多数の疑似システムを用いたシステム同定の統計力学
A statistical mechanical analysis of system identification using many false systems

三好 誠司
Seiji MIYOSHI
神戸高専 電子工学科
Kobe City College of Tech.
miyoshi@kobe-kosen.ac.jp

岡田 真人
Masato OKADA
東大 新領域，理研 脳総研
Univ. of Tokyo, RIKEN BSI
okada@k.u-tokyo.ac.jp

あらまし

同定を行いたい未知システムそのものの出力は観測できず，そのかわりにその未知システムと似た複数の疑似システムの出力が観測できるモデルについて統計力学的手法を用いた解析を行う．LMS アルゴリズムを用いる場合について理論的な解析を行った結果，ステップサイズパラメータが 1 より小さい場合には疑似システムの数が多いほど，また，疑似システムの多様性が豊かであるほど二乗平均誤差が小さくなるが，ステップサイズパラメータが 1 より大きい場合には逆であることが明らかになった．また，疑似システムの多様性が十分に豊かであるときには，ステップサイズパラメータが小さく，疑似システム数が大きい極限で完全なシステム同定が実現されることが明らかになった．

Abstract

We have analyzed a model in which an output of an unknown system itself can not be observed and outputs of false systems which are similar to the unknown system can be observed. Calculating the mean-squared error of an adaptive filter analytically using statistical mechanics, it has been proven that when the step-size parameter satisfies $\eta < 1$, the larger the number K and the diversity of the false systems are, the smaller the mean-squared error is. On the other hand, when the step-size parameter satisfies $\eta > 1$, the smaller the number and the diversity of the false systems are, the smaller the mean-squared error is. When the diversity of the false systems is rich enough, the unknown system is perfectly identified in the limit of $\eta \to 0$ and $K \to \infty$.

1　まえがき

観測データを用いてその背後にあるデータの生成過程を推定することは一般に学習と呼ばれる．教師つき学習においては観測データは教師の入出力であり，これは例題とも呼ばれる．学習はバッチ学習とオンライン学習 [1] に大別できる．バッチ学習においては与えられたいくつかの例題を繰り返し使用する．この場合，生徒が適切な自由度を持っていればすべての例題に正しく答えられるようになるが，それまでに長い時間が必要である．また，多くの例題を蓄えておくメモリが必要である．これに対してオンライン学習では一度使った例題は捨ててしまう．この場合，過去に使った例題に対して生徒が必ず正しく答えられるとは限らないが，多くの例題を蓄えておくためのメモリが不要であり，また時間的に変化する教師にも追随できるなどの利点がある．これまでに我々はオンライン学習の枠組みでいくつかのモデルの汎化能力について統計力学的手法を用いた解析を行ってきた [2, 3, 4, 5, 6]．

本論文では同様の手法を用いてシステム同定の問題 [7, 8] を議論する．すなわち，同定を行いたい未知システムそのものの出力は観測できず，そのかわりにその未知システムと似た複数の疑似システムの出力が観測できるモデルについて統計力学的手法を用いた解析を行う．適応フィルタの更新に LMS アルゴリズムを用いる場合について理論的な解析を行った結果，ステップサイズパラメータが 1 より小さい場合には疑似システムの数が多いほど，また，疑似システムの多様性が豊かであるほど誤差が小さくなるのに対して，ステップサイズパラメー

タが 1 より大きい場合には逆であることが明らかになる．また，疑似システムの多様性が十分に豊かであるときには，ステップサイズパラメータが小さく，疑似システム数が大きい極限で完全なシステム同定が実現されることが明らかになる．

2 モデル

本論文では 1 個の未知システム，K 個の疑似システム，1 個の適応フィルタを考える．信号処理の分野ではこれらのシステムやフィルタとしてタップ付き遅延線フィルタを考える場合が多いが，本論文ではいわゆる独立理論（the independent theory）[7, 8] に基づき，それらはすべて線形パーセプトロンであるとして議論を進める．線形パーセプトロンの動作については後で詳述する．

