

JAERI-M
89-175

炉雑音信号に対するGMDHに基づく
階層型モデル

1989年10月

林 光二・篠原 慶邦・鈴木 勝男・鍋島 邦彦

JAERI-Mレポートは、日本原子力研究所が不定期に公刊している研究報告書です。
入手の間合わせは、日本原子力研究所技術情報部情報資料課（〒319-11茨城県那珂郡東海村）
あて、お申しこしてください。なお、このほかに財団法人原子力弘済会資料センター（〒319-11茨城
県那珂郡東海村日本原子力研究所内）で複写による実費頒布をおこなっております。

JAERI-M reports are issued irregularly.

Inquiries about availability of the reports should be addressed to Information Division, Department
of Technical Information, Japan Atomic Energy Research Institute, Tokai-mura, Naka-gun,
Ibaraki-ken 319-11, Japan.

© Japan Atomic Energy Research Institute, 1989

編集兼発行 日本原子力研究所
印刷 日立高速印刷株式会社

炉雑音信号に対するGMDHに基づく階層型モデル

日本原子力研究所東海研究所原子炉工学部

林 光二・篠原 慶邦・鈴木 勝男・鍋島 邦彦

(1989年10月6日受理)

非線形炉雑音信号のモデリングを目的としたGMDHに基づく階層型モデルに関する研究を行っている。本報告は、その第1報であり、GMDHの原理と概要、階層モデルのフィッティング・アルゴリズムならびに線形階層モデルの理論的、数値处理的考察に関するものである。

フィッティング・アルゴリズムは、1入力1出力オープン・ループ型モデルについて処理段階毎に詳細に示し、さらに1変数型(自己回帰型)、多変数フィードバック型モデルへの拡張についても述べた。

次に、統計的信号処理の理論の立場から、Kolmogorov-Gabor型の基礎関数の係数を与える正規方程式を考察し、係数の推定式は高次の相関関数から成立っている事を明らかにした。

さらに、線形階層モデルとARモデルの関数を考察し、線形階層モデルはARモデルと等価な係数を持つこと、また、白色雑音源を仮定することにより、ARモデルと同様に多変数フィードバック・システムの同定とその分解が可能であること等を明らかにした。

最後に、幾つかの線形シミュレーション・データを用いて線形階層モデルのフィッティング試験を行い、係数値や予測誤差に関してARモデルと比較した。

Layered Model Based on GMDH for Reactor Noise Signals

Koji HAYASHI, Yoshikuni SHINOHARA, Katsuo SUZUKI and
Kunihiko NABESHIMA

Department of Reactor Engineering
Tokai Research Establishment
Japan Atomic Energy Research Institute
Tokai-mura, Naka-gun, Ibaraki-ken

(Received October 6, 1989)

The layered model based on GMDH (Group Method of Data Handling) has been studied for the modeling of non-linear reactor noise signals. This is the first report on the study which is concerned with the principle and the overview of the GMDH, the layered model and its fitting algorithm and the theoretical and numerical investigations of the linear layered model.

The fitting algorithm of the layered model with single-input and single-output is described in detail at each processing step and its expansions to univariate and multivariate layered models with feedbacks are also described.

Next, the normal equation which gives the coefficients of a fundamental function of Kolmogorov-Gabor type is considered from the view point of the statistical signal processing. It is found that the estimate of the coefficients consists of high order correlation functions.

Furthermore, the relationship between the linear layered model and AR (Autoregressive) model is considered. From the results, it is found that the linear layered model has coefficients equivalent to those of AR model. It is also found that, by assuming white noise sources, the linear layered model as well as the AR model is capable to identify and to decompose the subsystems in a multivariate feedback system.

Finally, the linear layered model is tested through numerical fitting calculations using some linear simulation data and is compared

with the AR model with respect to the coefficients and the prediction errors.

Keywords : GMDH, System Identification, Model Fitting, Reactor Noise Analysis, Non-linear Signals, Autoregressive Model

目 次

1. はじめに	1
2. GMDHの原理と統計的モデリング	4
2.1 従来の統計的同定計算アルゴリズムとの相違点	4
2.2 GMDHとニューラル・ネットワークの関係	5
2.3 非線形システムの並列化表現	7
2.3.1 非線形システムのVolterra級数展開による連続型表現	7
2.3.2 非線形システムの離散化表現	8
2.3.3 並列化表現	9
2.4 GMDHにおける基礎関数と階層モデル	10
2.4.1 基礎関数とモデルの階層化	10
2.4.2 GMDHにおける完全表現	12
2.5 具体的な基礎関数の導出	12
2.5.1 基礎関数の原型	13
2.5.2 実用的な基礎関数	15
2.6 階層モデルのフィッティング	19
2.6.1 入力データの組合せ発生と基礎関数のフィッティング	20
2.6.2 基礎関数の選択と中間データ	22
2.6.3 最終層の決定	26
2.7 発見的規範とパラメータ	27
2.7.1 発見的規範	27
2.7.2 経験的パラメータ	29
3. 時系列信号に対するアルゴリズム	31
3.1 1入力1出力型階層モデルの同定の問題	31
3.2 データ行列の生成	32
3.3 データ行列Zの行に関する前処理	33
3.3.1 データの相関関数の計算	33
3.3.2 弱い相関関係にある時間遅れ点の棄却	35
3.4 データ行列Z'または中間データ行列Z'の列に関する前処理	36
3.5 トレーニング・データに対する基礎関数のフィッティング	39
3.5.1 基礎関数	39
3.5.2 フィッティング	40
3.5.3 入力データ点の組合せの発生	40
3.5.4 基礎関数のパラメータの解法	41

3.5.5 全組合せのフィッティング	44
3.6 チェッキング・データによる基礎関数の候補の選択と層数の判定	44
3.6.1 評価関数の計算	44
3.6.2 基礎関数の候補の選択	45
3.6.3 層数の決定	46
3.7 中間データの生成と繰返し計算	47
3.8 完全表現の構築	48
3.9 フィードバック型および多変数システムのフィッティング	49
3.9.1 1変数フィードバック・モデル	49
3.9.2 多変数システムのモデリングの問題	49
3.9.3 多変数データ行列の生成	51
3.9.4 多変数データ行列の行に関する前処理	51
4. 基礎関数の係数推定と統計的関数	54
4.1 高次の相関関数とその性質	54
4.2 正規方程式の統計的意味	55
5. GMDHとARモデルの比較	59
5.1 フィッティング構造	59
5.2 基礎関数の係数とARモデルの係数	61
5.3 GMDHの完全表現とARモデルの関係	63
5.4 多変数フィードバック・システムの同定	66
5.4.1 フィードバック・システム	66
5.4.2 ARモデルによる表現	67
5.4.3 GMDHによる表現	70
5.5 スペクトル表現	72
6. フィッティングの実例	73
6.1 フィッティング・データ	73
6.1.1 1変数シミュレーション・データ	73
6.1.2 2変数シミュレーション・データ	73
6.2 モデルのフィッティング	74
6.3 1点予測	75
6.4 1変数モデルのフィッティング結果	76
6.5 2変数モデルのフィッティング結果	82
7. おわりに	88
参考文献	88

Contents

1. Introduction	1
2. Principle of GMDH and statistical modeling	4
2.1 Differences between GMDH and conventional statistical fitting algorithms	4
2.2 Relationship between GMDH and neural networks	5
2.3 Parallel expression of non-linear systems	7
2.3.1 Continuous expression of a non-linear system using Volterra series expansion	7
2.3.2 Discrete expression of the non-linear system	8
2.3.3 Parallel expression	9
2.4 Fundamental function and layered model in GMDH	10
2.4.1 Fundamental function and layered expression	10
2.4.2 Complete description in GMDH	12
2.5 Derivation of practical fundamental function	12
2.5.1 Basis of fundamental function	13
2.5.2 Practical fundamental functions	15
2.6 Fitting of layered model	19
2.6.1 Generation of combinations of input data and fitting of fundamental function	20
2.6.2 Selection of fundamental functions and intermediate data	22
2.6.3 Determination of final layer	26
2.7 Heuristics criteria and parameters	27
2.7.1 Heuristics criteria	27
2.7.2 Heuristics parameters	29
3. Algorithm for time series data	31
3.1 Identification problem of single-input/single-output layered model ...	31
3.2 Generation of data matrix	32
3.3 Preprocessing of row data in matrix Z	33
3.3.1 Calculation of correlation function of the data	33
3.3.2 Rejection of time lag points with weak correlations	35
3.4 Preprocessing of column data in matrix Z'	36
3.5 Fitting of fundamental function to training data	39
3.5.1 Fundamental function	39
3.5.2 Fitting	40
3.5.3 Generation of combinations of input data points	40

3.5.4 Solving method for parameters of fundamental function	41
3.5.5 Fitting of all combinations	44
3.6 Selection of candidates for fundamental functions and determination of number of layers using checking data	44
3.6.1 Calculation of evaluation function	44
3.6.2 Selection of candidates for fundamental functions	45
3.6.3 Determination of number of layers	46
3.7 Generation of intermediate data and iterative calculations	47
3.8 Building of complete description	48
3.9 Fitting of feedback and multivariate systems	49
3.9.1 Univariate feedback model	49
3.9.2 Problem of modeling for multivariate system	49
3.9.3 Generation of multivariate data matrix	51
3.9.4 Preprocessing of row data in multivariate data matrix	51
4. Estimation of coefficients of fundamental function and statistical functions	54
4.1 High order correlation functions and its characteristics	54
4.2 Statistical meanings of normal equation	55
5. Comparison with GMDH and AR model	59
5.1 Fitting structures	59
5.2 Coefficients of fundamental function and AR coefficients	61
5.3 Relationship between complete description of GMDH and AR model	63
5.4 Identification of multivariate feedback systems	66
5.4.1 Feedback system	66
5.4.2 Expression using AR model	67
5.4.3 Expression using GMDH	70
5.5 Spectral expression	72
6. Examples of fitting	73
6.1 Fitting data	73
6.1.1 Univariate simulation data	73
6.1.2 2-dimensional simulation data	73
6.2 Model fitting	74
6.3 One step prediction	75
6.4 Results of fitting to univariate simulation data	76
6.5 Results of fitting to 2-dimensional simulation data	82
7. Conclusion	88
References	88

1. はじめに

原子炉雑音解析の分野において、1970～1980年代は商業用発電炉群の本格的稼働と高速計算機の普及によって、多変数多サンプルデータの大規模な統計的処理の時代となり、膨大な観測データに基いて定常運転時におけるプラント動特性の同定や炉心内の熱水力学的な現象、炉内構造物の振動現象等の局所的な解析、診断が行える様になった。

その間、原子炉雑音解析に用いられる時系列解析の手法も、従来の古典的な相関スペクトル解析に代って、FFT(高速フーリエ変換)によるスペクトル解析、自己回帰モデルに基づくシステム同定とその応用解析手法等の線形システムを対象とした解析手法が、定常線形という枠内ではあるが、計算機の大幅な性能向上と共に十分な活路を見出してきた。

しかし、定常線形手法による信号解析が定着し、炉雑音信号の測定と解析の経験が蓄積されてくると、非線形現象の観測例が次第に増加し、また測定技術の進歩により非定常信号の観測もある程度可能となってきた。これらの解析は、システムの異常診断の見地から次第に重要になりつつあり、これまで種々の理由で後回しにされてきた非線形現象、非定常信号の解析が、いよいよ避けて通れない問題としてクローズアップされるようになってきた。例えば、オランダの商業用PWR Borssele炉で観測された2次系の負荷変動に起因した炉出力の過渡的変動現象¹⁾、原研のNSRRで観測される炉出力の非線形変動現象²⁾等、従来の定常線形手法では役不足といった例が目立つ様になり、どうしても非定常データや非線形データのモデリング手法が必要である。

このうち、非定常炉雑音信号の解析については、サンプル・データを局所定常とみなせる小区間に分けて解析する瞬時ARモデルを用いた方法等の研究が始められた^{3,4)}。一方、非線形炉雑音信号の解析に対しては、以前から古典的なスペクトル解析手法の1つであるバイスペクトルを用いた手法の適用が研究されてきた^{5,6)}が、この手法はモデリング手法ではないために、非線形現象の特徴を間接的に捕えることしかできず、その適用には限界がある。やはり、ARモデルによる定常線形システムのモデリングがこれだけ普及した今日では、その便利さは周知の事実であり、非線形解析の分野においても非線形信号の発生メカニズムを直接モデリングできる手法が待望される。

非線形システムの同定手法は、対象システムの構造が未知の場合には、一般的でかつ実用的な手法というものはほとんど知られておらず、わずかに、Volterra級数展開による方法とGMDH(Group Method of Data Handling)の2つが知られているだけである⁷⁾。このうち前者は、理論的には非線形信号の一般的なモデリングを可能とする手法であるが、その実現に当っては現在の計算機的能力を持ってしても計算効率の面から多大な困難が予想され、実用的ではない。後者は、ソ連のA.G. Ivakhnenkoらによって1968年に提唱され、前者の欠点であった計算効率をサイバネティクスの視点から改良を試みた手法であり、システムの入出力観測データに基づいてパラメトリック・モデルを同定する1種の統計的同定手法である⁸⁻¹¹⁾。

GMDHと多変量回帰モデルに代表される従来の統計的回帰モデルの同定手法との違いは、入出力データの情報処理方法とモデル構造の冗長性にある。GMDHにおいては、本質的複雑さ(定式化の困難といわゆる高次元の呪い)を持つシステムのモデリングを実現するために、サイバネティクス的な情報処理と階層型モデル構造が採用されている。これらは、今日盛んに研究されつつあるニューラル・ネットワーク¹²⁻¹⁴⁾の原型となったパーセプトロンのしきい値的信息選択機構と階層的構造に着想を得たものである。

GMDHにおける情報処理方法とモデル構造の特徴は、次の通りである。

- 1) 入力情報をその重要性に応じて取捨選択して使用する。
- 2) 1つのモデルを多数の微小モデルの階層的な結合(カスケード)で表現する。
- 3) 微小モデルは簡単な構造を持つ多入力1出力の関数で表現する。

この結果、以下の利点が生じる。

- 1) 多変数、非線形システムが簡単にモデリングできる。
- 2) 少ない入出力データでもモデリングが行える。
- 3) 多変数、高次の非線形のモデルが得られる割には計算量が少ない。
- 4) モデリングに際してモデル構造に関する先験的信息が必ずしも必要ではない。

また、欠点として次のものがあげられる。

- 1) 得られたパラメトリック・モデルの物理的な意味付けが困難である。
- 2) モデリングの成否を決める入力情報の取捨選択の基準や微小モデルの表現式の選択に関する理論的決定法が存在せず、もっぱら経験的に見出される基準(ヒューリスティック)が用いられるため、妥当な結果を得るのがやっかいである。
- 3) 得られるモデルが数学的に極値を持つか否か分からないから、気持ちが悪い。
- 4) 手法自身の概念が通常の数学的手法に比べて難解である。
- 5) 現在の線形系の同定手法に比べると費やすべき計算時間が大きい。

GMDHは、産み出されてから約20年を経過し、その間、様々な問題点の指摘と改良が為され、種々の分野でその応用が試みられてきた¹⁵⁻²⁹⁾。特に、局所モデルの選択とその集合である全体のモデルの決定に関しては、自己回帰モデルの分野で出現した赤池の情報量基準AIC³⁵⁾の適用によってある程度の客観的なモデリングができるようになった事²³⁻²⁵⁾は事実であるが、しかしこれはあくまでもできあがったモデルの統計的な評価の1手段に過ぎず、問題の核心となる入力情報の選択基準は、ヒューリスティックから離れる事ができなかった。

このため、日本における制御工学の分野での適用例は、そのほとんどが観測データの量が少なく、しかもこれといって有効な解析手法が見出せない分野に限られていた。原子力の分野でも事情は余り変わらず、化学関係で1例³²⁾、制御関係で1例³⁰⁻³¹⁾見られるだけである。炉雑音信号のモデリングに使用した例は全く知られていない。

そこで、改めてGMDHを時系列解析およびシステム解析の観点から再検討してみると、

- 1) GMDHの基本的計算アルゴリズムに、炉雑音データの様に連続にサンプルされた大容量のデータをどの様に取り入れるのか？
- 2) 同定されたモデル・パラメータと伝達関数との関係は？
- 3) 従来の線形系の手法の同定結果と比べて、GMDHの結果は何が違うか？

4) 非線形性の信号はどの程度モデリングでき、どの程度解析に役に立てられるか？

といった点が問題である事に気付く。

これらの問題点のうち、1)はそのアルゴリズムをきちんと整理すれば自ずと明らかにできよう。2)は理論的にフィッティング過程を議論することにより、ある程度の見通しを得ることができよう。3)、4)は、線形、非線形におけるシミュレーション・データや実測データによるフィッティングを通じて評価できるはずであり、少なくとも線形の場合は確かな結論を得ることができよう。

本論文の目的は、GMDHを非線形信号のモデリング手法と位置付けた場合に、これが非線形炉雑音データに対する有力な解析手法と成り得るか否かを明らかにすることにある。

このため、第2章においては、まず同定手法としてのGMDHを理解するために、従来の統計的同定手法や神経回路網の手法との関係を明らかにし、Ivakhnenkoによって提案されたオリジナル版のGMDHに沿ってその原理と概要を概説する。第3章においては、実際のデータ処理の立場から時系列データに対するアルゴリズムを整理する。

次に第4章では、統計的な信号処理の観点から、Kolmogorov-Gaborの多項式を基礎関数とするGMDHに関して導出される正規方程式の係数の統計的な性質についての理論的な考察を行う。第5章では、GMDHのフィッティング・モデルと物理的モデルとの関係を明らかにする目的で、線形GMDHを取り上げ、線形モデルの典型であるARモデルを対比させながら、GMDHのパラメータとモデル構造との関係を明らかにし、同時にフィードバック・システムの同定の問題を考える。

最後に第6章では、フィッティングの実例として、素性の分かっている線形のシミュレーション信号の幾つかを用いて実際にフィッティング計算を行い、その結果を検討し、GMDHがどの程度実用性を持っているか、またどのような問題点を持つかを議論する。

