AIC の利用による最小 2 乗パラメータ推定器の 時変忘却係数の一選定法

田 淵 敏 明*・古 本 奈奈代*

A Selection Method of Time-Variable Forgetting Factor of Least Squares Parameter Estimator by Using AIC

Toshiaki TABUCHI* and Nanayo FURUMOTO*

A selection method of time-variable forgetting factor is presented for least squares parameter estimator. The method is based on the monitoring of the statistical behavior of the residual. Then the forgetting factor is selected so as to adapt the residual variance calculated from the estimator to the actual one. AIC is used for the monitoring the statistical behavior of the residual.

Bounds on the time evolution of estimation error are discussed. It is also shown that the proposed method gives acceptable performance in the numerical experiments.

Key Words: parameter estimation, least squares, AIC, forgetting factor

1. はじめに

時間とともに緩やかに変動するパラメータをとまなう回帰型の線形システム

$$y(t) = \phi^T(t)\theta(t) + e(t) \quad (1)$$

のパラメータベクトル $\theta(t)$ を推定するための推定器として、忘却係数 $\lambda(t)$ (<1) を用いた逐次型最小 2 乗推定器がよく知られているが、それは次式で与えられる¹⁾。

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\phi(t)\varepsilon(t) \quad (2)$$

$$\varepsilon(t) = y(t) - \phi^T(t)\hat{\theta}(t-1) \quad (3)$$

$$\left. \begin{aligned} P(t) &= P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{r^2(t)} \\ P(t-1) &= \frac{P(t-1)}{\lambda(t)} \\ r^2(t) &= \phi^T(t)P(t-1)\phi(t) + 1 \end{aligned} \right\} \quad (4)$$

ここで、(1)式における $y(t)$ はスカラーの出力、 $\phi^T(t)$ は $1 \times n$ の回帰ベクトル、 $\theta(t)$ は $n \times 1$ の未知パラメータベクトル、 $e(t)$ は

$$e(t) \sim N[0, \sigma^2] \quad (5)$$

の正規性白色雑音である。また、(2)、(3)式において $\hat{\theta}(t)$ は $\theta(t)$ の推定値を、 $\varepsilon(t)$ は $y(t)$ の 1 段予測誤差を表わす。(4)式における $P(t)$ 、 $P(t|t-1)$ はそれぞれ $n \times n$ の行列である。

このとき、 $\lambda(t)$ の選定に際しては、 $\theta(t)$ の変動に対する推定器の機敏性と、ノイズ $e(t)$ に対する感度の、互いに相反する性質をいかに妥協させるかが問題になる。 $\theta(t)$ の変動状況が事前にわからない場合は、 $\lambda(t)$ の選定は非常に難しく、選定についての一般的な方法はないようである²⁾。Fortescue ら³⁾は、 $\lambda(t)$ の選定についての一つの方法を提案しているが、最終的には試行錯誤を必要とする設計パラメータが関係してくる。

本論では、できるだけ試行錯誤を必要としないような $\lambda(t)$ の選定について述べるものである。 $\lambda(t)$ は、(2)、(4)式からもわかるように、推定器ゲイン $P(t)$ と密接に関係し、推定器性能に大きい影響を及ぼすと考えられる。一方、推定器ゲインの中では、カルマンゲインは最小分散推定値を与えるという意味で最適ゲインである。カルマンゲインのもつ一つの性質は、 $\varepsilon(t)$ の実際の分散と、推定器より計算されるその分散 $\sigma^2 r^2(t)$ を等しくさせる、すなわち $E\{\varepsilon^2(t)|y^{t-1}\} = \sigma^2 r^2(t)$ 、($y^{t-1} = \{y(t-1), \dots, y(1)\}$)、を成立させることである⁴⁾。本手法は、推定器ゲインにこの性質をもたせるようにするため $P(t)$ が $t \rightarrow \infty$ で 0 になるのを防ぐだけでなく、 $E\{\varepsilon^2(t)|y^{t-1}\} = \sigma^2 r^2(t)$ が成立することを旨として $\lambda(t)$ を選定するものである。したがって、 $\varepsilon(t)$ の実際の分散の推定が必要になるが、これを AIC を用いて行う。AIC のパラメータ推定問題への利用は、対象としているシステムの次数決定がほとんどで

* 徳島文理大学工学部 香川県大川郡志度町堂林 1314-1

* Faculty of Engineering, Tokushima Bunri University, Ohkawa-gun, Kagawa
(Received August 20, 1990)
(Revised January 29, 1991)

あるが、本論文では、 $\varepsilon(t)$ の実際の分散推定に、したがって、推定器ゲインの調整に用いられる。

逐次型推定器の最大の利点は、 $\theta(t)$ が変動する場合にも、ある程度までそれを追跡することが可能になるというところにあるが、数値実験例では、本手法が良好な結果を与えることが示される。