未知システム，疑似システム，適応フィルタのタップ係数をそれぞれ $\boldsymbol{A}$，$\boldsymbol{B}_k$，$\boldsymbol{J}$ とする．ただし $k=1,\ldots,K$ である．なお，本論文では簡単のため未知システムのタップ係数，疑似システムのタップ係数，適応フィルタのタップ係数のことをそれぞれ単に未知システム，疑似システム，適応フィルタと呼ぶことにする．未知システム $\boldsymbol{A}=(A_1,\ldots,A_N)$，疑似システム $\boldsymbol{B}_k=(B_{k1},\ldots,B_{kN})$，適応フィルタ $\boldsymbol{J}=(J_1,\ldots,J_N)$ および入力 $\boldsymbol{x}=(x_1,\ldots,x_N)$ は N 次元ベクトルであり，$\boldsymbol{A}$ の各要素 A_i は平均 0，分散 1 のガウス分布にしたがい独立に生成され，不変であるとする．$\boldsymbol{B}_k$ の各要素 B_{ki} は平均 0，分散 1 のガウス分布にしたがい，不変であるとする．また，$\boldsymbol{B}_k$ と $\boldsymbol{A}$ の方向余弦は R_{Bk}，$\boldsymbol{B}_k$ と $\boldsymbol{B}_{k'}$ の方向余弦は $q_{kk'}$ であるとする．$\boldsymbol{J}$ の初期値 $\boldsymbol{J}^0$ の各要素 J_i^0 は平均 0，分散 1 のガウス分布にしたがい独立に生成されるものとし，$\boldsymbol{J}$ と $\boldsymbol{A}$ の方向余弦は R_J，$\boldsymbol{J}$ と $\boldsymbol{B}_k$ の方向余弦は R_{BkJ} であるとする．また，$\boldsymbol{x}$ の各要素 x_i は平均 0，分散 $1/N$ のガウス分布にしたがい独立に生成されるものとする．以上より，

$$\langle A_i\rangle=0,\qquad \left\langle (A_i)^2\right\rangle=1, \tag{1}$$

$$\langle B_{ki}\rangle=0,\qquad \left\langle (B_{ki})^2\right\rangle=1, \tag{2}$$

$$\left\langle J_i^0\right\rangle=0,\qquad \left\langle \left(J_i^0\right)^2\right\rangle=1, \tag{3}$$

$$\langle x_i\rangle=0,\qquad \left\langle (x_i)^2\right\rangle=\frac{1}{N}, \tag{4}$$

$$R_{Bk}=\frac{\boldsymbol{A}\cdot\boldsymbol{B}_k}{\|\boldsymbol{A}\|\|\boldsymbol{B}_k\|},\qquad q_{kk'}=\frac{\boldsymbol{B}_k\cdot\boldsymbol{B}_{k'}}{\|\boldsymbol{B}_k\|\|\boldsymbol{B}_{k'}\|}, \tag{5}$$

$$R_J=\frac{\boldsymbol{A}\cdot\boldsymbol{J}}{\|\boldsymbol{A}\|\|\boldsymbol{J}\|},\qquad R_{BkJ}=\frac{\boldsymbol{B}_k\cdot\boldsymbol{J}}{\|\boldsymbol{B}_k\|\|\boldsymbol{J}\|}. \tag{6}$$

ここで，$\langle\cdot\rangle$ は平均を表す．未知システム $\boldsymbol{A}$，疑似システム $\boldsymbol{B}_k$，適応フィルタ $\boldsymbol{J}$ および方向余弦 $q_{kk'},R_{Bk},R_J,R_{BkJ}$ の関係を図 1 に示す．

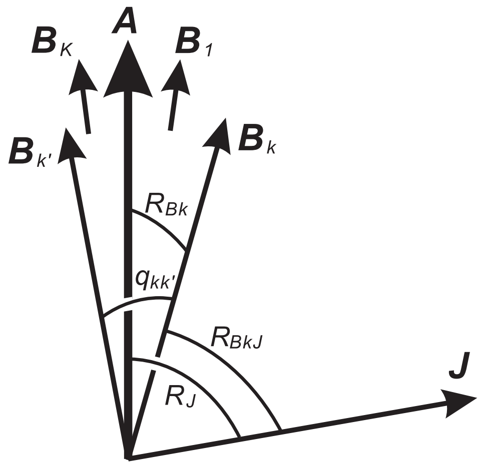

図 1: 未知システム $\boldsymbol{A}$，疑似システム $\boldsymbol{B}_k$，適応フィルタ $\boldsymbol{J}$．$q_{kk'},R_J,R_{Bk},R_{BkJ}$ は方向余弦である．