2. GMDHの原理と統計的モデリング

前節において、GMDHは1種の統計的同定手法であると述べた。本章では、まず、GMDHと従来の統計的同定手法との計算アルゴリズム上の相違点、ならびに今日盛んな研究分野であるニューラル・ネットワーク情報処理との関係を明らかにし、続いてGMDHの中核となる微小モデルの表現式の導出、係数の最小2乗解、微小モデルの階層的結合構造、フィッティング時における微小モデルの選択基準と全体モデルの構築、全体モデルの性質等について議論する。

2.1 従来の統計的同定計算アルゴリズムとの相違点

統計的同定手法とは、モデルの構造をある差分方程式や伝達関数の形に定式化し、システムの入出力データからモデル・パラメータの最小2乗解または最尤解を推定する方法である。

実際的同定計算において、従来の方法では、まず固定したモデル構造に基づいて解析的に最適解を導出し、これを直に入出力データに当てはめるという方法が採用されてきた。これを計算アルゴリズムの観点から見ると、データに対する解析解の計算が直ちにモデル・パラメータの推定になり、解析解とモデル・パラメータは1対1の関係にある構造となっていた。したがって、この種の計算アルゴリズムは、解析解が得易い簡単なシステムに対しては威力を発揮するが、多数の入力変数を持ちしかも非線形であるといった複雑なシステムに対しては、その適用が極端に難しくなってくる。この種の複雑なシステムを同定するために考え出された同定計算アルゴリズムがGMDHである。

GMDHは、モデルの構造に冗長性を与えるために、ある微小モデルを多数個カスケード型に結合したモデル構造を採用している。これが従来の方法との相違点の1つである。ここで、微小モデルとは、離散型システムを考えた場合に現れる時間遅れ点の幾つかを入出力データとする、多入力1出力型の仮想モデルで、その構造は最適解が容易に求まる簡単な構造を持つものとする。この時、全体モデルは、カスケードの第1段に多数の微小モデルが配置され、段を重ねていくと最終段には唯一の出力点を持つという構造となる。さらに、カスケードの段数を未知とすれば、同定対象のシステムの複雑さに合わせていくらかでもモデルの複雑さを増していくことが可能となる。

この種の計算アルゴリズムでは、少なくとも個々の微小モデルのパラメータ同定に関しては、最適解を解析的に求めておいて、これを個々の部分的な入出力データに当てはめるという、従来と同様の方法で容易に推定できる。しかし、1つのシステムの記述を微小に分割したために、全体モデルに全ての部分的入出力関係を盛り込むためには、非常に多数の微小モデルが必要となり、しかもカスケード段数も大きくなってしまおうという新しい問題を生じる。

これを解決するために、GMDHは、全体モデルに盛り込むべき部分的入出力関係の情報を、ある経験則に基づいて予測誤差の観点から取捨選択し、カスケード構造の簡素化を図った。これ

が、GMDHと従来の統計的同定手法とを明確に区別する大きな相違点である。この情報の取捨選択が介入する事により、最終的に得られるモデルは統計的に見て最小2乗解あるいは最尤解になるという数学的な補償を失うが、取捨選択を的確に行えばモデルを最適なものに近づけることができる。この取捨選択の基準は、モデル同定の成否に関わる重要な要素であるが、モデルの収束性といった数学的な議論だけで完全に決定できる性質のものではないため、GMDHにおいては、自然界に見られる情報の取捨選択に関する具体例、生物の神経回路網や植物の品種改良等にヒントを得て作られた経験的な基準を採用している。これが、GMDHがサイバネティクス的方法論の立場からデータ処理を行っているといわれる所以である。

2.2 GMDHとニューラル・ネットワークの関係

ここでは、今日盛んに研究されているネットワーク・メカニズムによる情報処理と、本報告の主題であるGMDHとの関係について簡単に触れておこう。

ネットワーク・メカニズムは、生物の脳に代表される神経回路網(ニューラル・ネットワーク)にヒントを得て産み出された、並列分散型の高度な情報処理を目的とする1種の情報処理手法である。生物の神経回路網における情報処理のメカニズムは、

- 1) 比較的簡単な演算機能を持つユニットで構成される。
 - 2) 多数のユニットが不規則に結合してネットワークを形成している。
 - 3) ユニット相互間の結合の強さが調整でき、教育と学習によって調整される。
 - 4) 情報処理は各ユニットで並列的に行われ、その結果は回路網全体に分散して蓄えられる。
- といったモデルで表わされると考えられ、このメカニズムを情報処理に利用できれば、従来のノイマン型の計算機と直列型の演算アルゴリズムでは実現できない高度の情報処理が可能になると期待される。

生物の神経回路網を模擬した最初のネットワーク・メカニズムは、1958年にRosenblattの提案したパーセプトロン(Perceptron; 学習するパターン識別機械)であった。この型のネットワーク・メカニズムは、古典的なメカニズムに属し、構造上階層的ネットワーク(層状ネットワーク)に分類される⁷⁾。パーセプトロンは、提案以後その基本能力と応用に関する種々の研究がなされたが、1969年にMinsky, Papertによってパーセプトロンの理論的な限界についての論文が発表されて、その流行にひとまず終止符が打たれた。

その後ネットワーク・メカニズムの研究は一時期停滞していたが、1982年にHopfieldによって相互結合型のネットワークが、また1983年にHintonらによってこれを確率的に拡張したボルツマン・マシン(Boltzmann Machine)が提唱され、再び活気を帯びた研究分野となっている。現在のところ、比較的簡単な情報処理ユニットを相互に結合した形態に関するものと、個々のユニットの結合を適切なものに調節する学習能力(自己組織能力)に関するものの2面から研究が進められている。

IvakhnenkoがGMDHを提唱した1968年は、パーセプトロンが普及しだした時期に当る。ここでは、GMDHがパーセプトロンからどのような影響を受けたかを知るために、パーセプトロンの情報処理のメカニズムを簡単に紹介しておこう。

パーセプトロンは、神経回路網を模擬した情報処理用メカニズムを実現するに当り、その構

造や機能を次の様に具体化したネットワーク・メカニズムである。

- 1) ネットワークは、3層の階層型(カスケード)であり、各層は適当な数のユニットからなる。同一層内のユニット間の結合はなく、となり合う層のユニット間では入力層から出力層の向きの1方向の結合だけが存在する。
- 2) ユニットの第1層には物理量変換用の1入力1出力型のしきい素子、また第2、第3層には線形加重和の機能を持つ多入力1出力型のしきい素子が用いられ、取り扱える信号は第1層の入力が実数、他の入出力は全て論理2値(0/1)である。
- 3) 第2、第3層のユニット間の結合だけが強さを調整でき、教育と学習により調整される。他の結合は固定である。
- 4) 学習は、正解と出力との誤差を学習信号に使う方法(誤り訂正型)によって行う。

若干の補則的な説明を加えると、この様な階層型ネットワークでは、メカニズムの動作に関して見ると、ネットワーク中を情報が入力層から出力層へ一度流れて状態変化が終る、という形態になっている。また、ユニットに論理演算型しきい素子が採用された理由は、パーセプトロンの開発目的がパターン認識にあり、論理型演算の方が都合が良かったからに過ぎない。後に実数型のユニットを用いたパーセプトロンも出現している。学習に関しては、ここに採用された方法は、パーセプトロン以降、2乗誤差最小化学習あるいは相関学習、確率的降下法による2乗誤差最小化学習へと拡張されたが、この種のアプローチはいずれも学習するユニット全てに正解を与える必要があるため、多層ネットワークの最終層のところでは学習できないという欠点があることが知られている。

一方のGMDHは、複雑なシステムの同定を目的として開発された計算メカニズムであり、ネットワーク・メカニズムに必要な要因を全て満足する訳ではないが、パーセプトロンと対比させることができる。

- 1) ネットワークは、多層の階層型(カスケード)である。同一層内のユニット間の結合はなく、となり合う層のユニット間では入力層から出力層の向きの1方向の結合だけが存在する。各層のユニット数は、入力層のデータ数に依存する。
- 2) ユニットの低次の多項式関数を用いる。変数の数が入力数となり、各変数の値を入力とし関数値を出力とする多入力1出力型の演算素子となる。
- 3) 各ユニット間の結合の強さは、入力に関しては実数パラメータによって、また出力に関してはOn/Offによって調整でき、学習によって調整する。(すなわち、微小モデルのパラメータ決定と採用すべき微小モデルの選択)
- 4) 学習(フィッティング)は、正解(出力の観測値)と計算出力(予測値)と相関または誤差を学習信号(判定)に使う方法によって、各層毎に段階的に行う。

GMDHに採用された計算メカニズムに関しては、各ユニットで簡単な情報処理を行い、これを多層化するというパーセプトロンのメカニズムがそのまま踏襲されている。

しかし、GMDHの目的は関数フィッティングにあるため、ユニットとしては線形加重和のしきい素子の代りに、多入力1出力の非線形関数を用いている。この結果、ユニットの入力対応の線形加重パラメータの代りに非線形関数パラメータが用いられ、ユニット間の結合は実数値データで行われる。さらにユニットの入力数は実際には2あるいは3と小さいものが用いられる

ため、複雑な関数を表現するためには層数が必要であり、また同一層内のユニット数は入力数によって制限を受けてしまう。

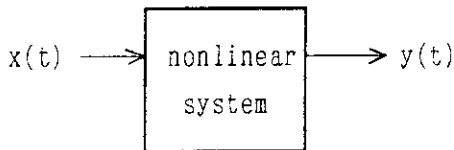
また学習(GMDHにおいては微小モデルのパラメータ決定と)に関しては、相関学習、後に2乗誤差最小化学習と類似の方法を用いている。パーセプトロンでは学習メカニズムからくる制約のために、最終層のユニットしか学習能力が無かった。これを回避するため、GMDHでは第1層から順番に各層毎に学習(係数決定と経路選択)を行っている。これが可能な理由は、パーセプトロンの学習が独立なデータを1セット入力する毎に行っているのに対して、GMDHの場合は、統計的平均に必要な全てのデータを一括して入力し、各層の学習毎に同じデータを繰返し用いているからである。

2.3 非線形システムの並列化表現

GMDHの階層型モデルは、並列に並んだ互に結合関係の無い独立な微小モデルが1つの層をなし、隣り合った層の微小モデルが互に結合して階層構造をつくっている。この階層モデルは、通常の統計的同定に用いられる単一構造のモデルを並列化し、これをさらに階層化したものと考えることができる。このうち、階層構造は、微小モデルを記述する基礎関数の性質で決る。

ここでは、GMDHの各論に入る前段階として、一般的な単一構造のモデルが、如何にして並列な微小モデルに分解されるかを、一般的な非線形システムを例にとって説明する。

2.3.1 非線形システムのVolterra級数展開による連続型表現



入力を $x(t)$ 、出力を $y(t)$ とする任意の非線形動特性システム(Fig.2.1)を考える。いま、出力 $y(t)$ の瞬時値は、Volterra級数展開によって、次の多項式に近似できるものとしよう。

$$y(t) = b_0 + b_1 y^1(t) + b_2 y^2(t) + b_3 y^3(t) + \dots$$

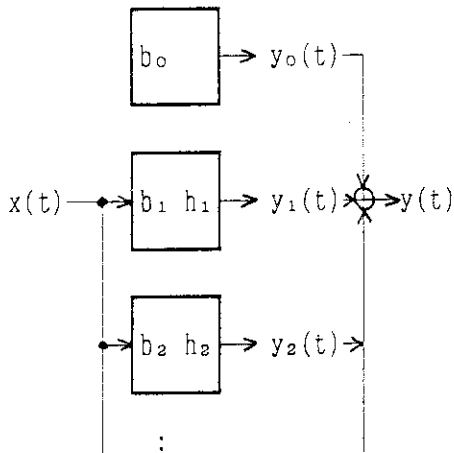
$$= \sum_{i=0}^{\infty} b_i y^i(t) \dots \dots \dots (2.1)$$

ここで、 $\{b_0, b_1, \dots\}$ は係数

さらに、出力 $y(t)$ は、システムの動特性上の遅れ時間 τ を考慮した入力 $x(t-\tau)$ を用いて次の様に表現できるとする。

$$y(t) = b_0 + b_1 \int_0^t h_1(\tau) x(t-\tau) d\tau + b_2 \int_0^t \int_0^t h_2(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) d\tau_1 d\tau_2 + b_3 \int_0^t \int_0^t \int_0^t h_3(\tau_1, \tau_2, \tau_3) x(t-\tau_1) x(t-\tau_2) x(t-\tau_3) d\tau_1 d\tau_2 d\tau_3 + \dots \dots \dots (2.2)$$

ここで、 $\{ h_1, h_2, \dots \}$ は各次数毎のインパルス応答関数を表わす。



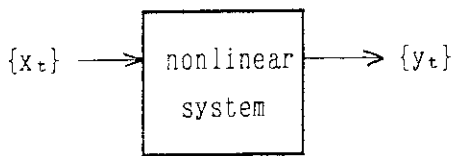
いま、(2.1)式の各項を $\{y_0, y_1, y_2, \dots\}$ と記すと、これらは(2.2)式から、

$$\begin{aligned}
 y_0(t) &= b_0 \\
 y_1(t) &= b_1 \int_0^t h_1(\tau) x(t-\tau) d\tau \\
 y_2(t) &= b_2 \int_0^t \int_0^t h_2(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) d\tau_1 d\tau_2 \\
 &\vdots
 \end{aligned}
 \tag{2.3}$$

Fig. 2.2 Volterra's series

と表わすことができ、非線形システムはFig.2.2に示すように、各次数のインパルス応答関数に対応するサブシステムの集合(総和)表現できることが分かる。

2.3.2 非線形システムの離散化表現



今度は、Fig.2.1 で取り扱った連続型の非線形システムを離散化し、入出力信号 $x(t)$ 、 $y(t)$ を時間間隔 Δt でサンプリングした離散型システムを考える。

いま、最大遅れ時間を $\tau (=M\Delta t)$ とした時、区間 $[t-\tau, t]$ の入力 $x(t)$ および時刻 t の出力 $y(t)$ のサンプル値を、

Fig. 2.3 Discrete nonlinear system $\{ x_{t-M}, \dots, x_t, y_t \}$ (2.4)

と置くと、(2.2)式の離散化表現は次式で与えられる。

$$\begin{aligned}
 y_t = & b_0 + b_1 \sum_{i=1}^M h_{1i} x_{t-i} + b_2 \sum_{i=1}^M \sum_{j=1}^M h_{2ij} x_{t-i} x_{t-j} + b_3 \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M h_{3ijk} x_{t-i} x_{t-j} x_{t-k} + \dots \\
 & \dots \dots \dots (2.5)
 \end{aligned}$$

係数 $\{b\}$ とインパルス応答 $\{h\}$ の積を係数 $\{a\}$ で表わすと

$$\begin{aligned}
 y_t = & a_0 + \sum_{i=1}^M a_{1,i} x_{t-i} + \sum_{i=1}^M \sum_{j=1}^M a_{2,ij} x_{t-i} x_{t-j} + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M a_{3,ijk} x_{t-i} x_{t-j} x_{t-k} + \dots \\
 & \dots \dots \dots (2.6)
 \end{aligned}$$

と与えられる。これが、Kolmogorov-Gaborの多項式である。

2.3.3 並列化表現

非線形システムの離散型表現(2.6)式を、演算の順序の観点から眺めてみると、各項の総和記号Σの中にある係数{ a }と観測データ{ x }の内積は、どれから先に計算しても同じなことに気付く。

いま、各項の内積の値を

$$\begin{aligned}
 y_{t0} &= a_0 \\
 y_{t1,i} &= a_i x_{t-i} \quad \dots\dots\dots (2.7) \\
 y_{t2,ij} &= a_{ij} x_{t-i} x_{t-j} \\
 &\vdots
 \end{aligned}$$

と置くと、非線形システムの離散型表現(2.6)式は、これら内積の総和として次ぎの様に書き換えられる。

$$y_t = y_{t0} + \sum_{i=1}^M y_{t1,i} + \sum_{i=1}^M \sum_{j=1}^M y_{t2,ij} + \dots \quad \dots\dots\dots (2.8)$$

各内積{ $y_{t0}, y_{t1,i}, y_{t2,ij}, \dots$ }は、ある特定の遅れ時間に対応する入力変数 x の成分と出力変数 y の部分的な成分との関係を表現していることから、これをある特定の入出力成分間に存在する微小なサブシステムと考えることもできる。この仮定の微小サブシステムは掛け算器と係数器とで構成されるが、その形は次数毎に異なり、Fig.2.4 に示す様に、第2項では1入力1出力型、第3項は2入力1出力型、第4項は3入力1出力型となる。

これら微小サブシステムは、異なる入力成分(またはその組合せ)毎に独立に存在するから、(2.8)式の表現は、第2項ではM個、第3項では M^2 個、第4項では M^3 個の微小サブシステムが並列に並んだシステムと見なせる。

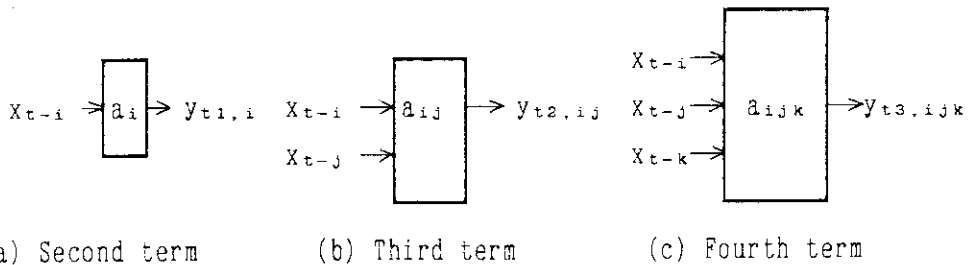


Fig. 2.4 Subsystems in the Kolmogorov-Gabor's equation

逆に、これらの微小なサブシステムを表現する局所モデルを個々に同定し、それらをもとど

おりに結線すれば、もとのシステムを表現するモデルを得ることが可能である。ただし、この考え方は、次数の増大とともにサブシステムの数と関数形の種類も増大するため、これを直接適用してシステム同定を行おうとすると、高々3次までの同定でもサンプル数の大きさ次第では、その係数の数は膨大なものとなり、実現性に乏しくなる。

