: Unification od discrete time explicit model reference adaptive control design, *Automatica* 17, 593/611(1981)

[25] Kawamura Y.: Direct synthesis of LQ regulator from inner product of response signals, 11th IFAC Symposium on system identification (SYSID'97) 1717/1222 (1997)

[26] Kawamura Y.: Extension of the Riccati Difference Equation from Robust Gradient Method, 計測自動制御学会論文集, 33-2, 103/108(1997).