本論文では，$N\to\infty$ の熱力学的極限を考えることにする．このとき，

$$\|\boldsymbol{A}\|=\sqrt{N},\quad \|\boldsymbol{B}_k\|=\sqrt{N},\quad \|\boldsymbol{J}^0\|=\sqrt{N},\quad \|\boldsymbol{x}\|=1. \tag{7}$$

となる．適応フィルタのノルム $\|\boldsymbol{J}\|$ は一般には時間の経過とともに変化するが，初期値 $\sqrt{N}$ に対する比を l^m とし，適応フィルタの長さと呼ぶことにする．すなわち，$\|\boldsymbol{J}^m\|=l^m\sqrt{N}$ である．m は時間ステップである．

未知システムの出力は $y^m+n_A^m$，疑似システムの出力は $v_k^m+n_{Bk}^m$，適応フィルタの出力は $u^ml^m+n_J^m$ である．ここで，

$$\begin{aligned}
y^m &= \boldsymbol{A}\cdot\boldsymbol{x}^m, && (8)\\
v_k^m &= \boldsymbol{B}_k\cdot\boldsymbol{x}^m, && (9)\\
u^ml^m &= \boldsymbol{J}^m\cdot\boldsymbol{x}^m, && (10)\\
n_A^m &\sim \mathcal{N}\left(0,\sigma_A^2\right), && (11)\\
n_{Bk}^m &\sim \mathcal{N}\left(0,\sigma_{Bk}^2\right), && (12)\\
n_J^m &\sim \mathcal{N}\left(0,\sigma_J^2\right). && (13)
\end{aligned}$$

である．$\mathcal{N}\left(0,\sigma^2\right)$ は平均 0，分散 σ^2 のガウス分布を表す．つまり，未知システムの出力，疑似システムの出力，適応フィルタの出力にはそれぞれ分散 $\sigma_A^2,\sigma_{Bk}^2,\sigma_J^2$ の互いに独立なガウス雑音が重畳されている．またこのとき，y，v_k，u は平均 0，分散 1 のガウス分布にしたがう確率変数となる．

いま，未知システム $\boldsymbol{A}$ と疑似システム $\boldsymbol{B}_k$ の誤差 ϵ_{Bk} を両者の出力の二乗誤差で定義する．すなわち，

$$\epsilon_{Bk}^m\equiv\frac{1}{2}\left(y^m+n_A^m-v_k^m-n_{Bk}^m\right)^2. \tag{14}$$

同様に，疑似システム $\boldsymbol{B}_k$ と適応フィルタ $\boldsymbol{J}$ の誤差 ϵ_{BkJ} を両者の出力の二乗誤差で定義する．すなわち，

$$\epsilon_{BkJ}^m \equiv \frac{1}{2}\left(v_k^m + n_{Bk}^m - u^m l^m - n_J^m\right)^2. \tag{15}$$

適応フィルタ $\boldsymbol{J}$ は入力 $\boldsymbol{x}$ とそれに対する K 個の疑似システム $\boldsymbol{B}_k$ の出力を順番に用いて最小二乗平均 (LMS) アルゴリズムにより学習を行うものとする．すなわち，

$$\boldsymbol{J}^{m+1} = \boldsymbol{J}^m - \eta \frac{\partial \epsilon_{BkJ}^m}{\partial \boldsymbol{J}^m} \tag{16}$$

$$= \boldsymbol{J}^m + \eta\left(v_k^m + n_{Bk}^m - u^m l^m - n_J^m\right)\boldsymbol{x}^m, \tag{17}$$

$$k = \mathrm{mod}\,(m, K) + 1. \tag{18}$$

ここで，η は適応フィルタのステップサイズパラメータであり定数とする．また，$\mathrm{mod}\,(m, K)$ は m を K で割った余りを表す．

更新則を一般化すると以下のように表せる．

$$\boldsymbol{J}^{m+1} = \boldsymbol{J}^m + f_k \boldsymbol{x}^m \tag{19}$$

$$= \boldsymbol{J}^m + f\left(v_k^m + n_{Bk}^m, u^m l^m + n_J^m\right)\boldsymbol{x}^m, \tag{20}$$