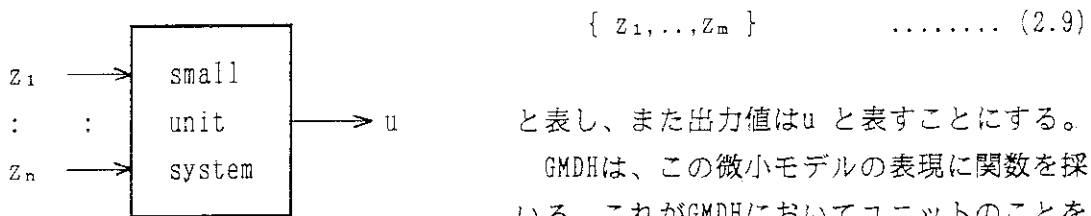
2.4 GMDHにおける基礎関数と階層型モデル

GMDHのモデル構造は、システム同定の観点から見れば微小モデルが、またネットワーク・メカニズムの観点から見れば並列演算要素としてのユニットが、階層的(カスケード)に結合したものとなっている。GMDHでは、この微小モデルあるいはユニットを「部分表現」と呼び、この表現に用いる関数を「基礎関数」と呼ぶ。またカスケード構造全体で表わされるモデルを「完全表現」と呼ぶ。

以下では、微小システムが並列に並んだ単層モデルを如何に階層モデル化していくか、さらに、GMDHの階層モデルはどのような構造になっているか、について述べる。

2.4.1 基礎関数とモデルの階層化

ここでは引続き、2.3.3節と同様にFig.2.3に示した離散型非線形システムを考えよう。入力変数 x の最大遅れ時間点 M までのサンプル値((2.4)式)の中から、 m 個の任意の遅れ点を、入力とするFig.2.5 の様な微小サブシステムを考える。ここで、任意の m 点の入力値は、



と表し、また出力値は u と表すことにする。

GMDHは、この微小モデルの表現に関数を採用している。これがGMDHにおいてユニットのことを基礎関数と呼ぶ所以である。

Fig. 2.5 A unit in GMDH

関数としては、一般に多変数の低次多項式が用いられるが、ここではGMDHにおけるモデルの階層化の

説明を行うために、関数の具体的な形は後回しにして、とりあえず関数に g という記号を用いておく。

$$u = g(z_1, \dots, z_n) \quad \dots \dots \dots (2.10)$$

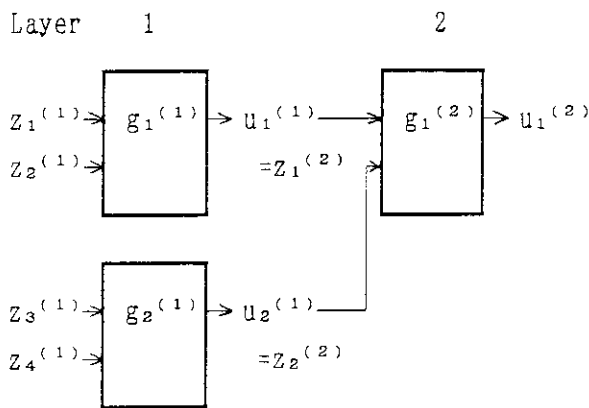
この基礎関数 g は、 n -入力、 1 -出力の微小モデルを表す。そこで、システム全体の入力としては、遅れ時間点が $M+1$ 点あったから、これを全て用いた並列化モデルでは微小モデルの数 N_u は、単純に計算すれば、

$$N_u = M^n \quad \dots \dots \dots (2.11)$$

と膨大な数になる。そこで、この数を制限するために通常は $n=2$ または3くらいが用いられる。ここでは、この Nu 個の微小モデルが単純に並列に並んだ構造を単層構造と呼ぶことにする。単層構造の並列化モデルは、各微小モデルの総和が出力を表す。

$$y_t = \sum_{i=1}^{Nu} u_i = \sum_{i=1}^{Nu} g_i(\{z\}; \{x_t, \dots, x_{t-m}\}) \quad \dots\dots\dots(2.12)$$

この単層構造のモデルで複雑なシステムを表現する方法としては、微小モデルの表現式をより複雑な形にする方法と、もう1つの方法としてモデル構造を階層化する方法がある。



階層化する方法で最も簡単な方法は、単層の並列化モデル中の各微小モデルの後に、さらに同じ形の微小モデルを配置する方法である。

この例をFig.2.6に示す。ここでは、議論を簡単にするために、微小モデルの入力数を2としている。図中、入出力 z, u と微小モデル g に付けられた記号(1)、(2)は層数を示す。

第1層の各微小モデルの出力 u を、第2層の各微小モデルの入力に結線することにより、モデル構造の階層化がはかれる。

一般に、この階層化を第 K 層まで行った階層化モデルでは、出力 y_t は第 K 層の Nu_k 個

Fig.2.6 Combination-paths between 1-st and 2-nd layer units.

の微小モデルの総和として次の様に表される。

$$y_t = \sum_{i=1}^{Nu_k} u_i^{(k)} = \sum_{i=1}^{Nu_k} g_i^{(k)}(\{z^{(k)}\}) \quad \dots\dots\dots(2.13)$$

ただし、第 K 層の微小モデルは、その入力として前層($K-1$)の各微小モデルの出力を用いている。したがって数学的には、階層構造は関数 g の中に同じ形式の関数を次々に入れていく、という入れ子型の関数となることがわかる。

$$u^{(k)} = g^{(k)}(z_1^{(k)}, \dots, z_n^{(k)}) = g^{(k)}(g_1^{(k-1)}, \dots, g_n^{(k-1)}) \quad \dots\dots\dots(2.14)$$

2.4.2 GMDHにおける完全表現

GMDHにおける階層モデルは、(2.13)式において $Nu_k=1$ と置いた場合、すなわち第K層には微小モデルが唯1つしか存在しない場合のモデルに相当する。

$$y_t = u^{(k)} = g^{(k)}(\{z^{(k)}\}) \dots\dots\dots(2.15)$$

これが、GMDHにおける完全表現である。

いま、基礎関数 g として2つの入力を持つ関数が採用されているとしよう。この時、第K層の入力 $z_1^{(k)}, z_2^{(k)}$ は、前層の2つの基礎関数の出力によって表され、以下、層をさかのぼる毎に、各入力とそれを表わす基礎関数は2倍ずつ増加し続け、最終的には第1層を形成する基礎関数の数は 2^{k-1} 個、またその入力データの個数は 2^k 個となる(Table 2.1参照)。

Fig.2.7に、 $k=5$ の完全表現の例を示す。部分表現が2のべき乗でピラミッド状に集って、完全表現を作っていく様子が分かる。仮に、基礎関数が3変数である時は、この形状は3のべき乗に、また4変数では、4のべき乗となっていく。

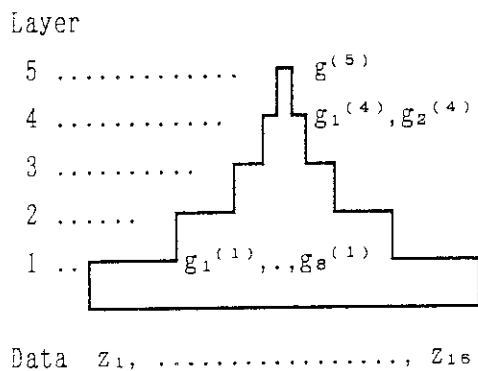


Table 2.1 Number of partial descriptions in each layer. (input=2)

Layer	Input data	Patial description
K	2	1
K-1	4	2
:	:	:
2	2^{K-1}	2^{K-2}
1	2^K	2^{K-1}

Fig.2.7 Complete description with Pyramid-like structure.

2.5 具体的な基礎関数の導出

ここでは、基礎関数の具体的な形について議論する。カスケードのユニットとしての基礎関数は、動特性モデルのフィッティングを目的とした場合に、理論的また実際の計算の見地から次の特徴を持つ必要がある。

- 1) 簡単な関数形であること(パラメータ同定が容易であること)
- 2) 入力数に対応する変数の数が2以上であること(階層化できること)、
- 3) 線形、低次の非線形の入出力関係が表現でき、その階層化とともにモデルがより詳細、より高次の非線形、より多変数なものへと拡張できること、
- 4) 複数の関数形のを混用せずに、一貫して唯1種類の関数形用いること。
- 5) 少ない数の基礎関数、層数で全体を十分表現できること。

上記の観点から検討してみると、Kolmogorov-Gabor(以下KGと略す)の多項式(2.6式)の第3項や第4項の内積表現(Fig.2.4(b),(c))は、基礎関数の候補として極めて有力であることが分かる。しかし、低次の内積を無視して高次の内積だけフィッティングする事はできないため、そのままの形では基礎関数にはならない。そこで、まず内積表現に少し工夫を加え、0次から数次までの内積を1つにまとめた基礎関数の原型を導出し、この性質を検討した後、実用的な基礎関数を紹介する。

2.5.1 基礎関数の原型

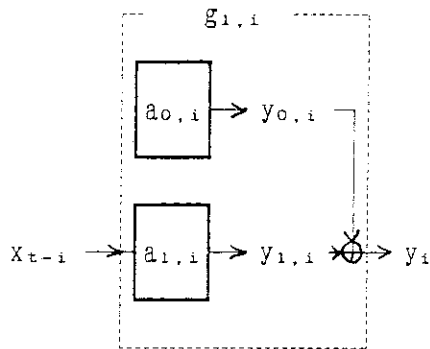
(1) KG多項式の1次近似

はじめに、(2.6)式を第2項で打ち切った以下の1次近似式を検討してみよう。

$$y_t \sim a_0 + \sum_{i=1}^M a_{1,i} x_{t-i} \quad \dots\dots\dots (2.16)$$

この式は、出力 y_t を x_t とすれば、DC値付きの自己回帰モデルと全く同じになる。上式中第1項は入力 x_{t-i} に依存しない定数項で、第2項は入力を x_{t-i} とする1入力1出力型のM個の微小サブシステムの総和と見なすことができる。

そこで、第1項を入力 x_{t-i} 毎に分解して第2項とまとめ、1入力1出力型の関数を作ってみよう。前式の第1項の係数の分解を



$$a_0 = \sum_{i=1}^M a_{0,i} \quad \dots\dots\dots (2.17a)$$

$$y_{0,i} = a_{0,i} \quad \dots\dots\dots (2.17b)$$

と定義し、これを第2項の内積 $y_{1,i}$ と組合せて、次の関数 g を作る。

$$g_{1,i}(x_{t-i}; a_{0,i}, a_{1,i}) = a_{0,i} + a_{1,i} x_{t-i} \quad \dots\dots\dots (2.18)$$

Fig.2.8 Block diagram of Eq.(2.18).

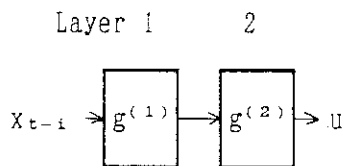
この関数は、Fig.2.8に示す様に、定数項をDCバイアス値として取込んだ1入力1出力型の微小サブシステムとなる。これを用いれば(2.16)式は、

$$y_t \sim \sum_{i=1}^M g_{1,i} = \sum_{i=1}^M (y_{0,i} + y_{1,i}) \quad \dots\dots\dots (2.19)$$

と書き換えられ、新たに定義した微小サブシステムM個の出力の総和としての表現を与える。

ここで導出した関数 g (2.18式)は、もとの近似関数がDC値付きの自己回帰モデルと同じである事からも分る様に、先に述べた基礎関数の条件2)を満たさず、このままでは階層化できない。

例えば、第2層の出力は第1層の出力を入力として形式的に



$$u^{(2)} = g^{(2)}(g^{(1)}) \dots\dots\dots (2.20)$$

ここで、

$$g_{1,i}^{(1)} = a_{0,i}^{(1)} + a_{1,i}^{(1)} x_{t-i}$$

$$g_{1,j}^{(2)} = a_{0,j}^{(2)} + a_{1,j}^{(2)} g_{1,i}^{(1)}$$

Fig.2.9 I/O paths of the function in Eq.(2.18) (2.20)式は、と計算できるが、その接続はFig.2.9に示す通りであり、

$$u^{(2)} = a_{0,j}^{(2)} + a_{1,j}^{(2)} (a_{0,i}^{(1)} + a_{1,i}^{(1)} x_{t-i})$$

$$= (a_{0,j}^{(2)} + a_{1,j}^{(2)} a_{0,i}^{(1)}) + (a_{1,j}^{(2)} a_{1,i}^{(1)}) x_{t-i}$$

$$= a_0' + a_1 \cdot x_{t-i} \dots\dots\dots (2.21)$$

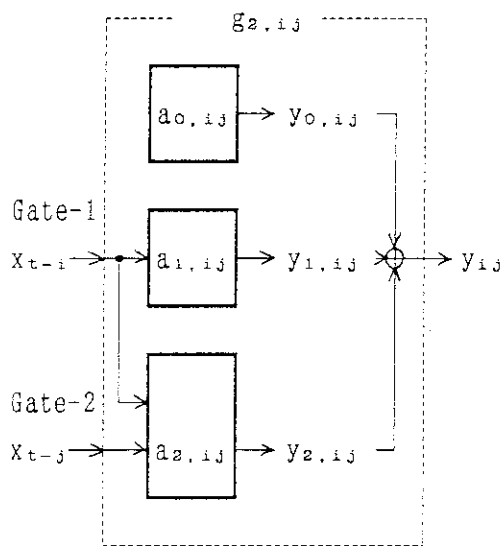
となり、その内容は第1層と全く同じである。すなわち、この関数gは入力を1つしか持たないために第2層以降の経路選択の自由度は無く、また新しい情報の縮約は起らない。

線形の基礎関数で多入力とする方法は、後に述べることにする。

(2) KG多項式の2次近似

次に、KG多項式(2.6)式の第3項で打ち切った2次近似式を検討してみよう。

$$y_t = a_0 + \sum_{i=1}^M a_{1,i} x_{t-i} + \sum_{i=1}^M \sum_{j=1}^M a_{2,ij} x_{t-i} x_{t-j} \dots\dots\dots (2.22)$$



上式中第1項、第2項は前に説明した通りであるが、第3項は2つの入力 x_{t-i} 、 x_{t-j} を持つ2入力1出力型の M^2 個のサブシステムの総和と見なすことができる。

そこで、第1、2項を2つの入力毎に分解して第3項にまとめた、2入力1出力型の関数を作ってみよう。各項はそれぞれ次の様に分解できる。

$$a_0 = \sum_{i=1}^M \sum_{j=1}^M a_{0,ij} \dots\dots (2.23a)$$

$$y_{0,ij} = a_{0,ij} \dots\dots (2.23b)$$

$$a_{1,i} = \sum_{j=1}^M a_{1,ij} \dots\dots (2.24a)$$

Fig.2.10 Block diagram of Eq.(2.25).

$$y_{1,ij} = a_{1,ij} X_{t-i} \dots\dots\dots (2.24b)$$

これらを第3項の内積 $y_{2,ij}$ と足し合わせたものとして、関数 g が次の様に定義される。

$$g_{2,ij}(X_{t-i}, X_{t-j} ; a_{0,ij}, a_{1,ij}, a_{2,ij}) \\ = a_{0,ij} + a_{1,ij} X_{t-i} + a_{2,ij} X_{t-i} X_{t-j} \dots\dots\dots (2.25)$$

この関数は、Fig.2.10に示す様に、定数項をDCバイアスとする2入力1出力型の微小サブシステムとなる。これを用いれば(2.16)式は、

$$y_t \sim \sum_{i=1}^M \sum_{j=1}^M g_{2,ij} = \sum_{i=1}^M \sum_{j=1}^M (y_{0,ij} + y_{1,ij} + y_{2,ij}) \dots\dots\dots (2.26)$$

となる。

ここで導出した関数 g (2.25式)は、2.5に示した基礎関数の条件のうち5)を除くほとんどを満足する。関数の形は十分簡単(パラメータは3つ)であるし、階層化も可能である。また、限定されてはいるが非線形を表現でき、さらに階層化した場合には、層が進むに従って情報の縮約が起ることが分る。(Fig.2.6参照)。したがって、(2.25)式は基礎関数としての資質を十分に備えていることが分る。

2.5.2 実用的な基礎関数

ここでは、KG多項式の1次および2次近似から導出された基礎関数の原型に、結合経路に関する考察を加え、新たに定義される実用的な基礎関数を紹介する。

(1) 非対称2入力型2次多項式

KG多項式の2次近似に基づく(2.25)式は、基礎関数に用いることができる。

$$G_{ij} = g_{2,ij} \dots\dots\dots (2.27)$$

この関数は、入力2、パラメータ3の関数であるから、次の様に記述できる。

$$G_{ij}(X_{t-i}, X_{t-j} ; c_1, c_2, c_3) \\ = c_1 + c_2 X_{t-i} + c_3 X_{t-i} X_{t-j} \dots\dots\dots (2.28a)$$

ここで、

$$c_1 = a_{0,ij} \\ c_2 = a_{1,ij} \\ c_3 = a_{2,ij} \dots\dots\dots (2.28b)$$

この関数は、Fig.2.10に示す様に2つの入力非対称であるため、M個の入力データに対するデータの組合せ数は $NP=M^2$ となり、出力 y_t の近似は(2.26式)と同様に M^2 個の総和で表される。

$$y_t \sim \sum_{i=1}^M \sum_{j=1}^M G_{ij} \dots\dots\dots (2.29c)$$

したがって、必要な基礎関数(ユニット)の数は、 $Nu=M^2$ となり、またフィッティング・パラメータの総数は、 $Nfp=3M^2$ となる。

(2) 対称2入力型2次多項式

(2.28)式の関数は、入力に対して非対称であった。そこで入力が互に逆の2つの関数を1つの関数にまとめてしまえば、必要な基礎関数の数やパラメータ数が削減できそうである。これを次の様に定義しよう。

$$G_{ij} = g_{2,ij} + g_{2,ji} \dots\dots\dots (2.30)$$

この関数は、(2.25)式を代入して整理すると、

$$\begin{aligned} G_{ij} &= (y_{0,ji} + y_{0,ij}) + (y_{1,ji} + y_{1,ij}) + (y_{2,ji} + y_{2,ij}) \\ &= (a_{0,ji} + a_{0,ij}) + a_{1,ji}X_{t-i} + a_{1,ij}X_{t-j} + (a_{2,ji} + a_{2,ij})X_{t-i}X_{t-j} \end{aligned} \dots\dots\dots (2.30b)$$

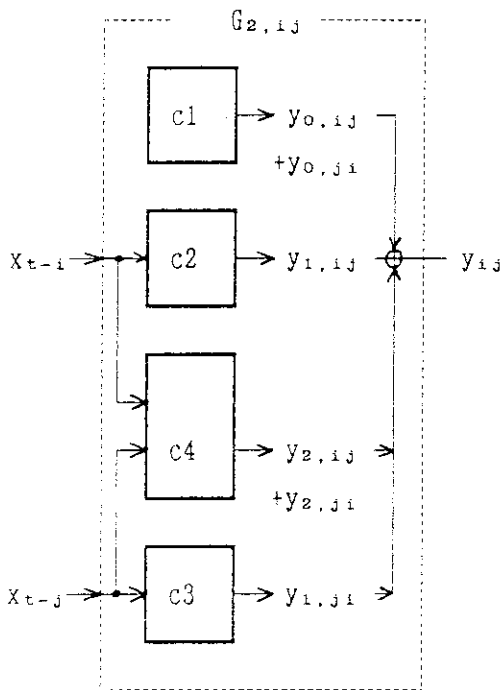


Fig.2.11 Block diagram of Eq.(2.31).