2. $\lambda(t)$ の選定指針

$\theta(t)$ は正規性の確率ベクトルとし、 $\phi(t)$ については、 $t \geq L$ に対して

$$\left. \begin{aligned} 0 < \alpha_1 \mathbf{I} &\leq \sum_{i=t-L+1}^t \phi(i) \phi^T(i) \quad (\text{a.s.}) \\ \mathbf{E} \left[\sum_{i=t-L+1}^t \phi(i) \phi^T(i) \right] &\leq \alpha_2 \mathbf{I} \end{aligned} \right\} \quad (6)$$

を仮定する。 $\lambda(t)$ については

$$0 < \lambda_{\min} \leq \lambda(t) \leq \lambda_{\max} < 1 \quad (7)$$

とする。ここで、 $\mathbf{I}$ は単位行列であり、 $\mathbf{E}$ は期待値である。

さて、 $\lambda(t)$ の選定指針を得るために、 $\theta(t)$ の変動状況が random walk

$$\theta(t) = \theta(t-1) + \mathbf{w}(t-1) \quad (8)$$

で表わされるとする。ここで、 $\mathbf{w}(t-1)$ は

$$\mathbf{w}(t-1) \sim N[0, \mathbf{Q}] \quad (9)$$

の白色雑音とし、 $\theta(0)$ 、 $e(t)$ とは統計的に独立とする。このとき、

$$\tilde{\theta}(t) := \theta(t) - \hat{\theta}(t) \quad (10)$$

$$\tilde{\theta}(t|t-1) := \theta(t) - \hat{\theta}(t-1) \quad (11)$$

$$= \tilde{\theta}(t-1) + \mathbf{w}(t-1) \quad (12)$$

として、関数

$$V(t) := \tilde{\theta}^T(t) \mathbf{P}^{-1}(t) \tilde{\theta}(t) \quad (13)$$

の時間進化を考える。ここで、 $:=$ は定義式である。

(2)～(4)式を用いるとき、 $V(t)$ はつぎのように書ける。

$$\begin{aligned} V(t) &= [\tilde{\theta}(t|t-1) - \mathbf{P}(t) \phi(t) \varepsilon(t)]^T \mathbf{P}^{-1}(t) \\ &\quad \times [\tilde{\theta}(t|t-1) - \mathbf{P}(t) \phi(t) \varepsilon(t)] \\ &= \lambda(t) V(t|t-1) + e^2(t) \\ &\quad - [1 - \phi^T(t) \mathbf{P}(t) \phi(t)] \varepsilon^2(t) \end{aligned} \quad (14)$$

ここで、

$$V(t|t-1) := \tilde{\theta}^T(t|t-1) \mathbf{P}^{-1}(t-1) \tilde{\theta}(t|t-1) \quad (15)$$

である。また、(14)式の右辺第3項の部分は、(4)式を用いるとき、つぎのように書ける。

$$1 - \phi^T(t) \mathbf{P}(t) \phi(t) = \frac{1}{r^2(t)} \quad (16)$$

$$r^2(t) = \phi^T(t) \frac{\mathbf{P}(t-1)}{\lambda(t)} \phi(t) + 1 \quad (17)$$

今、(16)、(7)式を考慮して、(14)式の両辺の期待値

をとるとき、次式を得る。

$$\mathbf{E}\{V(t)\} \leq \lambda_{\max} \mathbf{E}\{V(t|t-1)\} + \sigma^2 \left[1 - \mathbf{E} \left\{ \frac{\varepsilon^2(t)}{\sigma^2 r^2(t)} \right\} \right] \quad (18)$$

このとき、右辺第2項の $\mathbf{E}\{\varepsilon^2(t)/\sigma^2 r^2(t)\}$ に注目する。(4)式の代わりに、カルマンフィルタが用いられるならば、 $\sigma^2 r^2(t)$ は $\varepsilon(t)$ の分散に一致し⁴⁾、この項は1になる。その結果、(18)式の右辺第2項は0になる。本論ではこのことを参考に、 $\sigma^2 r^2(t)$ を推定器より計算される $\varepsilon(t)$ の分散と考え、(18)式の右辺第2項が0になるように、すなわち

$$\mathbf{E}\{\nu^2(t)\} = 1 \quad (19)$$

となることを目標に $\lambda(t)$ の選定を行う。このことは、推定器より計算される $\varepsilon(t)$ の分散 $\sigma^2 r^2(t)$ を、 $\varepsilon(t)$ の実際の分散に適合させようとするものである。ここで、

$$\nu(t) := \frac{\varepsilon(t)}{\sigma r(t)} \quad (20)$$

である。(19)式は $\mathbf{E}\{\nu^2(t)|y^{t-1}\} = 1$ と書けるので、(19)式が成立するには

$$\mathbf{E}\{\nu^2(t)|y^{t-1}\} = 1 \quad (21)$$

が成立すればよい。したがって、結局(21)式が満足されることを指針にして $\lambda(t)$ を選定することになる。

3. $\lambda(t)$ の選定

(21)式が満足されるように $\lambda(t)$ を選定することを考えると、まず $\nu(t)$ の実現値をもとに、何らかの方法で $\mathbf{E}\{\nu^2(t)|y^{t-1}\}$ を評価し、それが1に等しいかどうかを調べ、もし1に等しくなければ、1に等しくなるように $\lambda(t)$ の値を設定しなければならない。 $\nu(t)$ は(20)式で定義されるが、 $\nu(t)$ の実現値の計算ということからは、 σ はその推定値で代用するとしても、 $r^2(t)$ のほうに選定しようとしている $\lambda(t)$ 自身に関係しているので、 $r(t)$ の計算が行えず、結局(20)式の形では $\nu(t)$ の実現値が計算できない。そこで、(20)式に対応して