$$k = \mathrm{mod}\,(m, K) + 1. \tag{21}$$

ここで，f は更新量を表す関数であり更新則によって決定される．

また，未知システム $\boldsymbol{A}$ と適応フィルタ $\boldsymbol{J}$ の誤差 ϵ_J も両者の出力の二乗誤差で定義しておく．すなわち，

$$\epsilon_J^m \equiv \frac{1}{2}\left(y^m + n_A^m - u^m l^m - n_J^m\right)^2. \tag{22}$$

3 理論

3.1 二乗平均誤差

統計的学習理論の分野においては教師と生徒の誤差の平均（入力に関する平均）は汎化誤差 (Generalization Error) と呼ばれ，これを理論的に求めることが大きな目的である．汎化誤差は適応信号処理の分野においては二乗平均誤差 (Mean-Squared Error) に相当する．個々の疑似システム $\boldsymbol{B}_k$ の二乗平均誤差 ϵ_{Bkg}，適応フィルタ $\boldsymbol{J}$ の二乗平均誤差 ϵ_{Jg} はそれぞれ以下のように計算される．なお，以後は時間ステップを表す添字 m は特に必要な場合を除いて省略する．

$$\epsilon_{Bkg} = \int d\boldsymbol{x} dn_A dn_{Bk} P\left(\boldsymbol{x}, n_A, n_{Bk}\right)\epsilon_{Bk} \tag{23}$$

$$= \int dy dv_k dn_A dn_{Bk} P\left(y, v_k, n_A, n_{Bk}\right) \times \frac{1}{2}\left(y + n_A - v_k - n_{Bk}\right)^2 \tag{24}$$

$$= \frac{1}{2}\left(-2R_{Bk} + 2 + \sigma_A^2 + \sigma_{Bk}^2\right), \tag{25}$$

$$\epsilon_{Jg} = \int d\boldsymbol{x} dn_A dn_J P\left(\boldsymbol{x}, n_A, n_J\right)\epsilon_J \tag{26}$$

$$= \int dy du dn_A dn_J P\left(y, u, n_A, n_J\right) \times \frac{1}{2}\left(y + n_A - ul - n_J\right)^2 \tag{27}$$

$$= \frac{1}{2}\left(-2R_J l + (l)^2 + 1 + \sigma_A^2 + \sigma_J^2\right). \tag{28}$$

ここで積分の実行には，y, v_k, u が平均 0，分散 1 のガウス分布に従うこと，y と v_k の共分散が R_{Bk}，v_k と u の共分散が R_{BkJ}，y と u の共分散が R_J であること，および，n_A，n_{Bk}，n_J はいずれも他の確率変数とは独立であることを利用した．

3.2 巨視的変数の微分方程式とその解

解析を容易にするため，以下の補助的な巨視的変数を導入する．

$$r_J \equiv R_J l, \tag{29}$$

$$r_{BkJ} \equiv R_{BkJ} l. \tag{30}$$

今回，巨視的変数のダイナミクスを記述する連立微分方程式 [9, 10] を熱力学的極限における自己平均性に基づき以下のような決定論的な形で導出した [2]．ここでタップ数 N は疑似システムの数 K よりも十分大きいものとし，また，時間ステップ m をタップ数 N で正規化した値を時間 t としている．

$$\frac{dr_{BkJ}}{dt} = \frac{1}{K}\sum_{k'=1}^{K}\langle f_{k'} v_k\rangle, \tag{31}$$

$$\frac{dr_J}{dt} = \frac{1}{K}\sum_{k=1}^{K}\langle f_k y\rangle, \tag{32}$$

$$\frac{dl}{dt} = \frac{1}{K}\sum_{k=1}^{K}\left(\langle f_k u\rangle + \frac{1}{2l}\langle f_k^2\rangle\right). \tag{33}$$