となり、2入力4パラメータの関数となる。

$$\begin{aligned} G_{ij}(X_{t-i}, X_{t-j}; C_1, C_2, C_3, C_4) \\ = C_1 + C_2 X_{t-i} + C_3 X_{t-j} + C_4 X_{t-i} X_{t-j} \end{aligned} \dots\dots\dots (2.31a)$$

ここで、

$$\begin{aligned} C_1 &= a_{0,ji} + a_{0,ij} \\ C_2 &= a_{1,ji} \\ C_3 &= a_{1,ij} \\ C_4 &= a_{2,ji} + a_{2,ij} \dots\dots (2.31b) \end{aligned}$$

この式を図示するとFig.2.11になる。また出力 y_t の近似は、次の様に書き換えられる。

$$y_t \sim \sum_{i=1}^{M/2} \sum_{j=1}^{M/2} G_{ij} \dots\dots\dots (2.32)$$

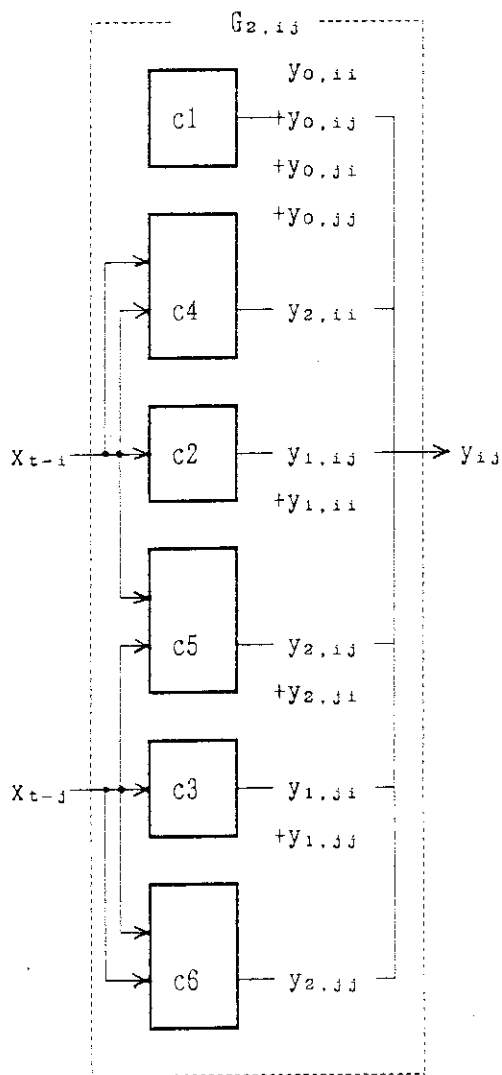
すなわちこの関数では、M個のデータに対するデータの組合せ数は $Np=M^2/2$ となり、必要な基礎関数(ユニット)の数も $Nu=M^2/2$ と半分になる。またパラメータは1つ増えるため、パラメータの総数は、 $Nfp=4M^2/2$ となる。

(3) 重複項付き対称入力型2次多項式

与えられた2つの入力から得られるだけの情報を取り出すという観点から、対称2入力型2次多項式(2.31式)をさらに発展させ、入力の対称性に加えて、2つの入力に同一のデータを印加した場合の2つを加えた、以下の形式のものが考えられる。

$$G_{1j} = g_{2,11} + g_{2,1j} + g_{2,j1} + g_{2,jj} \quad \dots\dots\dots (2.33)$$

先の例と同様に、上式の各項に(2.25)式を代入して整理すると、以下の2入力6パラメータの関数が得られる。この関数はKolmogorov-Gabor型の2次多項式と呼ばれ、一般にGMDHの基礎関数として非線形システムの同定に用いられている。



$$G_{1j}(X_{t-i}, X_{t-j}; C_1, C_2, C_3, C_4, C_5, C_6) = C_1 + C_2 X_{t-i} + C_3 X_{t-j} + C_4 X^2_{t-i} + C_5 X_{t-i}X_{t-j} + C_6 X^2_{t-j} \quad \dots\dots\dots (2.34a)$$

ここで、

$$\begin{aligned} C_1 &= a_{0,11} + a_{0,j1} + a_{0,1j} + a_{0,jj} \\ C_2 &= a_{1,11} + a_{1,j1} \\ C_3 &= a_{1,1j} + a_{1,jj} \\ C_4 &= a_{2,11} \\ C_5 &= a_{2,j1} + a_{2,1j} \\ C_6 &= a_{2,jj} \end{aligned} \quad \dots\dots\dots (2.34b)$$

この関数では、1つの入力の2乗積も考慮するため、M個のデータに対するデータの組合せ数は、 $Np=M(M-1)/2$ となる。この入力の組合せの全ての基礎関数の出力の総和を計算すると、

$$y_t < \sum_{i=1}^{M/2} \sum_{j=2}^{M/2} G_{1j} \quad \dots\dots\dots (2.35)$$

となり、毎回第4項、第6項の2乗積を計算す

Fig.2.12 Block diagram of Eq.(2.34).

るために、(2.26式)あるいは(2.32式)の場合に比べて2乗積の重複の分だけ大きくなる。この重複項は、より少ない数の基礎関数でシステムを表現する、という前に述べた基礎関数の条件5)の実現化の1手段である。

ここで、基礎関数の2つの入力の組合せ(i,j)を、便宜上p(p=1,...,Nu)という記号で表し、基礎関数もG_pと表すことにしよう。M(M-1)/2個の基礎関数から任意のNu個を取り出して、出力y_tを近似した時、

$$y_t \sim \sum_{k=1}^{Nu} G_p \quad \dots\dots\dots (2.36a)$$

が成り立ったとすると、必要な基礎関数の数Nuは、

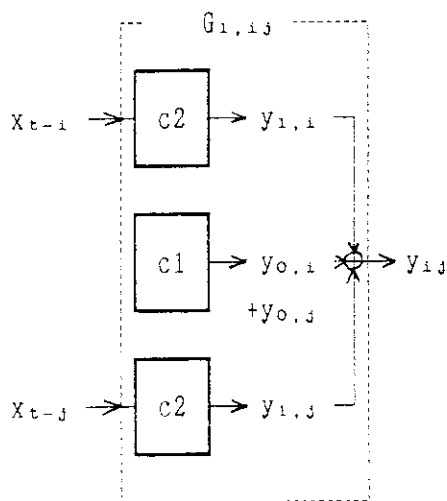
$$Nu < M(M-1)/2 \quad \dots\dots\dots (2.36b)$$

とさらに少なくなる。フィッティング・パラメータの総数は、N_fp=6Nuとなる。

(4) 2入力1次多項式

前節において、KG多項式の1次近似から導出される関数は1入力1出力型の関数であるため、階層構造にならないことを述べた。しかし、この関数はもともと並列化の過程で導出された関数であり、入力点の異なる幾つかの関数を合成(加算)して1つの関数に定義すれば、複数個の入力をもつ基礎関数とすることができる。ここでは、いま、2つの入力X_{t-i}, X_{t-j}に関する(2.18)式の合成を考えよう。

$$G_{1,ij} = g_{1,i} + g_{1,j} = (a_{0,i} + a_{0,j}) + a_{1,i} X_{t-i} + a_{1,j} X_{t-j} \quad \dots\dots\dots (2.37)$$



この関数Gは、係数を整理すると2入力3パラメータの関数となる。

$$G_{1,ij}(X_{t-i}, X_{t-j}; C_1, C_2, C_3) = C_1 + C_2 X_{t-i} + C_3 X_{t-j} \quad \dots\dots\dots (2.38a)$$

ここで、

$$\begin{aligned} C_1 &= a_{0,i} + a_{0,j} \\ C_2 &= a_{1,i} \\ C_3 &= a_{1,j} \quad \dots\dots\dots (2.38b) \end{aligned}$$

Fig.2.13 Block diagram of Eq.(2.38).

この式を関数化したのがFig.2.13である。M個の入力に対して、その組合せはN_p=M²あるが、2つ

の入力は対称であり、しかも内部の演算は線形であるため M^2 個の組合せのうちほとんどは重複する。したがって、重複を全く許さなければ基礎関数の数は、 $Nu=M/2$ となる。組合せ (i, j) を記号 p で表せば、(2.19)式は、次の様に書き換えられる。

$$y_t \sim \sum_{p=1}^{M/2} G_p \quad \dots\dots\dots (2.39)$$

また、フィッティング・パラメータの総数は、 $Nfp=3M/2$ となる。

この型の基礎関数は、1次のKolmogorov-Gabor型の多項式と呼ばれ、線形システムのフィッティングに用いられている。

2.6 階層モデルのフィッティング

GMDHにおける階層的モデルのフィッティングは、以下の手順で行われる。

1) データの予備的選別

与えられたデータから相関等の情報に基づいてフィッティングに用いるデータを選別しておく。

2) 入力データの組合せ発生

基礎関数の入力数に対応する全組合せを作る。

3) 基礎関数のフィッティング

入力データの組合せに対応する個々の基礎関数をそれぞれ独立にフィッティングする。

4) 基礎関数の選別

フィッティングされた基礎関数がある基準で選別し、候補を残す。

5) 中間入力データの計算

選別した基礎関数とそれに対応する入力データからその出力値を計算し、これを次層の入力データとする。

6) 繰り返し処理と最終層の決定

2~5の手順を各層毎に繰り返し、上位層の処理を進める。さらに、別の基準に基づいて階層構造の最大層数の決定を行う。

7) 完全表現の構築

最大層数が決定した段階で、最終層の基礎関数の候補から唯1つを選別し、この入力を第1層まで溯り、各層毎に採用すべき基礎関数を決定する。これらを完全表現とする。

上記の手順3)の、階層モデルの各基礎関数(カスケード構造の各ユニット)のフィッティングは、通常最小2乗法を用いて行われる。具体的には、まず基礎関数からパラメータに関する最小2乗解を与える理論式を導出し、この式を入出力データにフィッティングして各パラメータを決定する。個々のフィッティングに必要な入力データは、第1層では観測データを、また第

2層以降では前層の該当する基礎関数の出力の計算値を用いる。一方、フィッティングに必要な出力データは、どの層でも同一の観測データを繰返し用いる。

同一の出力観測データを繰返し用いる理由は、パーセプトロンが最終層でしか学習できないという制約に由来しており、類似のメカニズムを採用しているGMDHでは、同定処理を各層毎に分け、既知の下位層の結果を利用して1層ずつ上位層の処理をしていく、という逐次的な処理方法を取っている。

一方、階層構造の決定に関しては、最適な(またはそれに近い)階層構造をもたらす基礎関数の結合理論は知られておらず、基礎関数のフィッティングがそうであった様に、基礎関数の結合決定も1層ずつ逐次的に進めながら、全ての可能な結合に対するある種の検定を行って決定する必要がある。

手順2)の、与えられた入力データから考え得る全ての組合せを作り、これに対して基礎関数のフィッティングを行うという処理は、上記の逐次的処理を実現する上で重要である。この理由は、各層の基礎関数のフィッティングに際しては、最終層の層数や各層を形成する基礎関数の数が未知で、しかもどの基礎関数が上位層と結合するか事前には分からないから、全ての候補をフィッティングしておく必要があるのである。

ところが、このような逐次的処理による階層モデルの同定方法は、多層化を可能とする一方で、次の様な困難をもたらす。入力の組合せ数は上位の層へ向って基礎関数の入力数のべき乗で増大するから、多かれ少なかれモデルは発散してしまうのである。この問題を防止し、最適なモデルに近づける役目を担っているのが、手順1)、4)、6)である。

手順1)、4)は、ある判定基準に基づいてデータ点または基礎関数の数を選別(削減)し、階層構造に組み入れられるデータ点や基礎関数の候補を限定することによって、モデルの発散(計算量の増大)を抑制するとともに、同定されるモデルを最適なものに近づける機能を果たす。また、手順6)は、手順1)、4)とは異なる判定基準に基づき、階層モデルの候補の中から最終結果を決定する機能を果たす。

これらに使用される判定基準を「発見的規範(Heuristics)」と呼ぶ。発見的規範は、統計的情報理論に基づいて導出されたものではなく、あくまでも経験的に導かれたものである。この発見的規範は種々のものが提案されている。

手順4)、6)の基礎関数の取舍選択や最終層の決定には、通常、フィッティング後の基礎関数の予測誤差を用いた評価関数が用いられ、手順1)ではデータ点相互間の相関係数が用いられる。

2.6.1 入力データの組合せ発生と基礎関数のフィッティング

ここでは、手順2)と3)についてももう少し詳しく説明する。いま、あるシステムの入出力に関する観測データとして、時間間隔 Δt でサンプルされた離散型データ

$$\{ x_t, y_t \} (t=1, \dots, N) \quad \dots \dots \dots (2.40)$$

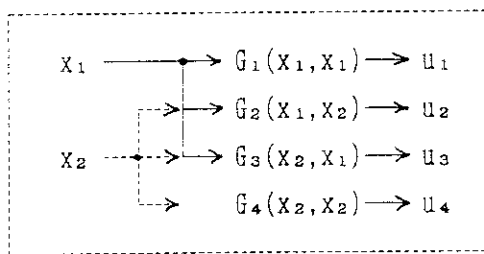
が与えられているとする。ただし、 N は十分大きな数とする。この時、このシステムを同定する問題を考えよう。システムに仮定される最大遅れ時間点を M とすれば、GMDHの階層モデルは、

(N-M)個のデータ・セットDt

$$D_t = \{ x_{t-1}, x_{t-2}, \dots, x_{t-M}, y_t \} \quad (t=M+1, \dots, N) \quad \dots\dots\dots(2.41)$$

に対してフィッティングを行えば良いことになる。GMDHの基礎関数に2入力型を採用すると、第1層の基礎関数の2つの入力には、それぞれ任意の2点が入力される。

$$\begin{aligned} z_i &= x_{t-i} \\ z_j &= x_{t-j} \quad (1 \leq i, j \leq M) \end{aligned} \quad \dots\dots\dots(2.42)$$



例として、最大遅れ時間点Mが2の場合を見てみよう。基礎関数の入力データは、全部で

$$\{ x_{t-1}, x_{t-2} \} \quad \dots\dots\dots(2.43) \quad (t=3, \dots, N)$$

の2点である。これらのデータ点を基礎関数に入力する場合は、その組合せは4通りあり、基礎関数、その出力はともに4つ存在する。データ点を x_1, x_2 と記すと、その関係はFig.2.14の様に図示できる。結合経路から見ると、各データ点は4個の基礎関数の中の3個に接続されているのが分かる。

一般に、データ点がM点与えられている場合には、その組合せ数 N_p 、各データ結合の結合経路の数 N_c は、それぞれ

$$N_p = M^2 \quad \dots\dots\dots(2.44a)$$

$$N_c = M+(M-1) \quad \dots\dots\dots(2.44b)$$

で与えられる。したがって、層を増す毎にこれらの数も増大し、何等かの方法で数を削減する必要がある訳である。数の削減は後に述べる事として、ここでは与えられた1組の入力データ (z_i, z_j) を如何にフィッティングするかについて述べる。

いま、2入力型のc個のパラメータを持つ基礎関数、

$$u_{ij} = G_{ij}(z_i, z_j ; c_{1,ij}, \dots, c_{c,ij}) \quad \dots\dots\dots(2.45)$$

を考える。(2.41)式のデータに対してフィッティングすると、第1層目の基礎関数 $G_{ij}^{(1)}$ の値(出力 u_{ij})とシステムの出力の観測値 y_t は、

$$y_t \sim \sum_i \sum_j G_{ij}^{(1)} \quad \dots\dots\dots(2.46)$$

の関係で表され、第1層の基礎関数は出力 y_t の1成分を表現しているのが分かる。GMDHの階層モデルのフィッティングでは、第2層以降基礎関数が、階層的結合を通じて下位層の基礎関数を取り込み、次第に出力 y_t に近づいて行く様にするのが目的であるから、個々の基礎関数のフィッティングの目標値は、システムの観測出力 y_t である。