$$\bar{\nu}(t) := \frac{\varepsilon(t)}{\sigma \bar{r}(t)} \quad (22)$$

として、(21)式の代わりに $\bar{\nu}(t)$ の分散を評価することを考える。ここで

$$\bar{r}^2(t) := \phi^T(t) \mathbf{P}(t-1) \phi(t) + 1 \quad (23)$$

であり、 $\bar{r}^2(t)$ は計算可能である。したがって、 $\bar{\nu}(t)$ のほうは σ の部分を推定値で代用するとき、計算可能になる。今、(21)式に対応して

$$\mathbf{E}\{\bar{\nu}^2(t)|y^{t-1}\} = \bar{r}^2(t) \quad (24)$$

とすると、(22)式の関係より

$$\mathbf{E}\{\varepsilon^2(t)|y^{t-1}\} = \sigma^2 \bar{r}^2(t) \bar{r}^2(t) \quad (25)$$

となる。一方、(18)式の右辺第 2 項を 0 にするためには

$$E\{\varepsilon^2(t)|y^{t-1}\} = \sigma^2 r^2(t) \quad (26)$$

となることが必要である。(25), (26)式より

$$r^2(t) = \tau^2(t) \bar{r}^2(t) \quad (27)$$

の関係が得られる。左辺の $r^2(t)$ は、(17)式で与えられるので、これを上式左辺に用いて $\lambda(t)$ について解くとき、次式を得る。

$$\lambda(t) = \frac{\bar{r}^2(t) - 1}{\tau^2(t) \bar{r}^2(t) - 1} \quad (28)$$

これが本論で選定される忘却係数である。ここで、 $\bar{r}^2(t)$ は計算可能であるので、 $\tau^2(t)$ がわかれば $\lambda(t)$ が求まる。すなわち、(18)式の右辺第 2 項が 0 になるように $\lambda(t)$ を選定することは、(24)式で与えられる $\tau^2(t)$ を評価してやればよいことになる。

ところで、(27)式より $\tau^2(t) = r^2(t)/\bar{r}^2(t)$ となるが、これは(17), (23)および(7)式より、1 より大である。すなわち、

$$\tau^2(t) > 1 \quad (29)$$

である。一方、 $\lambda(t)$ のほうは(7)式で表わされるような上下限が設定されているので、(28)式の関係より、 $\tau^2(t)$ のほうも

$$\tau_{\min}^2 \leq \tau^2(t) \leq \tau_{\max}^2 \quad (30)$$

とならなければならない。そして、 $\tau^2(t) = \tau_{\max}^2$ のときに $\lambda_{\min}$ が、 $\tau^2(t) = \tau_{\min}^2$ のときに $\lambda_{\max}$ が得られる。そこで、まず下限 $\tau_{\min}^2$ のほうは(29)式を考慮して

$$\tau_{\min}^2 = 1 + \delta \quad (0 < \delta \ll 1) \quad (31)$$

とする。 δ はたとえば $10^{-3} \sim 10^{-4}$ 程度の小さい値とする。つぎに、上限 $\tau_{\max}^2$ のほうは、 $\lambda_{\min}$ を与える $\tau^2(t)$ をもとにつぎのように評価される。

$$\begin{aligned} \tau^2(t) &= \frac{1}{\bar{r}^2(t)} \left[1 + \frac{\bar{r}^2(t) - 1}{\lambda_{\min}} \right] \\ &\leq 1 + \frac{1}{\lambda_{\min}} \end{aligned} \quad (32)$$

これより、

$$\tau_{\max}^2 = 1 + \frac{1}{\lambda_{\min}} \quad (33)$$

とする。

さて、(28)式より $\lambda(t)$ を計算するためには、 $\bar{v}(t)$ の実現値をもとに $\tau^2(t)$ を推定しなければならない。(22)式で与えられる $\bar{v}(t)$ は、前述のように σ が関係しているため、この部分を推定値 $\hat{\sigma}^2(t-1)$ で代用した

$$\hat{v}(t) := \frac{\varepsilon(t)}{\hat{\sigma}(t-1) \bar{r}(t)} \quad (34)$$

を考え、この分散を $\tau^2(t)$ の推定値とする。したがって、 $\hat{\sigma}^2(t)$ を求めておく必要がある。そこで、(19)式

に注目する。この式より

$$\sigma^2 = E \left\{ \frac{\varepsilon^2(t)}{\tau^2(t) \bar{r}^2(t)} \right\} \quad (35)$$