本論文では未知システム，疑似システム，適応フィルタは線形であるので，これらの連立微分方程式に現れるサンプル平均は以下のように容易に計算することがで

きる．

$$\langle f_k u\rangle = \eta\left(\frac{r_{BkJ}}{l} - l\right), \tag{34}$$

$$\langle f_k^2\rangle = \eta^2\left((l)^2 - 2r_{BkJ} + 1 + \sigma_{Bk}^2 + \sigma_J^2\right), \tag{35}$$

$$\langle f_k y\rangle = \eta\left(R_{Bk} - r_J\right), \tag{36}$$

$$\frac{1}{K}\sum_{k'=1}^{K}\langle f_{k'} v_k\rangle = \eta\left(-r_{BkJ} + \frac{1}{K}\sum_{k'=1}^{K} q_{kk'}\right). \tag{37}$$

本論文では未知システム $\boldsymbol{A}$，適応フィルタ $\boldsymbol{J}$ の初期値 $\boldsymbol{J}^0$ の各要素は平均 0，分散 1 のガウス分布にしたがい独立に生成され，また，$N\to\infty$ の熱力学的極限を考えているので，初期状態においてこれらは直交しており，

$$R_J^0 = 0 \tag{38}$$

である．また，

$$l^0 = 1 \tag{39}$$

である．式 (34)–(39) を用いて連立微分方程式 (31)–(33) は以下のように解析的に解ける．

$$r_{BkJ} = \frac{1}{K}\sum_{k'=1}^{K} q_{kk'}\left(1 - e^{-\eta t}\right), \tag{40}$$

$$r_J = \frac{1}{K}\sum_{k=1}^{K} R_{Bk}\left(1 - e^{-\eta t}\right), \tag{41}$$

$$\begin{aligned}(l)^2 &= \frac{1}{2-\eta}\left[2\left(1-\eta\right)\bar{q} + \eta\left(1 + \bar{\sigma_B^2} + \sigma_J^2\right)\right] \\ &\quad + \left[1 + \frac{1}{2-\eta}\left(\eta\left(1 + \bar{\sigma_B^2} + \sigma_J^2\right) - 2\bar{q}\right)\right] \\ &\quad \times e^{\eta(\eta-2)t} - 2\bar{q}e^{-\eta t},\end{aligned} \tag{42}$$

ここで，

$$\bar{q} = \frac{1}{K^2}\sum_{k=1}^{K}\sum_{k'=1}^{K} q_{kk'}, \tag{43}$$

$$\bar{\sigma_B^2} = \frac{1}{K}\sum_{k=1}^{K}\sigma_{Bk}^2, \tag{44}$$

である．

4　結果と議論

以下では疑似システムと未知システムの方向余弦，疑似システム間の方向余弦，および，それらに重畳される雑音の分散が一様である場合について計算を行う．すなわち，

$$R_{Bk} = R_B, \quad k = 1,\ldots,K, \tag{45}$$

$$q_{kk'} = \begin{cases} q, & k \neq k', \\ 1, & k = k', \end{cases} \tag{46}$$

$$\sigma_{Bk}^2 = \sigma_B^2, \tag{47}$$

であるとする．この場合，式 (43)，(44) は，

$$\bar{q} = q + \frac{1-q}{K}, \tag{48}$$

$$\bar{\sigma_B^2} = \sigma_B^2, \tag{49}$$

となる．

式 (28)，(29)，(40)–(49) を用いて理論的に計算される二乗平均誤差 ϵ_{Jg} のダイナミクスを計算機シミュレーションの結果と重ねて図 2 に示す．計算機シミュレーションは $N = 2000$ で実行し，二乗平均誤差は各時点で 10^4 個のランダム入力に対する誤差の平均を計算することにより計算した．疑似システムの二乗平均誤差 ϵ_{Bg} も重ねて描いた．また，このときの R_J と l のダイナミクスを図 3 に示す．