したがって、各層における個々の基礎関数のフィッティングは、発生させた任意の入力データの組 $\{z_i, z_j\}$ と観測出力 y_t に対して、

$$y_t \sim G_{ij}^{(k)}(z_i^{(k)}, z_j^{(k)}) \dots\dots\dots(2.47)$$

と行われる。

基礎関数 G_{ij} の c 個のパラメータ $\{c_{k,ij}\}$ は、データ・セット D_d に出力 y_t と近似誤差 $\varepsilon_{t,ij}$

$$\varepsilon_{t,ij} = y_t - G_{ij} \dots\dots\dots(2.48a)$$

を考え、これに最小2乗法を適用して、その2乗平均 ε^2_{ij} が最小となるように決められる方法が取られている。

$$\varepsilon^2_{ij} = \frac{1}{N-M} \sum_{t=1}^{N-M} (\varepsilon_{t,ij})^2 \rightarrow \min \dots\dots\dots(2.48b)$$

具体的には、パラメータに関する c 個の方程式

$$\frac{\partial \varepsilon^2_{ij}}{\partial c_{k,ij}} = 0 \quad (k=1, \dots, c) \dots\dots\dots(2.48c)$$

から、 c 個の最小2乗解を解析的に導出しておき、これらに $(N-M)$ 個のデータを適用して統計的に推定する。基礎関数に対応した最小2乗解は第3章で述べる。

2.6.2 基礎関数の選択と中間データ

ここでは、各層においてフィッティングされた基礎関数を、如何に取捨選択して最適な階層モデルを作るための候補を残すかについて述べる。

(1) 評価関数

いま、入出力データ・セット D_t (2.41式)に対してGMDHの階層モデルのフィッティングが開始されたとしよう。第1層において入力データの組合せ数が N_p 個あり、基礎関数 $G_{ij}^{(1)}$ も N_p 個フィッティングされたとする。

$$\{G_{ij}; c_{1,ij}, \dots, c_{c,ij}\}^{(1)} \quad ((i,j)=1, \dots, N_p) \dots\dots\dots(2.49)$$

フィッティングされたモデルの良否を判定するためには、一般にそのモデルの予測誤差に基づく評価関数が用いられる。GMDHの基礎関数の判定に用いられる評価関数の1つには、予測誤差(2.48a式)の2乗平均(2.48b式)を利用したものがある。

$$E_{i,j}(G_{i,j}) = \varepsilon^2_{i,j} \quad \dots\dots\dots(2.50)$$

この場合、 ε^2_p の計算に必要な各時点の予測誤差データ

$$\{ \varepsilon_{t,i,j} \} (t=1, \dots, N-M, (i,j)=1, \dots, N_p) \quad \dots\dots\dots(2.51)$$

は、個々の基礎関数(2.49式)を予測誤差の式(2.48a)に代入した式に対して、入出力データ・セット D_t を適用して計算される。この計算に用いる入出力データ・セット D_t は、統計的な見地(モデルのロバスト性等)から、フィッティングで用いたものとは独立なものが望ましい。

このため、GMDHでは(2.50)式を評価関数として用いた場合、始めに与えられる観測データ・セット D_t を、統計的時間平均の個数($N-M$)に関して、予めトレーニング用(フィッティング用)とチェック用(評価関数計算用)の2つに分けて用いる方法を取っている。

一方、別の評価関数として、時系列モデル一般に用いられている赤池の情報量基準AICを用いた例も見られる。AICは予測誤差の2乗平均値とパラメータ数に依存する関数であるが、パラメータ数を一定にした場合には、2乗平均項だけの関数となる。したがって、基礎関数の形を1種に固定してフィッティングした場合には、同一層内の基礎関数の評価は、単なる2乗平均(2.50式)による判定と一致する。

(2) データの分割法

基礎関数の評価関数に単純な予測誤差の2乗平均(2.50式)を用いた場合、通常、第1層のフィッティング前に、($N-M$)個のデータ・セット(2.41式)をトレーニング用とチェック用の2つに分割する操作が行われる。

(2.41)式の入出力データ・セット D_t は、時刻 t とそれに関係する全ての時間遅れ点を1セットとして、全部で($N-M$)個のデータ・セットから成立っている。モデルのパラメータは時刻に依存しないことを前提としているため、各時刻のデータ・セットは互に統計的に独立である。このため、統計的には分割はどのような方法を取っても良い。

しかし、現実には取り扱うデータの個数が有限、特にGMDHでは小数個の場合が多いため、分割の方法の違いでその結果が異なり、後の階層構造の決定に大きな影響を与えることが経験的に知られている。そこでGMDHでは、これを1つの発見的規範として、システム・モデリングの目的に合せた種々の方法が提唱されている訳である。

ここでは、一般に使用されている、処理前に入出力観測データの分散値を基準とした分割方法について述べる。この方法は、フィッティングの際にはデータの絶対値の大きいものを使った方が精度良く計算できるだろうし、また予測誤差に基づく評価関数の計算には小さいものの方が、精度が悪くなるためにその評価が容易になるだろう、という直感的、経験的な考え方に

基づいている。

データの分散値は、各時刻 t のデータ・セット(2.41式)に対して次の様に定義される。

$$\sigma^2_t = \frac{1}{N-M} \left((y_t - \mu_y)^2 + \sum_{i=1}^M (x_{t-i} - \mu_x)^2 \right) \quad \dots\dots\dots(2.52a)$$

ここで、

$$\mu_y = \frac{1}{N} \sum_{t=1}^N y_t, \quad \mu_x = \frac{1}{N} \sum_{t=1}^N x_t \quad \dots\dots\dots(2.52b)$$

この計算を、全てのデータ・セット($t=M+1, \dots, N$)について行い、各データ・セットに対応した($N-M$)個の分散値を得る。

$$\{ \sigma^2_t \} \quad (t=1, \dots, N-M) \quad \dots\dots\dots(2.53)$$

データをトレーニング用とチェックング用に分割する比率($N_t:N_c$)は、分割基準の採用と同様に任意である。ここでは、簡単に1:1(半分づつの大きさ)に分割してみよう。先に得られた分散値を相互に比較し、大きい順($q=1, \dots$)にデータ・セット D_t を並べ換える。ここで、 q 番目に対応する時刻を t_q と表せば、並べ換えたデータ・セットは、

$$\{ D_{t_q} \} \quad (q=1, \dots, N-M) \quad \dots\dots\dots(2.54)$$

と表され、これを前半と後半に分割すれば良い。

$$\begin{aligned} (t_q=1, \dots, N-M/2) & \quad ; \text{ トレーニング用} \\ (t_q=(N-M)/2+1, \dots, N-M) & \quad ; \text{ チェックング用} \end{aligned} \quad \dots\dots\dots(2.55)$$

以上のデータ分割操作は、分割比($N_t:N_c$)によって決る境界値を σ^2_{lim} と置けば、これをしきい値とした判定操作と等価である。

$$\begin{aligned} \{ \sigma^2_t \} \geq \theta_2 & \quad ; \text{ トレーニング用} \\ & < \theta_2 & \quad ; \text{ チェックング用} \end{aligned} \quad \dots\dots\dots(2.56)$$

ここで、 $\theta_2 = \sigma^2_{lim}$

(3) 基礎関数の取捨選択と中間データの計算

各層毎に行われるフィッティングされた基礎関数の取捨選択は、先に述べた様に、GMDHでは最大階層数が未知である事に起因するモデルの発散を回避し、上位層での計算処理を軽減する

とともに、得られる階層モデルを最適なものに近づける事が目的である。

いま、 k 番目の層において、入力データ $z_m^{(k)}$ が M 個あり、ここから入力データの組合せ数 N_p 個の全てに対して、トレーニング・データから N_p 個の基礎関数 $G_{ij}^{(k)}$ がフィッティングされ、またチェック・データからその予測出力 $u_{ij}^{(k)}$ と予測誤差 $\varepsilon_{t,ij}^{(k)}$ が計算されたとしよう。

$$\begin{aligned} \{ z_1, \dots, z_M, y_t \}^{(k)} &\rightarrow \{ G_{ij} \}^{(k)} && \dots\dots\dots (2.57) \\ (t=\text{トレーニング用}) &&& ((i,j)=1, \dots, N_p) \\ \{ z_1, \dots, z_M, y_t \}^{(k)}, \{ G_{ij} \}^{(k)} &\rightarrow \{ u_{ij} \}^{(k)}, \{ \varepsilon_{t,ij} \}^{(k)} \\ (t=\text{チェック用}) &&& ((i,j)=1, \dots, N_p) \end{aligned}$$

これらの基礎関数の取捨選択は、評価関数 $Ev_{ij}(G_{ij}^{(k)})$ の値を基準に行われる。問題は階層モデルの構築に用いる候補を幾つ残すかである。組合せの数 N_p に対して、残すべき候補の数を N_s 、また完全表現に組み込まれる最終的な数を N_d とすると、取捨選択の数を決める N_s は、明らかに

$$N_p^{(k)} \geq N_s^{(k)} \geq N_d^{(k)} \quad \dots\dots\dots (2.58)$$

を満足する必要がある。Fig.2.15は、層数5の階層モデルを例に取って、フィッティング過程における基礎関数の選択を図示したものである。図中A、B、Cが(2.58)式の関係を表している。

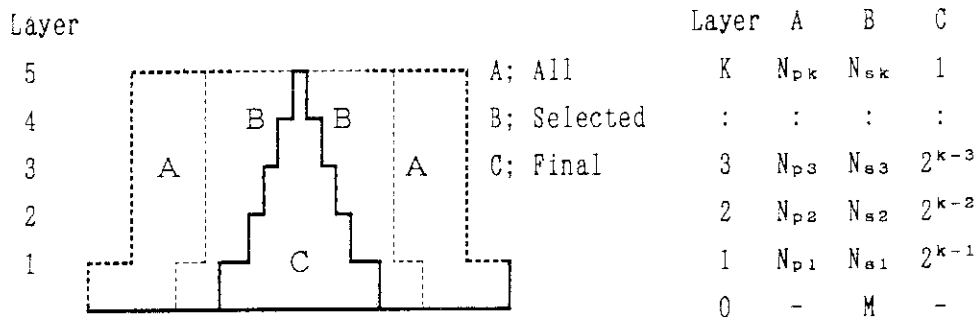


Fig.2.15 Fitting process in GMDH.

N_p は、第1層では入力データ、第2層では中間データの2乗(重複のない組合せ)に依存し、 N_s は層数の増加とともに2のべき乗で減少していく。したがって N_s は、大きすぎれば計算量が増大しモデルも収束できなくなるし、また小さすぎれば最適モデルとなる前に階層数に制限を受けてしまう。しかし、最適な決定方法は知られていないため、この決定も半経験的な発見的自己規範の範疇となる。これは第3章で述べる。

一旦選択数 N_s が決れば、取捨選択の作業は容易である。いま、評価関数の計算が為され、フィッティングされた N_p 個の基礎関数が適合の良さの順($p=1, \dots$)に並べ換えられたとする。この

時、基礎関数の取捨選択は評価関数の良い順 p で N_s 個選択される。

$$\{ G_p \}^{(k)} \quad (p=1, \dots, N_s) \quad \dots\dots\dots(2.59)$$

以上の基礎関数の取捨選択操作は、候補の比率(N_s/N_p)によって決る評価関数の値の境界値を EV_{lim} と置けば、これをしきい値とした判定操作と等価である。

$$\begin{aligned} \{ EV_p(G_p) \}^{(k)} \leq \theta_3^{(k)} & \quad ; \text{ 選択} & \dots\dots\dots(2.60) \\ & > \theta_3^{(k)} & \quad ; \text{ 棄却} \end{aligned}$$

ここで、 $\theta_3^{(k)} = EV_{lim}^{(k)}$

次層の入力データとなる中間データは、この選択された基礎関数 $\{G_p\}$ の出力値 $\{u_p\}$ である。 u_p は、(2.57)式の評価関数の計算の際に、チェックング・データに関しては既に計算済みであり、基礎関数の選択と同時に残すべきデータは決定される。一方、トレーニング・データに関しては、選択された基礎関数を用いて出力値を計算し、これを次層のデータとする。

$$\begin{aligned} \{ z_1, \dots, z_M, y_t \}^{(k)}, \{ G_p \}^{(k)} & \rightarrow \{ u_p \}^{(k)} & \dots\dots\dots(2.61) \\ (t = \text{トレーニング用} + \text{チェックング用}) & \quad (p=1, \dots, N_s) \end{aligned}$$

2.6.3 最終層の決定

最終層の層数 K の決定は、階層モデルの予測誤差に基づいて行われる。したがって、基本的には基礎関数の選択と同様に、基礎関数の予測誤差を調べれば良い。評価関数は、基礎関数の選択に使用したものと同じものが使用でき、例えば予測誤差の2乗平均であれば(2.50)式が使える。

最終層の決定は、同一層の中で良いものを複数個選択するという基礎関数の選択操作とは異なり、これまでにフィッティングした基礎関数全部を対象に唯1つの基礎関数を選択することある。基礎関数の取捨選択操作の段階で、同一層内の基礎関数の順序付けは完了しており、最良の基礎関数は既に分かっており、しかもフィッティングは階層数を増しながら逐次的に行われるから、最終層の決定は現在層(k)と前層($k-1$)の最良の基礎関数の評価関数を比較して、極小点を探せば良いことになる。

前層の最良基礎関数の評価関数 $EV_{min}^{(k-1)}$ をしきい値とすれば、現在層の値 $EV_{min}^{(k)}$ がそれよりも大きい時、前層の層数($k-1$)を最終層とする。

$$EV_{min}^{(k)} \geq \theta_4^{(k)} \quad \dots\dots\dots(2.62)$$

ここで、 $\theta_4^{(k)} = EV_{min}^{(k-1)}$

2.7 発見的規範とパラメータ

2.7.1 発見的規範

GMDH型の階層モデルのフィッティングでは、先に述べたように、ある与えられた形の個々の基礎関数に対しては最小2乗解や最尤解を得ることができるが、それらを結合した階層モデルの理論的最適解が得られないため、それらを如何に結合して最適にするかが問題となる。

多数のフィッティングされた基礎関数から候補を複数個選択し、それを上位層に繋いで徐々に希望する階層構造(カスケード)を作っていくというGMDH特有の操作は、自然界における植物の種子を人為的に選択して、世代交代を繰返しながら次第に希望の品種を作り上げるという、品種改良に似ている。品種改良では、希望の品種を作り上げることを目的とした種子の交配と選択に関する豊富な経験と知識が蓄積されており、数々の成功例が知られている。

A. Ivakhnenkoは、品種改良と階層モデルのフィッティング過程との類似性に着目して、品種改良における経験と知識を整理し、これを階層モデルフィッティングに役立てる目的で、次の仮説を立てた。

品種改良の仮説

- 1) 各世代の種播き用種子の選択に際しては、最適な一定の割合が存在し、その増減は品種改良の悪化、遅れをもたらす。
- 2) 品種改良は少なくとも数世代以上かかる。
- 3) 長期の品種改良は退化をもたらす。
- 4) 問題が複雑であればあるほど、淘汰には多くの世代を必要とする。

この仮説を階層モデルのフィッティングにあてはめると、次の様になる。

- 1) データや部分表現の取舍選択には、最適な一定の割合があり、その増減は得られるモデルの悪化、最適モデルへの収束の遅れをもたらす。
- 2) 階層モデルのフィッティングには、少なくとも数層以上行う必要がある。
- 3) 層数には最適値が存在し、これを越えると表現すべき情報に比較してモデル構造の複雑さが増すばかりで、誤差も増加していく(表現の退化)。
- 4) 対象となるシステムが複雑であればある程、大きな層数が必要となる。

この中で1)は、始めに与えるデータの数や逐次的フィッティングの過程で基礎関数の候補を選択する割合の決定に重要な示唆を与え、また2)、4)は基礎関数の形の決定に対して重要な示唆を与える。さらに、3)は最適層数の決定に重要な示唆を与える。

すなわち、逐次的な階層モデルのフィッティングにおいては、各ステップにおいてある一定の候補数を確保する必要があること、基礎関数の形により必要とする層数が増減するから問題に適した形を選ぶ必要があること、層数は必要最小限に抑える必要がある等である。

階層モデルのフィッティング過程において、これらの指針を与えるために導入されたのが、発見的規範である。発見的規範は、階層モデルの発散(計算量の増大)を抑制し、同定されるモ

デルを最適解に近づける機能を果たす。

GMDHでは、以下の5種の発見的規範が用いられる。(記号Hは発見的規範を表す)。

- (H-1) データの予備的選択 (遅れ時間点の棄却)
- (H-2) データの分割 (トレーニング用とチェック用への分割)
- (H-3) 基礎関数の形の決定
- (H-4) 基礎関数の候補の選択 (次層の入力データの組合せ数の制限)
- (H-5) 層数(最終層)の決定 (計算の停止)

各発見的規範は、これまで種々のものが提案されてきたが、以下では主として本論文に採用されたものについて簡単に説明する。

(1) 基礎関数の結合に関する発見的規範

発見的規範のうち(H-1)(H-4)(H-5)の3つは基礎関数の結合に関する規範であるため、ここではこれらをまとめて説明する。

GMDHの完全表現は、最終層に唯1個の部分表現を置き、これを頂点とするピラミッド型に部分表現を並べた形で表現される。逆に、完全表現を得るためには、系統的な結合経路の選択(データまたは部分表現の選択)が必要である。この結合経路の選択は次の3つの段階がある。

- 1) 第1層の入力データの取捨選択
- 2) 各層の部分表現(出力される中間データ)の取捨選択
- 3) 最終層の決定(完全表現に用いる部分表現の選択)

第1層の入力データの取捨選択に関しては、入力データ(変数、時間遅れ点の数)の数によって、組合せの総数が決り、フィッティングすべき部分表現の数を決るため、この数を取捨選択することは、入力データと部分表現の経路を遮断することに相当する。

各層の部分表現の取捨選択に関しては、部分表現の数がその出力である中間データの数を決めることから、この数を取捨選択することは、その出力経路と次層の部分表現の入力を遮断することに相当する。

また、層数(または最終層)の決定は、最終層内で唯一の部分表現を選択することであり、これによって、それより下位の層において残された部分表現も、最終層の部分表現と結合しているものだけが選択され、全結合経路が確定するということを意味する。