を得る。ここで、右辺を標本平均で置き換えて、 $\tau^2(t)$ は $\hat{\tau}^2(t)$ で代用するとき、 $\hat{\sigma}^2(t)$ がつぎのように求まる。

$$\begin{aligned} \hat{\sigma}^2(t) &= \frac{1}{t} \sum_{i=1}^t \frac{\varepsilon^2(i)}{\hat{\tau}^2(i) \bar{r}^2(i)} \\ &= \left(1 - \frac{1}{t} \right) \hat{\sigma}^2(t-1) + \frac{1}{t} \frac{\varepsilon^2(t)}{\hat{\tau}^2(t) \bar{r}^2(t)} \end{aligned} \quad (36)$$

次章では、(34)式を用いて $\tau^2(t)$ を推定する。

4. AIC の利用による $\tau^2(t)$ の推定

$\hat{v}(t)$ の分散を $\tau^2(t)$ の推定値とするのであるが、これは $\hat{v}(t)$ の確率分布を推定し、その分散を $\tau^2(t)$ の推定値とすることと同じである。 $\hat{v}(t)$ の確率分布を推定する方法として、いくつかの確率分布のモデルを用意し、 $\hat{v}$ の M 個の実現値 $\hat{v}_{i-M+1} := \{\hat{v}(t), \dots, \hat{v}(t-M+1)\}$ がどのようなモデルに最もよく適合するかを評価することが考えられる。もし、 $\theta(t)$ が一定であることが事前にわかっているならば、 $\lambda(t) \equiv 1$ とすればよく、このときの $\bar{v}(t)$ は y^{t-1} が与えられたとき、 $\bar{v}(t) \sim N[0, 1]$ である。また、 $\theta(t)$ が変動する場合は、 $\bar{v}(t)$ は $N[0, 1]$ に従うとは限らない。このことを参考に、 $\hat{v}(t)$ の確率分布モデルとして、つぎの二つのモデルを用意する。

$$\left. \begin{aligned} &N[0, 1] \\ &N[0, \sigma_v^2(t)] \end{aligned} \right\} \quad (37)$$

そして、 $\hat{v}_{i-M+1}$ がよりよく適合するほうのモデルを $\hat{v}(t)$ の確率分布であると推定し、その分散を $\tau^2(t)$ の推定値とする。

ところで、データのモデルへの適合のよさの判断基準として、AIC を用いることが考えられる。そこで、 $\hat{v}_{i-M+1}$ を(37)式における $N[0, 1]$ のほうのモデルをあてはめたときの AIC を $AIC(0, 1)$ と書き、 $N[0, \sigma_v^2(t)]$ のほうのモデルをあてはめたときの AIC を $AIC(0, \sigma_v^2(t))$ と書くとき、これらはそれぞれつぎのようになる(導出については、付録 1 参照)。

$$\left. \begin{aligned} AIC(0, 1) &= \text{const} + M[\hat{\mu}^2(t) + \hat{\sigma}_v^2(t)] \\ AIC(0, \sigma_v^2(t)) &= \text{const} + M[\log\{\hat{\mu}^2(t) \\ &\quad + \hat{\sigma}_v^2(t)\} + 1] + 2 \end{aligned} \right\} \quad (38)$$

ここで

$$\left. \begin{aligned} \text{const} &:= M \log 2\pi \\ \hat{\mu}(t) &:= \frac{1}{M} \sum_{i=t-M+1}^t \hat{v}(i) \\ \hat{\sigma}_v^2(t) &:= \frac{1}{M} \sum_{i=t-M+1}^t [\hat{v}(i) - \hat{\mu}(t)]^2 \end{aligned} \right\} \quad (39)$$

である。(38)式の値の小さいほうを MAICE (Mini-

mum AIC Estimate) と記すとき, (37) 式のモデルの内 で $\hat{\mathbf{d}}_{t-M+1}$ がよりよく適合するモデルは, MAICE を与えるほうのモデルである. すなわち,

$$\hat{\mathbf{d}}(t) \sim \begin{cases} N[0, 1] & \text{if MAICE} = \text{AIC}(0, 1) \\ N[0, \sigma_p^2(t)] & \text{if MAICE} = \text{AIC}(0, \sigma_p^2(t)) \end{cases} \quad (40)$$

である. このとき, (30) 式を考慮して, $\hat{v}^2(t)$ の推定値がつぎのように得られる.

$$\hat{v}^2(t) = \begin{cases} \tau_{\min}^2 & \text{if MAICE} = \text{AIC}(0, 1) \\ \hat{\sigma}_p^2(t) & \text{if MAICE} = \text{AIC}(0, \sigma_p^2(t)) \\ & \text{and } \tau_{\min}^2 \leq \hat{\sigma}_p^2(t) \leq \tau_{\max}^2 \\ \tau_{\min}^2 & \text{if MAICE} = \text{AIC}(0, \sigma_p^2(t)) \\ & \text{and } \hat{\sigma}_p^2(t) < \tau_{\min}^2 \\ \tau_{\max}^2 & \text{if MAICE} = \text{AIC}(0, \sigma_p^2(t)) \\ & \text{and } \hat{\sigma}_p^2(t) > \tau_{\max}^2 \end{cases} \quad (41)$$

その結果, 得られる $\lambda(t)$ は

$$\lambda(t) = \frac{\hat{r}^2(t) - 1}{\hat{v}^2(t) \hat{r}^2(t) - 1} \quad (42)$$

となる.