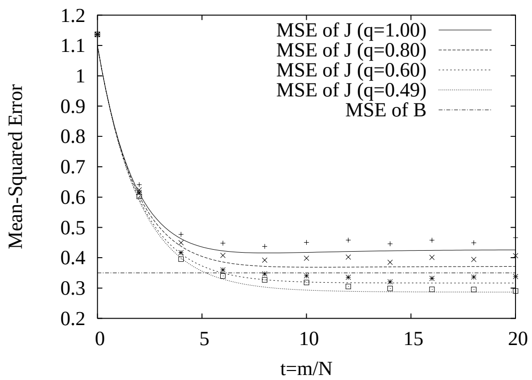

図 2: 二乗平均誤差 ϵ_{Jg} のダイナミクス．理論と計算機シミュレーション．q 以外の条件は $\eta = 0.3, K = 3, R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である．

これらの図において曲線は理論計算の結果を，$+, \times, *, \square$, ■などの印は計算機シミュレーションの結果を表す．また，q 以外の条件は共通で $\eta = 0.3, K = 3, R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である．これらを見ると以下のことがわかる．図 2 より，q が小さいほど，すなわち，疑似システムの多様性が豊かであるほど適応フィルタの二乗平均誤差 ϵ_{Jg} は小さい．特に $q = 0.6$ と $q = 0.49$ の場合には $t = 5$ 付近で適応フィルタの二乗平均誤差が疑似システムの二乗平均誤差よりも小さくなっている．すなわち，このモデルの適応フィルタは疑似システムの出力だけを更新に用いるにもかかわらず，疑似システムよ

りも高性能になりうる．また，図 3 より，疑似システムの多様性が豊かであるほど未知システムと適応フィルタの方向余弦 R_J は大きく，適応フィルタの長さ l は小さいことがわかる．なお，図 2，図 3 において q の最小値 (0.49) を $R_B(0.7)$ の 2 乗としている理由については後述する．

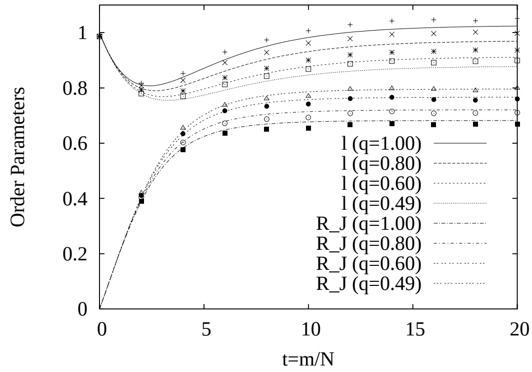

図 3: R_J と l のダイナミクス．理論と計算機シミュレーション．q 以外の条件は $\eta = 0.3, K = 3, R_B = 0.7, \sigma_A^2 = 0.0, \sigma_B^2 = 0.1, \sigma_J^2 = 0.2$ である．

図 2，3 を見ると，適応フィルタの二乗平均誤差 ϵ_{Jg} や R_J, l は $t = 20$ でほぼ定常値に達しているように見えるが，今回巨視的変数が解析的に得られているのでこれらの $t \to \infty$ におけるふるまいについては理論的な洞察が可能である．すなわち，式 (40)–(42) の指数関数のべきの符号に着目することにより $0 < \eta < 2$ でなければ適応フィルタの二乗平均誤差 ϵ_{Jg} や長さ l は発散することがわかる．$0 < \eta < 2$ の場合については，式 (40)–(42) において $t \to \infty$ とすることにより二乗平均誤差 ϵ_{Jg} や R_J, l の定常値は以下のように容易に得られる．

$$r_{BkJ} \to q + \frac{1-q}{K}, \tag{50}$$

$$r_J \to R_B, \tag{51}$$

$$l^2 \to \frac{1}{2-\eta} \times \left(2(1-\eta)\left(q + \frac{1-q}{K}\right) + \eta\left(1 + \sigma_B^2 + \sigma_J^2\right)\right). \tag{52}$$

式 (50)–(52) より以下のことがわかる．$\eta = 1$ の場合には長さ l の定常値は疑似システム数 K や疑似システム間の類似度 q に依存しない．よってこの場合，二乗平均誤差 ϵ_{Jg} や方向余弦 R_J の定常値は K や q に依存しない．$0 < \eta < 1$ の場合には q が小さいほど，また，K が大きいほど l と ϵ_{Jg} の定常値は小さくなり，R_J の定常値は大きくなる．$1 < \eta < 2$ の場合には逆に q が小さ