上の3つの問題に対して、種々の発見的規範があるが、本論文では以下のものを採用する。

(H-1) 第1層の入力データの予備的選択;

相関関数の値の大小関係を判定の基準とする。

- 相関の値が大きい順に(時間遅れ点、変数から成る組合せ総数:選択数)の割合で残す。

(H-4) 層内の部分表現の選択;

部分表現の予測誤差を判定の基準とする。

- 中間層; 予測誤差が小さい順に(組合せ総数: 選択数 N_s)の割合で残す。
- 最終層; 予測誤差が最小となるもの1個を選択する。

(H-5) 層数(最終層)の決定;

部分表現の予測誤差の各層毎の最小値を並べた曲線において最初に極小点が見られる層を最終層とする。

(2) 基礎関数の決定に関する規範

GMDHの骨子となる基礎関数を何にするかは種々のものが提案されている。

通常の階層モデルのフィッティングにおいては、計算開始前にある1つの基礎関数形を与え、これをフィッティング過程の最初から最後まで一貫して使用する。この場合、始めに与えた基礎関数形が適切かどうかは結果を見て評価する以外にはなく、結果が思わしくない時には基礎関数形を試行錯誤的に取り替えながらフィッティング計算を繰り返す必要がある。

一方、基礎関数形を可変としたものも提案されている^{23), 25)}。これはKolmogorov-Gaborの多項式を基本として、多項式の項数を徐々に増やしていく方法であり、その評価関数としてはAICが用いられている。

本論文では、簡単化のために基礎関数形はフィッティング過程においては固定とした。

(3) データの分割

データの分割に関する規範は、階層モデルのフィッティング過程において、基礎関数のパラメータ推定用(トレーニング用)とその評価関数の計算用(チェック用)にデータを分割する際の判定に用いられるもので、この分割操作によってもフィッティングの結果が左右されるという経験則から産み出されたものである。データの分割に関する規範も問題に応じて種々のものが提案されている。しかし、パラメータ推定と評価関数の計算に独立なデータを要しない場合(例えば評価関数にAICを用いたもの等)には、この規範を使う必要はない。

本論文では、簡単化のために、Ivakhnenkoによって初期に提案された次のものを採用する。

(H-2) データ分割は、観測データの分散値の大小関係を判定の基準とする。

(S/N比の意味で)分散値の大きいグループをトレーニング・データに、小さいグループをチェックング・データに分割する。

2.7.2 経験的パラメータ

本節の最後として、GMDHに用いられる経験的パラメータについて述べる。発見的規範は本来、処理の過程を通じて見出す規範であるが、(H-1)~(H-4)の発見的規範には、これを補助する経験的パラメータの入力が必要である。

規範	入力	内容
----	----	----

- ; (I-1) 解析すべき時間遅れ点の範囲(これは、相関の補助的なもの)
- (H-1) ; (I-2) 相関の値(または、棄却すべき割合点(数全点数:強い点数))
- (H-2) ; (I-3) トレーニング・データとチェックング・データの割合(Lt:Lc)
- (H-3) ; (I-4) 基礎関数形式と次数p
- (H-4) ; (I-5) 部分表現の選択割合(組合せ総数:選択数Ns)

(I-1)は、どの手法でも用いられているパラメータで、この値を大きく設定すると、通常はモデルに反映する周波数レンジが大きくなるが、反面、計算効率が悪くなる。これは、相関関数の計算のために用いる補助的なパラメータである。

(I-2)は、遅れ時間点の取捨選択のしきい値として用いる相関関数の値で、通常、統計的な推定精度から考えて、相関があるらしいと見られる値を設定しておく。

(I-3)は、取り扱うデータの性質とモデリングの目的によって異なる。(H-2)で用いている分散値による判別では、割合をまず(1:1)に設定して完全表現を同定した後、これを変えて何度か同定を繰返して、層数が少なくなる様に試行錯誤的に分割数を探す。

(I-4)は、基礎関数として何を与えるか、というものである。Kolmogorov-Gaborの多項式の場合は、問題によって、最大次数を設定しておく。

(I-5)は、中間データの取捨選択の割合で、フィッティングの前に用意したデータの数(変数の数、時間遅れ点の範囲)やシステムの複雑さ(線形、非線形)等のシステムに関する事前情報を基に、最大層数がどのくらいになるかを見積り、逆算して大まかに設定する。これも、試行錯誤的なパラメータである。

3. 時系列信号に対するアルゴリズム

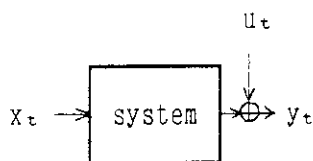
本章では、時系列データに対するGMDHのフィッティング・アルゴリズム^{4),5)}について述べる。アルゴリズムは、大別して下記の7つの処理ステップで構成されている。

- (1) データ行列Zの生成
- (2) データ行列Zの行に関する前処理 ; 相関関数に基づく入力データ点
(時間遅れ点)の取捨選択
- (3) データ行列Zまたは中間データ行 ; 分散値に基づく行列の分割
列Z'の列に関する処理 (トレーニング・データ行列Z_Tおよびチェッキング・データ行列Z_Cの生成)
- (4) トレーニング・データ行列Z_Tを ; 入力データ点の全組合せに対応する全基礎関数
用いた処理 (部分表現)のフィッティング
- (5) チェッキング・データ行列Z_Cを ; 2乗平均予測誤差に基づく部分表現の候補の取捨
用いた処理 選択および最終層の判定
- (6) 中間データ行列Z'の生成と繰返 ; 上位層の計算に必要な入力データの準備と第2層
し計算(4~6) 以降の計算
- (7) 完全表現の構築 ; 最終的な各層の部分表現の決定

そこで、以下の3.2~3.8節では、1入力1出力型の階層型モデルのフィッティングを例として、7つの処理ステップを詳細に説明し、3.9節では、多入力1出力型、1変数および多変数フィードバック型の場合の補則説明を行う。

3.1 1入力1出力型階層モデルの同定の問題

ここではまず、入力を x_t 、出力を y_t 、付加雑音を u_t とする、Fig.3.1の離散型1入力1出力システムを考えよう。これは、前章の始めに考えた連続型のシステムFig.2.1に対応するもので、Fig.2.3と同じものである。ここでは、システムが非線形か否かは問わない。



いま、これらの入出力に関する観測データとして、時間間隔 Δt でサンプリングされたN個の時系列データ

Fig.3.1 Single-input/single-output system

$$\{ x_i, y_i ; i=1, \dots, N \} \dots \dots \dots (3.1)$$

が与えられているものとする。この時、これらの観測データに基づいて1入力1出力型モデルを同定するのがここでの問題である。具体的には、システムの最大遅れ時間点をMとするとき、時刻tの出力 y_t とそれに対応する過去M点の入力 $\{x_{t-1}, \dots, x_{t-M}\}$ の関係を記述するモデルf

$$y_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-M}) + \varepsilon_t \quad \dots \dots \dots (3.2)$$

を同定する問題となる。ここで ε_t は残差成分を表す。

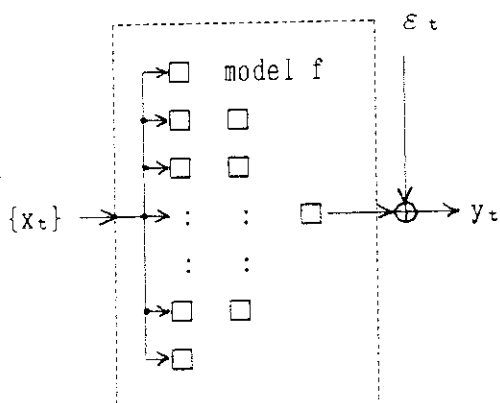


Fig.3.2 Single-input/single output GMDH model.

Fig.3.1に示した1入力1出力システムに対するGMDHに基づく同定では、Fig.3.2に示す様なモデル階層型の内部構造を持ったモデルが仮定される。モデル上では、 ε_t は入力 $\{x_t\}$ から計算された出力 y_t の予測誤差を表す量であり、これは本来入力とは無相関な量である。

したがって、この同定が成功するか否かはFig.3.1における付加雑音 u_t の特性に依存する。一般に、 u_t が白色雑音であれば通常の最小2乗法に基づく同定手法が適用可能であると同様に、GMDHに基づく階層モデルの適用も可能となる。

逆に、 u_t が入力と相関を持つ場合には、同定結果は偏り誤差を持つようになる。この場合は、後

に述べる様に、ARモデルと同様に、出力 y_t を入力変数に取り込んだフィードバック・モデルを作る必要がある。

3.2 データ行列の生成

ここでは、階層モデルのフィッティング過程の準備として、(3.1)式で与えられたM個の観測データを、統計的に独立な複数のサンプルを作りデータ行列を生成する。

まず、必要な初期パラメータとして、最大遅れ時間点Mを定義しておく。

(I-1) 解析すべき時間遅れの範囲を決定する最大遅れ時間 τ_{max} 、離散表現では最大遅れ点M ($M = \tau_{max} / \Delta t$)を、信号の性質に基づいて決めておく。

次に、フィッティング計算に便利な様に、時系列データを整理してデータ行列を作る。

(3.2)式に示した様に、時刻tにおける出力 y_t が過去M点の入力値 $\{x_{t-1}, \dots, x_{t-M}\}$ を用いたモデルによって記述できるとするならば、その逆に、それら入出力の観測値からモデルを同定できるはずである。具体的には、モデルがp個のパラメータで記述される場合には、確定的なシステムではp組の独立な観測データがあればシステムを同定できる。すなわち、同定に必要な基本的な観測データの組

$$\{ y_t, x_{t-1}, \dots, x_{t-M} \} \dots\dots\dots(3.3)$$

がp組あれば良い。

本アルゴリズムの対象である確率過程の場合には、pよりも十分大きいL組の独立な観測データが与えられていれば、原理的には、L元連立方程式を統計的に解くことによって同定が可能となる。統計的な性質は時刻に不変であるとすれば、(3.1)式に与えられたN個の連続した観測データから得られる独立な観測データの組Lは、次の様に定義できる。

$$L = N-M \dots\dots\dots(3.4)$$

ただし、 $t = 1+M, \dots, L+M$

そこで、次に定義する出力ベクトルY、入力データ行列Zを作成する。

$$Y = (y_{L+M} \quad y_{L+M-1} \quad y_{L+M-2} \quad \dots\dots\dots y_{M+1})^t \dots\dots\dots(3.5)$$

$$Z = \begin{pmatrix} x_{L+M} & x_{L+M-1} & x_{L+M-2} & \dots\dots\dots x_{M+1} \\ : & : & : & : \\ x_L & x_{L-1} & x_{L-2} & \dots\dots\dots x_1 \end{pmatrix} \dots\dots\dots(3.6)$$

行列Zの各列は、独立な観測データの組を与える。

3.3 データ行列Zの行に関する前処理

GMDHに基づく階層モデルの同定では、各層毎に候補となる基礎関数を探すために、与えられた入力データ点の全ての組合せに対応した基礎関数のフィッティング計算を行う必要がある。そこで、事前に、入出力変数間の相関の度合いに基づきフィッティングに用いるデータ点を減らしておく事ができれば、後の基礎関数のフィッティングの作業量をかなり軽減できる。

ここでは、GMDHに基づく階層モデルのフィッティングに必要な前処理として、データ行列Zの行(時間遅れ点)に関する取捨選択操作を行う。

具体的には、まず出力に対する入力変数の各遅れ時間点の相関を計算しこれに基づいて取捨選択を行い、弱い相関を持つ時間遅れ点(行データ)を棄却し行列Zを縮小する。

3.3.1 データの相関関数の計算

はじめに、出力点に対する各遅れ時間点の相関値を計算する。(相互)相関関数は、通常

$$\begin{aligned} R_{yx_m} &= E(x_{t-m} y_t) \\ &= E(x_t y_{t+m}) \quad (m=0, \dots, M) \dots\dots\dots(3.7) \end{aligned}$$

あるいは

$$R_{yx_m} = \lim_{N \rightarrow \infty} \frac{1}{N-m} \sum_{t=1}^{N-m} x_t y_{t+m} \quad \dots\dots\dots(3.8)$$

(m=0, ..., M)

と定義され、サンプル数Nが有限な実際の推定計算では、

$$R_{yx_m} = \frac{1}{N-m} \sum_{t=1}^{N-m} x_t y_{t+m} \quad \dots\dots\dots(3.9)$$

(m=0, ..., M)

が良く用いられる。この式による推定値は、偏り誤差の点では不偏推定量となっているが、分散誤差の点では、統計的平均の長さ(N-m)は時間遅れmの増加とともに減少していくため、全サンプル数Nが有限一定で小さい場合には、分散誤差はmの増加とともに大きくなる。

各時間遅れ点の分散誤差を一律に揃えようとする場合には、偏り誤差を多少犠牲にして、統計的平均の長さをL=N-Mと固定してしまう方法がある。これが、次の定義である。

$$R_{yx_m} = \frac{1}{L} \sum_{t=1}^L x_t y_{t+m} \quad \dots\dots\dots(3.10)$$

(m=0, ..., M)

上式は、相関関数ベクトルR、出力データ・ベクトルY、(M+1 x L)の入力データ行列Zを用いた行列表現を用いて、次の様に見えることができる。

$$R = \frac{1}{L} Z Y \quad \dots\dots\dots(3.11)$$

ここで、

$$R = (R_{yx_0} \ R_{yx_1} \ R_{yx_2} \ \dots\dots\dots R_{yx_M})^T \quad \dots\dots\dots(3.12)$$

行列Zの要素をz_{mn}とし、ベクトルYの要素をy_mとすると、(3.10)式は

$$R_{yx_m} = \frac{1}{L} \sum_{n=1}^L z_{mn} y_m \quad \dots\dots\dots(3.13)$$

で計算されることになる。

次に、求めた相関関数の相関の度合いを、観測データの計測単位に無関係に比較するために、これを規格化し無次元化する。相関関数の規格化は、通常、次式によって行われる。

$$R'_{yx_m} = \frac{E(x_t y_{t+m})}{\sqrt{E(x_t x_t) E(y_{t+m} y_{t+m})}}$$

$$= \frac{R_{yx_m}}{\sqrt{R_{xx_0} R_{yy_0}}} \quad \dots\dots\dots(3.14)$$

上式分母の2つの項 R_{yy_0} 、 R_{xx_0} は時間遅れ0点の自己相関関数、すなわちデータ x, y の分散値である。これらの推定には、データ長 N が比較的大きい場合は(3.9)式が利用できる。また、データ長 N が小さい場合には(3.8)式が利用できる。ただし、(3.10)式の統計平均の区間は $[1, N-M]$ で定義されているため、 R_{xx_0} の推定のときは良いが、 R_{yy_0} の場合は(3.14)式を見ると分かる様に $[1+m, N-M+m]$ の区間の平均を行う必要がある。

以上を整理し、行列 Z 、ベクトル Y の要素を用いた計算式に書き直すと以下の様になる。

$$R'_{yx_m} = \frac{R_{yx_m}}{\sqrt{R_{xx_0} R'_{yy_0}}} \dots\dots\dots(3.15)$$

ここで、

$$R'_{yy_0} = \frac{1}{L} \sum_{t=1}^L y_{t+m} y_{t+m} \dots\dots\dots(3.16a)$$

$$R_{xx_0} = \frac{1}{L} \sum_{n=1}^L z_{m,n} z_{m,n} \dots\dots\dots(3.16b)$$

($m=0, \dots, M$)

3.3.2 弱い相関関係にある時間遅れ点の棄却

ここでは、前節の演算で得られた規格化相関関数 R'_m を用いて、遅れ時間区間 $[1, M]$ の範囲から、弱い相関関係にある遅れ時間点のデータを棄却する処理を行う。

(I-2) 棄却すべき相関の度合い R_{max} を解析前の事前情報や経験から決定し、これを棄却判定のしきい値 θ_r とする。

$$\theta_r \leftarrow R_{max}$$

$$0 < R_{max} < 1 \dots\dots\dots(3.17)$$

このしきい値により、各遅れ時間 m における規格化相関関数 R'_m の判別を行い、

$$| R'_m | \geq \theta_r \dots\dots\dots(3.18)$$

($m=1, \dots, M$)

を満たす時間遅れ点だけを取り出す。遅れ時間 $m=0$ は、現在の時刻点を表し当然相関値は1であるから、この検索から除外する。(3.5)、(3.6)、(3.11)式から明らかな様に、遅れ時間 m の相関関数を棄却することは、出力ベクトル Y の m 番目の要素、ならびにデータ行列 Z の m 番目の行データを棄却する事に等しい。

検索の結果、 $(M-M')$ 個の遅れ時間点が棄却され、次の M' 個の遅れ時間点が残ったとしよう。

$$\{ 2 \dots\dots\dots M-1 \} \dots\dots\dots(3.19a)$$

これらの点を新たに次の記号で表すものとしよう。

$$\{ r_1 r_2 \dots\dots\dots r_{M'} \} \dots\dots\dots(3.19b)$$

この棄却操作により、行列Zは下線で示す行だけが残される。

$$Z = \left(\begin{array}{cccccc} X_{L+M} & X_{L+M-1} & X_{L+M-2} & \dots\dots\dots & X_{M+1} \\ \underline{X_{L+M-1}} & \underline{X_{L+M-2}} & \underline{X_{L+M-3}} & \dots\dots\dots & \underline{X_M} \\ : & : & : & & : \\ \underline{X_{L+1}} & \underline{X_L} & \underline{X_{L-1}} & \dots\dots\dots & \underline{X_2} \\ X_L & X_{L-1} & X_{L-2} & \dots\dots\dots & X_1 \end{array} \right) \begin{array}{l} r_1 \\ : \\ r_{M'} \end{array} \dots\dots\dots(3.20a)$$

棄却の結果、行列Zは行について縮小され、(M' x L)行列となる。これを行列Z'と置くと、この行列の要素は、(3.19b)式の時間遅れ点の記号を用いて次の様に表される。

$$Z' = \left(\begin{array}{cccc} X_{L-r_1} & X_{L-1-r_1} & \dots\dots\dots & X_{1-r_1} \\ : & : & & : \\ X_{L-r_{M'}} & X_{L-1-r_{M'}} & \dots\dots\dots & X_{1-r_{M'}} \end{array} \right) \dots\dots\dots(3.20b)$$