本結果をまとめると, アルゴリズムの手順はつぎのようになる.

- 1) 初期値として, $\hat{\theta}(0)$, $\mathbf{P}(0)$, $\hat{\sigma}^2(0)$, M , $\lambda_{\min}$, δ の値を設定する.
 - 2) (3), (23), (34) 式より, $\varepsilon(t)$, $\hat{r}^2(t)$, $\hat{\mathbf{d}}(t)$ を計算するとともに, $\hat{\mathbf{d}}(t)$ を記憶する.
 - 3) もし, $t \geq M$ ならば, (38), (39) 式より MAICE を評価し, (41), (42) 式より $\hat{v}^2(t)$, $\lambda(t)$ を計算する. もし, $t < M$ ならば, $\hat{v}^2(t) \equiv 1$, $\lambda(t) \equiv 1$ とする.
 - 4) (4) 式より $\mathbf{P}(t)$ を計算し, (2) 式より $\hat{\theta}(t)$ を計算する. また, (36) 式より $\hat{\sigma}^2(t)$ を計算する.
 - 5) $t = t + 1$ として, 2) にもどる.
- 次章では, この収束性について考察する.

5. 収束性

$\lambda(t)$ として (42) 式を用いるとき, (18) 式に対応する関係は, 次式で与えられる.

$$\begin{aligned} E\{V(t)\} &\leq \lambda_{\max} E\{V(t|t-1)\} + \sigma^2 \left[1 - E\left\{ \frac{\varepsilon^2(t)}{\sigma^2 \hat{r}^2(t)} \right\} \right] \\ &= \lambda_{\max} E\{V(t|t-1)\} \\ &\quad + \sigma^2 \left[1 - E\left\{ \frac{\varepsilon^2(t)}{\sigma^2 \hat{v}^2(t) \hat{r}^2(t)} \right\} \right] \end{aligned} \quad (43)$$

ここでは, (43) 式の上限を評価することにより, 推定器の収束性を考える. そのために, まず (6) 式のもとで成立するつぎの不等式を用いる (導出については, 付録 2 参照).

$$\begin{aligned} \alpha_3 \mathbf{I} &\leq \mathbf{P}^{-1}(t) \quad (\text{a.s.}) \quad (t \geq L) \\ E\{\mathbf{P}^{-1}(t)\} &\leq \alpha_4 \mathbf{I} \end{aligned} \quad (44)$$

ここで,

$$\left. \begin{aligned} \alpha_3 &:= \alpha_1 \lambda_{\min}^L \frac{1 - \lambda_{\min}^L}{1 - \lambda_{\min}^L} \\ \alpha_4 &:= \lambda_{\max}^L \left(\frac{1}{\beta} + \frac{\alpha_2}{1 - \lambda_{\max}^L} \right) \end{aligned} \right\} \quad (45)$$

であり, s は t/L の整数部である. さて, (12) および (44) 式の第 2 式を用いるとき, (43) 式の右辺第 1 項は, つぎのように評価される.

$$\begin{aligned} E\{V(t|t-1)\} &= E\{\hat{\theta}^T(t|t-1) \mathbf{P}^{-1}(t-1) \hat{\theta}(t|t-1)\} \\ &= E\{V(t-1)\} + \text{Tr}[E\{\mathbf{P}^{-1}(t-1)\} \mathbf{Q}] \\ &\leq E\{V(t-1)\} + \alpha_4 \text{Tr}(\mathbf{Q}) \end{aligned} \quad (46)$$

ここで, $\text{Tr}(\mathbf{Q})$ は, 正方行列 $\mathbf{Q}$ の跡和である. つぎに, (43) 式の右辺第 2 項は, (30), (19) 式を用いてつぎのように評価される.

$$\begin{aligned} E\left\{ \frac{\varepsilon^2(t)}{\sigma^2 \hat{v}^2(t) \hat{r}^2(t)} \right\} &\geq E\left\{ \frac{\varepsilon^2(t)}{\sigma^2 \tau_{\max}^2 \hat{r}^2(t)} \right\} \\ &= E\left\{ \frac{\tau^2(t)}{\tau_{\max}^2} \nu^2(t) \right\} \\ &\geq \frac{\tau_{\min}^2}{\tau_{\max}^2} \\ &= \frac{\lambda_{\min}(1 + \delta)}{1 + \lambda_{\min}} \end{aligned} \quad (47)$$

(47) 式は, (31), (33) 式を用いることにより得られる. (46), (47) 式を (43) 式に用いるとき, $E\{V(t)\}$ はつぎのようになる.