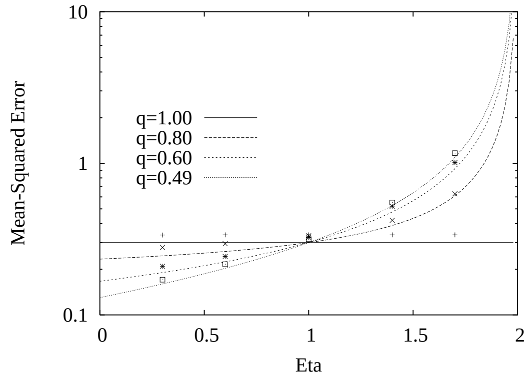

図 4: $K = 3$ の場合の二乗平均誤差 ϵ_{Jg} の定常値．理論と計算機シミュレーション．K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = \sigma_B^2 = \sigma_J^2 = 0.0$ である．

いほど，また，K が大きいほど l と ϵ_{Jg} の定常値は大きくなり，R_J の定常値は小さくなる．

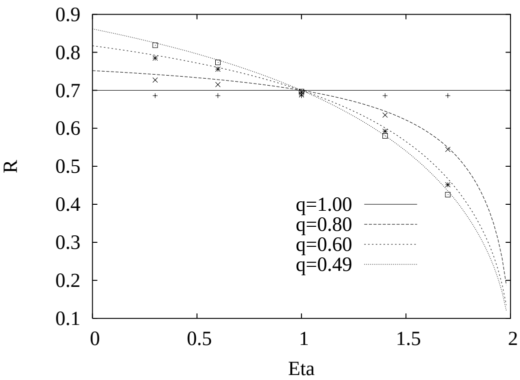

図 5: $K = 3$ の場合の R_J の定常値．理論と計算機シミュレーション．q 以外の条件は K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = \sigma_B^2 = \sigma_J^2 = 0.0$ である．

すなわち，$\eta < 1$ の場合には疑似システムの数が多いほど，また，疑似システムの多様性が豊かであるほど適応フィルタは高性能になれるのに対し，$\eta > 1$ の場合には逆に適応フィルタが高性能になるためには疑似システムの数は少ない方が良く，また，疑似システムの多様性は乏しい方が良い．

式 (52) より $\eta \to 0, K \to \infty$ の極限で $l \to \sqrt{q}$ となるのでこのとき式 (29)，(51) より $R_J \to R_B/\sqrt{q}$ となることがわかる．ところで一般に次元が大きいあるベクトル $\boldsymbol{X}$ があって，$\boldsymbol{X}$ との方向余弦が R_0 であるような二個のベクトル $\boldsymbol{Y}$ と $\boldsymbol{Z}$ を独立に生成したとき，$\boldsymbol{Y}$ と $\boldsymbol{Z}$ の方向余弦は $q_0 = R_0^2$ である [5]．このことから，未知システムとの方向余弦が R_B 一定という拘束条件のもとで疑似システムたちが独立に生成されたほど十分な多様性を

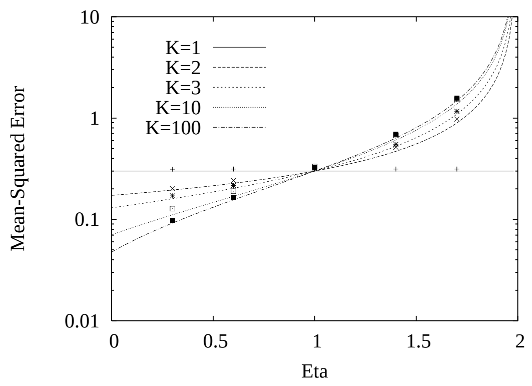

図 6: $q = 0.49$ の場合の二乗平均誤差 ϵ_{Jg} の定常値．理論と計算機シミュレーション．q, K 以外の条件は $R_B = 0.7, \sigma_A^2 = \sigma_B^2 = \sigma_J^2 = 0.0$ である．

持っているとき，ステップサイズパラメータ η が小さく，かつ，疑似システム数 K が多い極限で，またそのときに限り，適応フィルタと未知システムの方向余弦 R_J は雑音の分散によらず 1 になる．R_J が 1 になれば，その後，適応フィルタのすべてのタップ係数を l で割ることにより，完全なシステム同定が実現されることになる．