となる。元の行列Zの要素をz(i,j)と表わせば、上式は次の様になる。

$$Z' = \left(\begin{array}{cccc} z(r_1,L) & z(r_1,L-1) & \dots\dots\dots & z(r_1,1) \\ : & : & & : \\ z(r_{M'},L) & z(r_{M'},L-1) & \dots\dots\dots & z(r_{M'},1) \end{array} \right) \dots\dots\dots(3.20c)$$

3.4 データ行列Z'または中間データ行列Z'の列に関する前処理

ここでは、前節で作られたデータ行列Z'、または後節で作られる中間データ行列を2つの行列に分割する操作を行う。2つの行列とは、トレーニング・データ行列Z_Tとチェックング・データ行列Z_Cのことであり、前者は、部分表現の候補となる基礎関数のフィッティング用に、また後者は、得られたフィッティング関数に対する評価関数の計算用に用いられるデータ行列である。

この前処理は、本アルゴリズムに使用した評価関数の計算がフィッティングに用いるデータと独立なデータを必要とするために事前に行っておく処理であり、AICの様な特別に独立なデータを必要としない評価関数が適用できる場合には、この前処理は必要ない。

さて、ここではデータ行列を上記の2つのデータ行列に分割する方法を考えよう。この分割

は、観測されたサンプル・データの集合をその目的に応じて2つに分ける操作である。

3.2節のデータ行列の生成のところで述べた様に、データ行列Zの行は同定で考慮される全ての遅れ時間点が並び、また列は独立な観測データの組が並ぶ様に作られている。この関係は、相関関数に基づく前処理が施された行列Z'も同じであり、また後に述べる中間データ行列Z'も同様に作られている。したがって、データの集合の分割操作は行列上では列に関する操作を行うことに相当する。

データ行列Z'の列に関する操作の性質、ならびにトレーニング・データ行列Z_Tとチェック・データ行列Z_Cへの分割に際して課せられる条件は、次の通り。

- (1) 行列Z'の列は互に独立なデータの組であるため、列の入替え操作が可能。
したがって、分割に際して列の選択は任意である。
- (2) パラメータ推定あるいはその評価関数の計算に必要な列の数は、パラメータ数p以上あれば良い。したがって、行列Z_Tの列数L_t、Z_Cの列の数L_cの決定は、次の範囲内で任意である。

$$\begin{aligned} L_t &\geq p \\ L_c &\geq p \\ L_t + L_c &\leq L \end{aligned} \quad \dots\dots\dots(3.21)$$

ただし、統計的な演算精度を確保する意味から、両者ともできるだけ大きい値を取る必要がある。

一方、データ分割(クラス分け)に関する基準は、同定の成否を左右するためこれまでに種々のものが提案されているが、本アルゴリズムでは次のものを採用する。

(H-2) パラメータ決定(トレーニング・データ)には、分散値の意味でS/Nの良いデータ(列)を選択し、フィッティング後の判定(チェック・データ)には、残りの(S/Nの悪い)データを選択する。分割比(L_t:L_c)は、層の数が最小となる様に、試行錯誤的に決定する。

そこでまず分散値を計算する。データ行列Z'の第n列のデータの分散σ²_nは、(3.20c)式の要素の表現を用いて、次の様に定義される。

$$\sigma^2_n = \frac{1}{M'} \sum_{m=1}^{M'} z(r_{m,n})^2 \quad \dots\dots\dots(3.22)$$

(n=1, ..., L)

また、分散ベクトルΣを

$$\Sigma = (\sigma^2_1 \ \sigma^2_2 \ \dots\dots \ \sigma^2_L)^t \quad \dots\dots\dots(3.23)$$

と定義し、データ行列Z'の第n列ベクトルをZ'(n)と記すと、これは次式で計算できる。

$$\Sigma = \frac{1}{M'} Z'(n) Z'(n)^t \dots\dots\dots(3.24)$$

次に、得られた各列の分散値の大小判定を行い、各列を2つのクラスに分類する。ここで、これに必要なパラメータLt、Lcを事前に与えておこう。

(I-3) トレーニング・データの数LtをLの大きさと解析前の事前情報や経験から決定し、また、チェックング・データの数LcをL-Ltとしておく。

分割パラメータLt、Lcは、分割の比率だけを与える。したがって、分割操作は、得られた各列の分散値 $\sigma^2_n(n=1, \dots, L)$ の大きい順に並べ、大きい方からLt個をトレーニング・データに、残りのLc個をチェックング・データに分類する操作となる。

この操作は、LtとLcによって決まる分散値のしきい値を見つけ出す事に他ならず、分割比による分類は、しきい値による判定と等価である。まず、LtとLcによって決まる分散の境界値 σ^2_{lim} を、分割判定のしきい値 θ_v とする。

$$\begin{aligned} \theta_v &\leftarrow \sigma^2_{lim} \\ \sigma^2_{min} &< \sigma^2_{lim} < \sigma^2_{max} \end{aligned} \dots\dots\dots(3.25)$$

このしきい値に基づいて各列の分散値 σ^2_n を判定し、行列Z'の列番号n(n=1, ..., L)を2つに分類する。行列Z TとZ Cの行番号をそれぞれt, cで記せば、判定操作は以下の様に表現できる。

$$\begin{aligned} \{ \sigma^2_n \} \geq \theta_v &\rightarrow \text{トレーニング・データ} \{ t(1) \ t(2) \ \dots \ t(L_t) \} \\ &< \theta_v &\rightarrow \text{チェックング・データ} \{ c(1) \ c(2) \ \dots \ c(L_c) \} \end{aligned} \dots\dots\dots(3.26)$$

以上の操作を(3.20a)式のデータ行列Zに対して施した場合、下式の様に、枠で囲んだ列データがトレーニング・データに、またそれ以外の列がチェックング・データに分類される。

$$Z = \begin{pmatrix} \boxed{X_{L+M}} & \boxed{X_{L+M-1}} & \boxed{X_{L+M-2}} & \dots & X_{M+1} \\ \boxed{X_{L+M-1}} & \boxed{X_{L+M-2}} & \boxed{X_{L+M-3}} & \dots & X_M \\ \vdots & \vdots & \vdots & & \vdots \\ \boxed{X_{L+1}} & \boxed{X_L} & \boxed{X_{L-1}} & \dots & X_2 \\ \boxed{X_L} & \boxed{X_{L-1}} & \boxed{X_{L-2}} & \dots & X_1 \end{pmatrix} \begin{matrix} r_1 \\ \vdots \\ r_{M'} \end{matrix} \dots\dots\dots(3.27)$$

t(Lt) c(Lc) t(Lt-1) c(1)

分割によって、下記に示した大きさ($M' \times L_t$)を持つ行列 $Z T$ と大きさ($M' \times L_c$)を持つ行列 $Z C$ が新しく生成される。

$$Z T = \begin{pmatrix} X_{t(L_t)-r_1} & X_{t(L_t-1)-r_1} & \dots & X_{t(1)-r_1} \\ : & : & & : \\ X_{t(L_t)-r_{M'}} & X_{t(L_t-1)-r_{M'}} & \dots & X_{t(1)-r_{M'}} \end{pmatrix} \dots\dots\dots(3.28)$$

$$Z C = \begin{pmatrix} X_{c(L_c)-r_1} & X_{c(L_c-1)-r_1} & \dots & X_{c(1)-r_1} \\ : & : & & : \\ X_{c(L_c)-r_{M'}} & X_{c(L_c-1)-r_{M'}} & \dots & X_{c(1)-r_{M'}} \end{pmatrix} \dots\dots\dots(3.29)$$

これら2つの行列は、行列 Z の要素 $z(i, j)$ を用いて表わせば次の様に書き直せる。

$$Z T = \begin{pmatrix} z(r_1, t(L_t)) & z(r_1, t(L_t-1)) & \dots & z(r_1, t(1)) \\ : & : & & : \\ z(r_{M'}, t(L_t)) & z(r_{M'}, t(L_t-1)) & \dots & z(r_{M'}, t(1)) \end{pmatrix} \dots\dots\dots(3.30a)$$

$$Z C = \begin{pmatrix} z(r_1, c(L_c)) & z(r_1, c(L_c-1)) & \dots & z(r_1, c(1)) \\ : & : & & : \\ z(r_{M'}, c(L_c)) & z(r_{M'}, c(L_c-1)) & \dots & z(r_{M'}, c(1)) \end{pmatrix} \dots\dots\dots(3.30b)$$

最後に、データ行列 Z' の分割にともない、(3.5)式に示されている出力ベクトル Y も、分割しておく必要がある。ベクトル Y は、(3.29)式と(3.30)式に対応して次の様に定義される。

$$Y T = (X_{t(L_t)} \ X_{t(L_t-1)} \ \dots \ X_{t(1)})^T \dots\dots\dots(3.31a)$$

$$Y C = (X_{c(L_c)} \ X_{c(L_c-1)} \ \dots \ X_{c(1)})^T \dots\dots\dots(3.31b)$$

3.5 トレーニング・データに対する基礎関数のフィッティング

ここでは、オリジナル・データまたは中間データから作られたトレーニング・データを用いて、各層毎に行われる基礎関数の候補のフィッティングについて説明する。

3.5.1 基礎関数

基礎関数の関数形として、本アルゴリズムでは2.5.2節に述べた幾つかの関数形を用いている。この関数形は種々のものが提案されており、これが同定結果の成否を左右することも知られているが、仮に異なった関数形の基礎関数を採用した場合には、原則として基礎関数のパラメータの解法の部分さえ書き換えれば、その他の部分はそのまま使用できる。

基礎関数の形式は、事前に次の様に決めておく。

(H-3) 基礎関数として、Kolmogorov-Gaborの多項式から導出された

(1) 2入力1次多項式 (2.38)式

(2) 2入力2次多項式 (2.34)式

のいずれか1つを採用する。

(I-4) 基礎関数の次数は、あらかじめ、1次または2次を選択しておく。

3.5.2 フィッティング

(2.38)式または(2.34)式の基礎関数をフィッティングする問題は、第1層においては2つの時刻 $t_1, t_2 (< t)$ における入力の観測値 x_{t_1}, x_{t_2} を用いて出力の観測値 y_t を近似する問題となる。

$$y_t \sim G(x_{t_1}, x_{t_2}) \quad \dots\dots\dots(3.32a)$$

また、第2層以降は2つの中間変数のデータ点 z_{n_1}, z_{n_2} を用いて、 y_t を近似する問題となる。

$$y_t \sim G(z_{n_1}, z_{n_2}) \quad \dots\dots\dots(3.32b)$$

整理すれば、各層における基礎関数の各候補のフィッティングは、3つのデータ点から p 個のパラメータ $\{c\}$ を決定する問題となる。

$$\begin{aligned} \{ y_t, x_{t_1}, x_{t_2} \} \text{ or } & \rightarrow \{ c_1, \dots, c_p \} \quad \dots\dots\dots(3.33) \\ \{ y_t, z_{n_1}, z_{n_2} \} & \end{aligned}$$

ただし ($t=1, \dots, L$)

各基礎関数のパラメータ数 p は、1次多項式(2.38式)の場合は3個、2次多項式(2.34式)の場合は6個である。

3.5.3 入力データ点の組合せの発生

(2.34)式および(2.38)式の基礎関数の2つの入力データ点 $\{x_{t_1}, x_{t_2}\}$ または $\{z_{n_1}, z_{n_2}\}$ は、データ行列 Z' の行番号 $\{r_1, \dots, r_M\}$ に対応する M' 個のデータ点から抜き出された任意の2点である。さらに、これらの基礎関数の2つの入力は対称にできており、2つの入力端のどちらにデータを入力しても結果は同じであるし、見掛け上2つの入力端点に同じデータが入力された場合の計算も、基礎関数内部では片側の入力データだけに基づいて行う様に考慮されている。

したがって、2つの入力データ点の組合せ (n_1, n_2) は、次の範囲で考えれば良い。

$$r_1 \leq n_1 < n_2 \leq r_M \quad \dots\dots\dots(3.34)$$

		n_2			
		r_2	r_3	r_M
r_1	n_1	○	○	○
r_2			○	○
⋮					⋮
r_M					○

Fig.3.3 All combinations of the input data for $G(z_{n1}, z_{n2})$.

上式に従って実際に可能な組合せ(n_1, n_2)を図示したものが、Fig.3.3である。これらの組合せに対応する全ての基礎関数の候補をフィッティングする必要がある。

この場合、入力データ点の組合せ(n_1, n_2)の総数 NPは

$$NP = \frac{(M'-1)M'}{2} \dots\dots\dots(3.35)$$

と与えられる。

3.5.4 基礎関数のパラメータの解法

ここでは、トレーニング・データ(行列 Z^T 、ベクトル Y^T)から基礎関数のパラメータを求める方法について述べる。

本アルゴリズムにおける基礎関数のパラメータ決定では、2乗平均の意味で予測誤差が最小となる様にパラメータを最小2乗法を適用して決定する方法を採用している。具体的には、2乗平均誤差を与える式に基礎関数を代入してパラメータに関する偏微分方程式を作り、その最小2乗解を与える式を導出し、これにトレーニング・データを適用してパラメータを推定する。

はじめに時刻 t における出力の観測値 y_t に対する予測誤差 ε_t を定義しよう。いま、時刻 t に対応する入力データ点の組合せが NP 個あり、これに対応する基礎関数を $\{ G_{qt}; q=1, \dots, NP \}$ と記すと、予測誤差もまた q 通り定義できる。これを ε_{qt} と記せば次の様に定義できる。

$$\varepsilon_{qt} = y_t - G_{qt} \dots\dots\dots(3.36)$$

($q=1, \dots, NP, t=t(1), \dots, t(L_t)$)

トレーニング・データ $Z^T (t=t(1), \dots, t(L_t))$ に対する予測誤差の2乗平均値 ε^2_q は、次式で定義される。

$$\varepsilon^2_q = \frac{1}{L_t} \sum_{n=1}^{L_t} (\varepsilon_{qt(n)})^2$$

$$= \frac{1}{L_t} \sum_{n=1}^{L_t} (y_{t(n)} - G_{qt(n)})^2 \dots\dots\dots(3.37)$$

($q=1, \dots, NP$)

ここで、基礎関数 G_q の各トレーニング・データに対する近似値 $G_{qt(n)}$ をベクトル G

$$G_q = (G_{qt(1)} \ G_{qt(2)} \ \dots \ G_{qt(L_t)})^t \quad \dots \dots \dots (3.38)$$

と記せば、(3.37)式の2乗平均 ε^2_q は、(3.31)式の出カベクトル Y^T とベクトル G を用いて、以下の様 ε^2_q に書き直すことができる。

$$\varepsilon^2_q = \frac{1}{L_t} (Y^T - G_q)^t (Y^T - G_q) \quad \dots \dots \dots (3.39)$$

($q=1, \dots, NP$)

これら2乗平均 ε^2_q を最小にする様に基礎関数の係数 $\{c_{q1}, \dots, c_{qp}\}$ を決定するのが、ここでの問題である。(3.37)式を最小にするには、全ての係数 $\{c_m; m=1, \dots, p\}$ に渡って、その編微分が0になれば良い。ここで、解法は入力データの組合せ q は依存しない一般的なものであるの ε^2 で、以下の説明では記号 q を省略する。

$$\frac{\partial \varepsilon^2}{\partial c_m} = 0 \quad \dots \dots \dots (3.40)$$

($m=0, \dots, p$)

以下では、基礎関数に2次の多項式(2.34)式を用いた場合の最小2乗解を導出してみよう。(2.34)式の基礎関数 G_{ij} は、出力点 y_t と2つの入力データ点 $\{x_{t-i}, x_{t-j}\}$ に対して定義されたものである。トレーニング・データの場合、2つの入力データ点の表現は $\{x_{t(n)-i}, x_{t(n)-j}\}$ と時刻の表現が複雑になるので、以下簡単に時刻 $t(n)$ を n 、入力を $\{x_i, x_j\}$ と記すことにする。また中間データの場合は、 x を z で読み換えることとする。この様にすると、(2.34)式は次の様に書ける。

$$G_n = G_{ij} = c_1 + c_2 x_i + c_3 x_j + c_4 x_i^2 + c_5 x_i x_j + c_6 x_j^2 \quad \dots \dots \dots (3.41)$$

2乗平均(3.37)式に、上の基礎関数を代入すると次の様に展開できる。

$$\varepsilon^2 = \frac{1}{L_t} \sum_{n=1}^{L_t} (y_n^2 - 2y_n G_n + G_n^2)$$

$$= \frac{1}{L_t} \sum_{n=1}^{L_t} [y_n^2 - 2y_n (c_1 + \dots + c_6 x_j^2) + (c_1 + \dots + c_6 x_j^2)^2] \quad \dots \dots \dots (3.42)$$

この式を(3.40)式に代入し6つの係数に対する変微分を計算し、6元連立方程式を導出する。
第1番目の係数 c_1 に対する編微分は、次の通り。

$$-\frac{1}{L_t} \sum_{n=1}^{L_t} 2y_n + \frac{1}{L_t} \sum_{n=1}^{L_t} 2(c_1 + \dots + c_6 x^2_j) = 0 \quad \dots\dots\dots(3.43a)$$

これを整理して

$$\frac{1}{L_t} \sum_{n=1}^{L_t} y_n = \frac{1}{L_t} \sum_{n=1}^{L_t} (c_1 + \dots + c_6 x^2_j) \quad \dots\dots\dots(3.43b)$$

統計的平均操作を便宜上 $E(\cdot)$ という記号で表せば、

$$E(y_n) = E(c_1) + E(c_2 x_i) + E(c_3 x_j) + E(c_4 x^2_i) + E(c_5 x_i x_j) + E(c_6 x^2_j) \quad \dots\dots\dots(3.43c)$$