$$\begin{aligned} E\{V(t)\} &\leq \lambda_{\max}^L E\{V(0)\} \\ &\quad + \frac{1}{1 - \lambda_{\max}^L} \left[\alpha_4 \text{Tr}(\mathbf{Q}) + \frac{\sigma^2}{1 + \lambda_{\min}} \right] \end{aligned} \quad (48)$$

一方, 上式左辺は, (44) 式の第 1 式を考慮して

$$\begin{aligned} E\{V(t)\} &= E\{\hat{\theta}^T(t) \mathbf{P}^{-1}(t) \hat{\theta}(t)\} \\ &\geq \alpha_3 E\{\|\hat{\theta}(t)\|^2\} \end{aligned} \quad (49)$$

である. (49) 式を, (48) 式の左辺に用いて, また $V(0) = \hat{\theta}^T(0) \mathbf{P}^{-1}(0) \hat{\theta}(0) = \|\hat{\theta}(0)\|^2 / \beta$ を考慮して

$$\begin{aligned} E\{\|\hat{\theta}(t)\|^2\} &\leq \frac{1}{\alpha_3} \left[\frac{\lambda_{\max}^L}{\beta} E\{\|\hat{\theta}(0)\|^2\} \right] \\ &\quad + \frac{1}{\alpha_3(1 - \lambda_{\max}^L)} \left\{ \alpha_4 \text{Tr}(\mathbf{Q}) + \frac{\sigma^2}{1 + \lambda_{\min}} \right\} \end{aligned} \quad (50)$$

を得る. ここで, α_3 , α_4 は (45) 式で与えられる. これより, $\hat{\theta}(t)$ の収束性については, $\hat{\theta}(t)$ の 2 乗平均が, (50) 式を満足する領域に収束する. (45) 式の α_3 を (50) 式に代入するとき, 漸近的には

$$E\{\|\hat{\theta}(t)\|^2\} \leq \frac{1 - \lambda_{\min}^L}{\alpha_1 \lambda_{\min}^L (1 - \lambda_{\max}^L)} \left\{ \alpha_4 \text{Tr}(\mathbf{Q}) + \frac{\sigma^2}{1 + \lambda_{\min}} \right\} \quad (51)$$

となる.

6. 数値実験例

$\theta(t)$ が変動しているかどうかは、一般には事前にわからないということを想定して、実験は $\theta(t)$ が変動する場合と、しない場合の二つについて同一のアルゴリズムを用いて行った。

【 $\theta(t)$ が変動する場合⁵⁾】

この場合は、

$$\left. \begin{aligned} y(t) &= \phi^T(t) \theta(t) + e(t) \\ \phi^T(t) &= [y(t-1) \ y(t-2) \ y(t-3) \ y(t-4)] \end{aligned} \right\} \quad (52)$$

の 4 次の AR 過程で、 $\theta(t)$ は

$$\theta(1) = [0.3 \ -0.28 \ 0.46 \ -0.1]^T \quad (53)$$

を初期値として、200 刻ごとに各要素を 10% のステップ状に変動させた。

【 $\theta(t) \equiv \theta$ — 一定の場合】

この場合は、(52) 式と同じであるが、

$$\theta^T = [0.7 \ -0.6 \ 1.0 \ -0.3] \quad (54)$$

とした。いずれの場合も

$$\left. \begin{aligned} \hat{\theta}(0) &= \mathbf{0}, \quad \mathbf{P}(0) = 10^2 \mathbf{I}, \quad \sigma^2 = 1, \quad \delta^2(0) = 0.1 \\ \delta &= 10^{-4}, \quad \lambda_{\min} = 0.9 \end{aligned} \right\} \quad (55)$$

とし、AIC を用いるときの $\hat{\theta}$ の個数 M の値は $M=30$ とした。実験は、1,000 本の見本過程について行い、 $\theta(t)$ の推定精度は、次式で評価した。

$$\hat{\theta}_j(t) = \frac{1}{1000} \sum_{m=1}^{1000} \hat{\theta}_j^{(m)}(t) \quad (j=1, 2, 3, 4) \quad (56)$$

ここで、 $\hat{\theta}_j^{(m)}(t)$ は、 m 番目の見本過程に基づく $\theta_j(t)$ の推定値である。

このとき、Fig. 1, Fig. 2 はそれぞれ $\theta(t)$ が時変の場合と、一定の場合の推定結果を (56) 式で評価したものである。いずれの場合もおおむね良好な結果が得られていることがうかがえる。

Fig. 1 Estimation results (case of time-varying parameters)

つぎに、AIC を用いる際の $\hat{\theta}$ の個数 M の値が推定精度にどのような影響を及ぼすかについて見たのが、Fig. 3, Fig. 4 である。この場合も、1,000 本の見本過程について実験を行い、推定精度は次式で評価した。

$$d(t) = \frac{1}{1000} \sum_{m=1}^{1000} \|\theta(t) - \hat{\theta}^{(m)}(t)\| \quad (57)$$