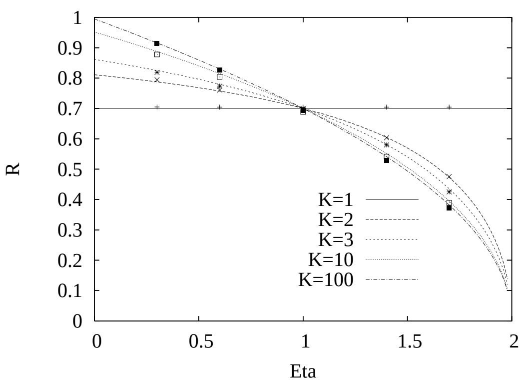

図 7: $q = 0.49$ の場合の R_J の定常値．理論と計算機シミュレーション．q, K 以外の条件は K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = \sigma_B^2 = \sigma_J^2 = 0.0$ である．

適応フィルタのステップサイズパラメータ η と二乗平均誤差 ϵ_{Jg}，方向余弦 R_J の定常値の関係の計算例を図 4〜図 7 に示す．これらのうち，図 4，5 は $K = 3$ 固定で q を変えた場合であり，図 6，7 は $q = 0.49$ 固定で K を変えた場合である．K, q 以外の条件は $R_B = 0.7, \sigma_A^2 = \sigma_B^2 = \sigma_J^2 = 0.0$ である．また，計算機シミュレーションは $\eta = 0.3, 0.6, 1.0, 1.4, 1.7$ で実行し，いずれの場合においても十分定常に達していると判断される $t = 20$ の値をプロットしている．

これらの図から以下のことがわかる．ステップサイズパラメータ η が小さいほど二乗平均誤差 ϵ_{Jg} は小さく，方向余弦 R_J は大きい．また，$\eta = 2$ で ϵ_{Jg} が発散し，R_J がゼロになる相転移現象が起こることが確認される．$\eta < 1$ の場合には，疑似システムの数 K が多いほど，また，疑似システム間の類似度 q が小さいほど（疑似システムの多様性が豊かであるほど）二乗平均誤差 ϵ_{Jg} は小さくなり，方向余弦 R_J は大きくなる．これに対し，$\eta > 1$ の場合には逆になっている．

謝辞

本論文の一部は科学研究費補助金（基盤 (C) 15500151，基盤 (C) 18500183，特定領域 18079003，特定領域 18020007，基盤 (C) 16500093 ）によるものであり，ここに感謝いたします．

参考文献

[1] Saad, D. (ed.), On-line Learning in Neural Networks, Cambridge University Press, (1998)

[2] 岡田 真人，原 一之，三好 誠司，“［チュートリアル講演］アンサンブル学習 ”，信学技報，NC2003-35, pp.7-12, 2003.7

[3] Miyoshi,S., Hara,K. and Okada,M., “Analysis of ensemble learning using simple perceptrons based on online learning theory”, Phys. Rev. E, 71, 036116. March 2005.

[4] Miyoshi,S. and Okada,M., “Analysis of on-line learning when a moving teacher goes around a true teacher”, Journal of Physical Society of Japan, Vol.75, No.2, 024003, Feb. 2006.

[5] Miyoshi,S. and Okada,M., “Statistical mechanics of online learning for ensemble teachers”, Journal of Physical Society of Japan, Vol.75, No.2, 044002, Apr. 2006.

[6] Miyoshi,S., Uezu,T. and Okada,M., “Statistical mechanics of online learning for ensemble teachers”, Journal of Physical Society of Japan, Vol.75, No.2, 084007, Aug. 2006.

[7] Haykin, S., Adaptive Filter Theory, 2nd ed., Prentice-Hall, (1991)

[8] Haykin, S., 武部 幹 訳, 適応フィルタ入門，現代工学社，(1987)

[9] 西森 秀俊，“スピングラス理論と情報統計力学，” 岩波書店，東京,1999.

[10] Nishimori, H., “Statistical Physics of Spin Glasses and Information Processing: An Introduction,” Oxford University Press, (2001)