Fig. 2 Estimation results (case of constant parameters)

Fig. 3 Time evolutions of parameter estimation errors (case of time-varying parameters)

Fig. 4 Time evolutions of parameter estimation errors (case of constant parameters)

Fig. 3 は, $\theta(t)$ が時変の場合の結果であるが, $M=10$ の場合を除いて, M の値が大きくなっても時間経過とともに推定精度はあまり変化しなくなることがうかがえる. Fig. 4 は, $\theta(t)$ が一定の場合の結果であり, Fig. 3 と同様に $M=10$ の場合を除いて, M の値が大きくなっても時間経過とともに推定精度はあまり変化しないことがうかがえる. これらより, M の値があまり小さくは推定精度が劣化するが, あまり大きくしても推定精度はそれほど変化しないこと, M の値を大きくすると計算量も増えることから, M の値は 30 前後が適当なところであると考えられる.

7. おわりに

本論文では, 時変パラメータを推定するための逐次型最小 2 乗推定器に用いる忘却係数を選定するための一つの手法について述べた. その方法は, $\varepsilon(t)$ の統計的挙動に注目し, 推定器から計算される $\varepsilon(t)$ の分散が, その実際の分散に一致するように $\lambda(t)$ を選定するものであった. $\varepsilon(t)$ の統計的挙動の監視は, AIC を用いることにより行われた.

得られた忘却係数を用いるとき, 推定器の収束性は, パラメータ変動が random walk でモデル化される範囲では, ある領域に収束することが示された. また, 数値実験例では, 良好な結果が得られたことが確認された.

しかしながら, スカラー値の忘却係数は, $\theta(t)$ の要素ごとの推定精度の調整はできない. $\theta(t)$ の推定精度を要素ごとに調整するためには, 忘却係数を用いる場合は, ベクトル型の忘却係数⁶⁾を用いるなど, 別の手法を考えることが必要である.

参考文献

- 1) L. Ljung: System Identification, 305/307, Prentice-Hall (1987)
- 2) 中溝高好: 信号解析とシステム同定, 139/140, コロナ社 (1988)
- 3) T. R. Fortescue, L. S. Kershenbaum and B. F. Ydstie: Implementation of Self-Tuning Regulators with Variable Forgetting Factors, Automatica, **17**-6, 831/835 (1981)
- 4) J. M. Mendel: Lessons in Digital Estimation Theory, 140/147, Prentice-Hall (1987)
- 5) S. Dasgupta and Y. Huang: Asymptotically Convergent Modified Recursive Least-Squares with Data-Dependent Updating and Forgetting Factor for Systems with Bounded Noise, IEEE Trans. Information Theory, **IT-33**-3, 383/392 (1987)
- 6) 芳谷直治: ベクトル型忘却係数を用いたパラメータ逐次推定とその実プラントへの適用, 計測自動制御学会論文集, **25**-5, 579/585 (1989)
- 7) 坂本, 石黒, 北川: 情報量統計学, 107/113, 共立出版

(1983)

- 8) R. M. Canetti and M. D. Espana: Convergence Analysis of the Least-Squares Identification Algorithm with a Variable Forgetting Factor for Time-Varying Linear Systems, Automatica, **25**-4, 609/612 (1989)

《付録 1 (38) 式の導出⁷⁾》

よく知られているように, AIC は次式で与えられる.

$$\begin{aligned} \text{AIC} = & -2 \times (\text{モデルの最大対数尤度}) \\ & + 2 \times (\text{モデルの自由パラメータ数}) \end{aligned} \quad (\text{A.1.1})$$

データ $\hat{y}_{t-M+1}$ が与えられたときの, モデル $N[0, 1]$, $N[0, \sigma_\varepsilon^2(t)]$ の最大対数尤度をそれぞれ $L(0, 1)$, $L(0, \sigma_\varepsilon^2(t))$ と書くとき, これらはつぎのようになる.

$$L(0, 1) = -\frac{M}{2} \log 2\pi - \frac{M}{2} \{\hat{\mu}^2(t) + \hat{\sigma}_\varepsilon^2(t)\} \quad (\text{A.1.2})$$

$$\begin{aligned} L(0, \sigma_\varepsilon^2(t)) = & -\frac{M}{2} \log 2\pi - \frac{M}{2} \log \{\hat{\mu}^2(t) \\ & + \hat{\sigma}_\varepsilon^2(t)\} - \frac{M}{2} \end{aligned} \quad (\text{A.1.3})$$

一方, モデル $N[0, 1]$ に対する自由パラメータ数は 0, モデル $N[0, \sigma_\varepsilon^2(t)]$ に対するそれは 1 である. これらおよび (A.1.2), (A.1.3) 式を (A.1.1) 式に代入するとき, (38) 式が得られる.

《付録 2 (44) 式の導出》

文献 8) と同様の方法で, まず (4) 式より

$$\begin{aligned} \mathbf{P}^{-1}(t) &= \lambda(t) \mathbf{P}^{-1}(t-1) + \phi(t) \phi^T(t) \\ &= \gamma(t, 0) \mathbf{P}^{-1}(0) + \sum_{i=1}^t \gamma(t, i) \phi(i) \phi^T(i) \end{aligned} \quad (\text{A.2.1})$$

である. ここで,

$$\gamma(t, i) := \begin{cases} \prod_{j=i+1}^t \lambda(j) & \text{for } i \leq t-1 \\ 1 & \text{for } t=i \end{cases} \quad (\text{A.2.2})$$

であり, $\gamma(t, i)$ はつぎのように書ける.

$$\gamma(t, i) = \gamma(t, j) \gamma(j, i) \quad \text{for } i \leq j \leq t \quad (\text{A.2.3})$$

また, (A.2.2), (7) 式より, $i \leq t-1$ に対して

$$\lambda_{\min}^t \leq \gamma(t, i) \leq \lambda_{\max}^{t-i} \quad (\text{A.2.4})$$

である.

さて, 今 s を t/L の整数部として, (A.2.1) 式の下限は $t \geq sL$ に対してつぎのように評価できる.

$$\mathbf{P}^{-1}(t) \geq \gamma(t, 0) \mathbf{P}^{-1}(0) + \sum_{i=t-L+1}^t \gamma(t, i) \phi(i) \phi^T(i)$$

$$\begin{aligned}
& + \cdots + \sum_{i=t-sL+1}^{t-(s-1)L} \gamma(t, i) \phi(i) \phi^T(i) \\
& = \gamma(t, 0) \mathbf{P}^{-1}(0) + \sum_{i=t-L+1}^t \gamma(t, i) \phi(i) \phi^T(i) \\
& + \cdots + \gamma(t, t-(s-1)L) \sum_{i=t-sL+1}^{t-(s-1)L} \gamma(t-sL, i) \phi(i) \phi^T(i) \quad (\text{A. 2. 5})
\end{aligned}$$

ここで, (A. 2. 4), (6) 式を考慮して

$$\begin{aligned}
\mathbf{P}^{-1}(t) & \geq \alpha_1 \lambda_{\min}^L \mathbf{I} + \alpha_1 \lambda_{\min}^{2L} \mathbf{I} + \cdots + \alpha_1 \lambda_{\min}^{sL} \mathbf{I} \\
& = \alpha_1 \lambda_{\min}^L \frac{1 - \lambda_{\min}^{sL}}{1 - \lambda_{\min}^L} \mathbf{I} \quad (\text{a.s.}) \quad (\text{A. 2. 6})
\end{aligned}$$

を得る. これは, (44) 式の第 1 式である.

$\mathbf{P}^{-1}(t)$ の上限は, つぎのように評価できる. まず, (A. 2. 5) 式と同様の手順で

$$\begin{aligned}
\mathbf{P}^{-1}(t) & \leq \gamma(t, 0) \mathbf{P}^{-1}(0) + \sum_{i=t-L+1}^t \gamma(t, i) \phi(i) \phi^T(i) \\
& + \cdots + \sum_{i=t-(s+1)L+1}^{t-sL} \gamma(t, i) \phi(i) \phi^T(i)
\end{aligned}$$

$$\begin{aligned}
& = \gamma(t, 0) \mathbf{P}^{-1}(0) + \sum_{i=t-L+1}^t \gamma(t, i) \phi(i) \phi^T(i) \\
& + \cdots + \gamma(t, t-sL) \sum_{i=t-(s+1)L+1}^{t-sL} \gamma(t-sL, i) \phi(i) \phi^T(i) \quad (\text{A. 2. 6})
\end{aligned}$$

となる. ここで, (A. 2. 4), (6) 式および $\mathbf{P}(0) = \beta \mathbf{I}$ を考慮して, (A. 2. 6) 式の両辺の期待値をとるとき, つぎのようになる.

$$\begin{aligned}
\mathbf{E} \{ \mathbf{P}^{-1}(t) \} & \leq \frac{\lambda_{\max}^t}{\beta} \mathbf{I} + \alpha_2 \lambda_{\max}^L \mathbf{I} \\
& + \alpha_2 \lambda_{\max}^{2L} \mathbf{I} + \cdots + \alpha_2 \lambda_{\max}^{(s+1)L} \mathbf{I} \\
& = \frac{\lambda_{\max}^t}{\beta} \mathbf{I} + \alpha_2 \lambda_{\max}^L \frac{1 - \lambda_{\max}^{(s+1)L}}{1 - \lambda_{\max}^L} \mathbf{I} \\
& \leq \frac{\lambda_{\max}^t}{\beta} \mathbf{I} + \frac{\alpha_2 \lambda_{\max}^L}{1 - \lambda_{\max}^L} \mathbf{I} \\
& = \lambda_{\max}^L \left(\frac{1}{\beta} + \frac{\alpha_2}{1 - \lambda_{\max}^L} \right) \mathbf{I} \quad (\text{A. 2. 7})
\end{aligned}$$

これは, (44) 式の第 2 式である.