となり、各係数 $\{c_m\}$ は、平均操作に関して不変であるから分離でき、次式が導出される。

$$E(y_n) = c_1 + c_2 E(x_i) + c_3 E(x_j) + c_4 E(x^2_i) + c_5 E(x_i x_j) + c_6 E(x^2_j) \quad \dots\dots\dots(3.43d)$$

以下、第2番目以降の係数についても順番に編微分して行き、導出された式を、行列を用いて表わすと次の行列式が得られる。

$$B = CA \quad \dots\dots\dots(3.44)$$

$$A = (c_1 \ c_2 \ c_3 \ c_4 \ c_5 \ c_6)^t \quad \dots\dots\dots(3.45)$$

$$B = (E(y_n) \ E(y_n x_i) \ E(y_n x_j) \ E(y_n x^2_i) \ E(y_n x_i x_j) \ E(y_n x^2_j))^t \quad \dots\dots\dots(3.46)$$

$$C = \begin{pmatrix} 1 & E(x_i) & E(x_j) & E(x^2_i) & E(x_i x_j) & E(x^2_j) \\ E(x^2_i) & E(x_i x_j) & E(x^3_i) & E(x^2_i x_j) & E(x_i x^2_j) \\ E(x^2_j) & E(x^2_i x_j) & E(x_i x^2_j) & E(x^3_j) \\ E(x^4_i) & E(x^3_i x_j) & E(x^2_i x^2_j) \\ * & E(x^2_i x^2_j) & E(x_i x^3_j) \\ E(x^4_j) \end{pmatrix} \quad \dots\dots\dots(3.47)$$

ここで、Aは係数ベクトル、Bは出力 y_n と入力データ点との積を表わすベクトル、Cは係数に掛る入力データ点の積からなるデータ行列であり、対角項を中心とする対象行列となっている。

基礎関数の各係数は、導出された正規方程式(3.44)式を解くことによって求めることができる。

$$A = C^{-1}B \quad \dots\dots\dots(3.48)$$

ただし、上式が成立つためには、行列Cが正則である必要がある。この逆行列の計算には一般に普及した計算方法が利用できるし、特に対象性を利用した効率的なアルゴリズムを用いれば計算精度を向上させることも可能である。また、逆行列の計算が、数値的に不安定となり失敗した場合には、計算を打ち切る。

3.5.5 全組合せのフィッティング

前節で述べたパラメータの解法を用いて、全組合せNPについてのフィッティングを行う。その結果として、 $p \times NP$ 個の係数が得られる。

$$\{ G_q(C_{q1}, \dots, C_{qp}) \} \quad \dots\dots\dots(3.49)$$

$$(q=1, \dots, NP)$$

GMDHに基づく階層モデルの同定計算においては、逆行列の計算過程において多数個ある基礎関数の候補の幾つかが計算に失敗したとしても、それが直ちにモデル・フィッティングの失敗に繋がる訳ではない。本アルゴリズムでは、逆行列計算で解が得られなかった基礎関数の数をNFとすると、次段以降の処理では基礎関数の総数NPを修正し、NPからこれを減じた数を新しく組合せ数NPと置く。

$$NP \leftarrow NP - NF \quad \dots\dots\dots(3.50)$$

3.6 チェッキング・データによる基礎関数の候補の取捨選択と層数の判定

ここでは、先に求めた基礎関数に対してチェックング・データを用いて評価関数を計算し、これに基づき完全表現の一部を構成する部分表現の候補を選択し、さらにより高い層への計算を続けるか否かの判定を行う。

3.6.1 評価関数の計算

本アルゴリズムでは、基礎関数に対する評価関数 B_v として予測誤差の2乗平均(2.50)式を採用している。 B_v の計算に必要な予測誤差は、前節で得られた各組合せの基礎関数とチェックング・データ用出力ベクトル Y_C 、データ行列 Z_C から計算される。

まず、各基礎関数の予測誤差系列を計算してみよう。チェックング・データの出力の観測値 $y_{t(c)}$ ($c=1, \dots, L_c$)に対する予測値は、フィッティングされたNP個の基礎関数 G_{nq} のそれぞれに

対して、チェックング・データ行列Z Cの列データ(c=1,...,Lc)から作られる入力データ点の組{x_{t(c)-i}, x_{t(c)-j}}あるいは{z_i, z_j}を代入して計算される。ここでも、時刻の表現を簡単にするためにt(c)を単にnあるいはi, jと記す。観測値y_nと区別するためにその予測値をu_{nq}と記すと、これは以下の様に計算される。

$$u_{nq} = G_{nq}(X_i, X_j ; C_{q1}, \dots, C_{qp}) \dots\dots\dots(3.51)$$

(n=1, \dots, L_c, q=1, \dots, NP)

得られた予測値系列は、行列表現すれば、(NP x Lc)行列Uとなる。

$$U = \begin{matrix} u(1,c(L_c)) & u(1,c(L_c-1)) & \dots & u(1,c(1)) \\ : & : & & : \\ u(NP,c(L_c)) & u(NP,c(L_c-1)) & \dots & u(NP,c(1)) \end{matrix} \dots\dots\dots(3.52)$$

このとき、各組合せの基礎関数に対応する予測誤差の2乗平均は、次の様に表わせる。

$$\varepsilon^2_q = \frac{1}{L_c} \sum_{n=1}^{L_c} (y_n - u_{nq})^2 \dots\dots\dots(3.53)$$

(q=1, \dots, NP)

上式は、(3.31)式の出カベクトルYC、予測値行列Uの第q番目の行ベクトルU_qを用いると、次の様にも表現できる。

$$\varepsilon^2_q = \frac{1}{L_t} (YC - Uq^t)^t (YC - Uq^t) \dots\dots\dots(3.54)$$

(q=1, \dots, NP)

こうして得られた2乗平均値ε²_qが組合せqに対する評価関数Ev_qの値となる。

3.6.2 基礎関数の候補の選択

本アルゴリズムにおける基礎関数の候補の取捨選択は、次の原理に基づいて行う。

- (H-4) 部分表現の候補の選択は、2乗平均の意味で最も予測誤差の小さいものからNs個選ぶ。
- (I-5) 選択数Nsは、事前の情報や経験から決めておく。

まず、前節で得られたNP個の基礎関数に対応する予測誤差の2乗平均値

$$\{ \varepsilon^2_1 \ \varepsilon^2_2 \ \varepsilon^2_3 \ \dots \ \varepsilon^2_{NP} \} \dots\dots\dots(3.55)$$

を大小比較し、予測誤差の小さいものから順に並べ、Ns個選び出す。予測誤差に対応する予測値行列Uの行番号をsとすれば、選択された行番号は以下の様に書ける。

$$\{ s(1) \ s(2) \ s(3) \ \dots \ s(Ns) \} \dots\dots\dots(3.56)$$

この操作は、全体の数Npと選択数Nsの比率によって決まる2乗平均のしきい値に基づいて判定を行うことと等価である。最小値から数えてNs番目の2乗平均値が選択判定のしきい値 θ_s となる。

$$\begin{aligned} \theta_s &\leftarrow \varepsilon^2_{s(Ns)} \\ \varepsilon^2_{min} &< \varepsilon^2_{s(Ns)} < \varepsilon^2_{max} \end{aligned} \dots\dots\dots(3.57)$$

このしきい値に基づいて各基礎関数の予測誤差が判定され、下式を満足する基礎関数だけが候補として残され、それ以外は棄却される。

$$\begin{aligned} \{ \varepsilon^2_q \} &\leq \theta_s \\ &(q = 1, \dots, NP) \end{aligned} \dots\dots\dots(3.58)$$

こうして、次の基礎関数が候補として残される。

$$\{ G_{s(i)} ; i=1, \dots, Ns \} \dots\dots\dots(3.59)$$

3.6.3 層数の決定

GMDHに基づく階層モデルのフィッティングにおいては、最適な層数Kは未知であり、フィッティングを1層ずつ進めながら、その結果から決定しなければならない。したがって、層数の決定は、同時に層を進めるか否かの判定でもある。層数Kは、モデルのパラメータの1つであるため、基礎関数の係数の決定と同様に評価関数によって決定される。

本アルゴリズムでは、評価関数として基礎関数と同様に予測誤差の2乗平均値を用いている。ただし、用いる2乗平均値は層内で最小値を示した値が用いられ、隣り合う層間でこれが比較される。その判定は、次の方針で行われる。

- (H-5) 次層に計算を進めるか否かの判定は、現在の層(k)と前層(k-1)の最小予測誤差(2乗平均値)の大小比較を通じて行う。
- 現層(k)が小さい場合は、層を進める。
 - 現層(k)が大きい場合は、前層(k-1)を最終層とし、計算を止める。

この判定は、言換えれば各層にまたがって予測誤差の最小値を探索する操作であり、第1番目の極小点を探索する操作に他ならない。

この操作は、しきい値を置く事と等価である。層打切り判定のしきい値を θ_L とすれば、このしきい値により、各層の最小予測誤差を判定し、下式を満足する層数 k を最終層に決定する。

$$\{ \varepsilon^2_{\min}(k) \} \leq \theta_L \quad \dots\dots\dots(3.60)$$

$$(k = 1, 2, \dots)$$

3.7 中間データの生成と繰返し計算

第1層の基礎関数のフィッティングとその選択には、オリジナル・データ(3.1)式から作られた出力ベクトル Y (3.5)式、データ行列 Z' (3.20)式を用いた。これに対して、第2層以降では、データ行列 Z' はオリジナル・データの代りにフィッティングされた基礎関数の候補を用いて計算した予測値に入替えて使用する。

予測系列の計算は、3.6節の評価関数の計算においてチェック・データ $Z_C(c=1, \dots, L_c)$ に対しては(3.51)式を用いて行ったものと全く同じである。ただし、ここで準備すべきデータは、トレーニング・データ Z_T を含めた全体のデータ($t=1, \dots, L$)に関する計算値であり、しかも使用する基礎関数は、選択された $G_s(s=1, \dots, N_s)$ に限定される。この場合の予測値 u_{ts} は、次の様に定義される。

$$u_{ts} = G_{ts}(X_i, X_j; C_{s1}, \dots, C_{sp}) \quad \dots\dots\dots(3.61)$$

$$(t=1, \dots, L, s=1, \dots, N_s)$$

得られた予測値系列は、行列表現すれば、 $(N_s \times L)$ 行列 U' となる。

$$U' = \begin{Bmatrix} u(1,L) & u(1,L-1) & \dots & u(1,1) \\ \vdots & \vdots & & \vdots \\ u(N_s,L) & u(N_s,L-1) & \dots & u(N_s,1) \end{Bmatrix} \quad \dots\dots\dots(3.62)$$

予測値行列 U' をデータ行列 Z' とする。トレーニング・データとチェック・データの列番号($t(1), \dots, t(L_t)$)、($c(1), \dots, c(L_c)$)は第1層で既に求まっており、第2層以降これを変更せずに計算を進める。

$$Z' \leftarrow U' \quad \dots\dots\dots(3.63)$$

$$Z_T, Z_C \leftarrow Z'$$

以降、最適層数が決定されるまで、3.5~3.7節の処理を繰り返す。

3.8 完全表現の構築

最終層の判定が下ると、これまでのフィッティングの結果を総合して、完全表現を構築することができる。完全表現の構築は、次の方針で行う。

(H-6) 最終層で選択されたNs個の基礎関数の中で、評価関数が最小値を示す唯一つの基礎関数を、完全表現における最終層Kの部分表現として選択する。

すなわち、最終層Kの唯一の部分表現が決定できれば、その部分表現に採用された2つの入力点から前層の2つの部分表現が決定でき、同様にして下位の層に溯ることにより、第1層までの全ての部分表現が決定できる。完全表現は、モデルが階層構造(すなわち入れ子式の関数)のために、一挙に数式で表現することは困難であるが、おおよそ以下の通りである。

一般に、第K層のi番目の基礎関数を $G_i^{(K)}$ と記し、その2つの入力を $\{z_{i,1}^{(K)}, z_{i,2}^{(K)}\}$ と記せば、完全表現は1つの基礎関数と2つのデータによって、次の様に表せる。

$$y_t = G_i^{(K)}(z_{i,1}^{(K)}, z_{i,2}^{(K)}) + \varepsilon_t \quad \dots\dots\dots(3.64)$$

もし、層数が $K>1$ の場合には2つのデータに対応する前層(第K-1層)の基礎関数が存在し、これに対応する(i,1)番目、(i,2)番目の基礎関数を $\{G_{i,1}^{(K-1)}, G_{i,2}^{(K-1)}\}$ 、またこれらの入力データ点の番号を同様の方法で定義すると、

$$\begin{aligned} z_{i,1}^{(K)} &= G_{i,1}^{(K-1)}(z_{i,1,1}^{(K-1)}, z_{i,1,2}^{(K-1)}) \quad \dots\dots\dots(3.65) \\ z_{i,2}^{(K)} &= G_{i,2}^{(K-1)}(z_{i,2,1}^{(K-1)}, z_{i,2,2}^{(K-1)}) \end{aligned}$$

と書け、完全表現は第K層から第K-1層までの3つの基礎関数と4つのデータを用いて表せる。

$$y_t = G_i^{(K)}(G_{i,1}^{(K-1)}(z_{i,1,1}^{(K-1)}, z_{i,1,2}^{(K-1)}), G_{i,2}^{(K-1)}(z_{i,2,1}^{(K-1)}, z_{i,2,2}^{(K-1)})) + \varepsilon_t \quad \dots\dots\dots(3.66)$$

同様に、K-2層の基礎関数が存在すれば($K>2$)、

$$\begin{aligned} z_{i,1,1}^{(K-1)} &= G_{i,1,1}^{(K-2)}(z_{i,1,1,1}^{(K-2)}, z_{i,1,1,2}^{(K-2)}) \quad \dots\dots\dots(3.67) \\ z_{i,1,2}^{(K-1)} &= G_{i,1,2}^{(K-2)}(z_{i,1,2,1}^{(K-2)}, z_{i,1,2,2}^{(K-2)}) \\ z_{i,2,1}^{(K-1)} &= G_{i,2,1}^{(K-2)}(z_{i,2,1,1}^{(K-2)}, z_{i,2,1,2}^{(K-2)}) \\ z_{i,2,2}^{(K-1)} &= G_{i,2,2}^{(K-2)}(z_{i,2,2,1}^{(K-2)}, z_{i,2,2,2}^{(K-2)}) \end{aligned}$$

となり、完全表現は、7つの基礎関数と8つのデータを用いて表されることになる。以下同様

にして第1層まで代入して、完全表現を得る。

第1層のデータは、基礎関数の番号に対応した任意の遅れ時間 m_1, m_2 を持つオリジナル・データ点である。

$$\begin{aligned} z_{i, \dots, 1}^{(1)} &= X_{t-m_1} \\ z_{i, \dots, 2}^{(1)} &= X_{t-m_2} \end{aligned} \dots\dots\dots(3.68)$$

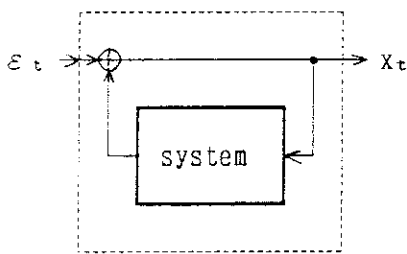
3.9 フィードバック型および多変数システムのフィッティング

3.2~3.8節では1入力1出力型階層モデルの同定用アルゴリズムについて説明した。ここでは、フィードバック型および多変数型モデルの同定について補則説明をする。

3.9.1 1変数フィードバック・モデル

システムの入力値がある時間遅れを持って再び自分自身のシステムの入力に印加されるという自己回帰(AR)型の信号の場合、1変数型の階層モデルが定義できる。

自己回帰型変数 x の観測値を $\{ x_t; t=1, \dots, N \}$ とする時、GMDHに基づく階層モデル f は、(3.2)式の左辺 y_t を x_t で置き換えた次式で表現できる。



$$x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-m}) + \epsilon_t \dots\dots\dots(3.69)$$

Fig.3.4 Univariate GMDH model. そのシステム構造は、ARモデルと同様に、Fig.3.4に示すものが仮定される。 ϵ_t はシステムを駆動する雑音源で、白色またはそれに近いものである。

このモデルの同定には、(3.5)式に与えた出力ベクトル Y を x で置き換えた、

$$Y = (x_{L+M} \ x_{L+M-1} \ x_{L+M-2} \ \dots \ x_{M+1})^t \dots\dots\dots(3.70)$$

を使用すれば良い。

3.9.2 多変数システムのモデリングの問題

多入力1出力システムに対しても、階層型モデルが定義できる。 $(K+1)$ 変数のシステムの入出力の観測値として、以下のものが与えられているとしよう。

$$\begin{aligned} \{ y_t, x_{it} \} & \dots\dots\dots(3.71) \\ (i=1, \dots, K, t=1, \dots, N) \end{aligned}$$

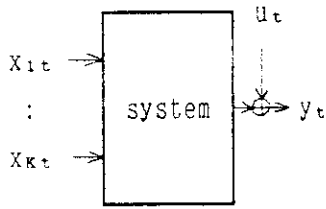


Fig.3.5 Multiple-input/ single-output system

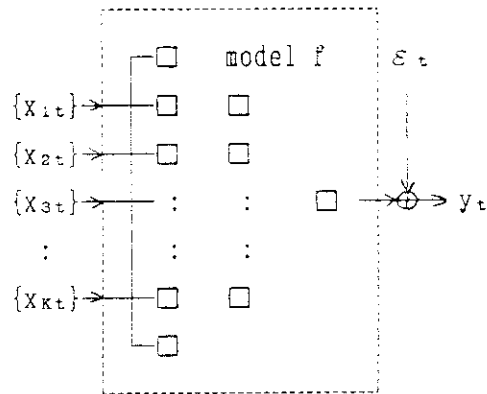


Fig.3.6 Multiple-input/ multiple-output GMDH model.

システムの出力 y_t がフィードバック経路を持たないならば、このシステムはFig.3.5に示すK入力1出力モデルとなり、このとき、階層モデル f は次式で表現される。

$$y_t = f(X_{1,t-1}, X_{1,t-2}, \dots, X_{1,t-M}, \dots, X_{K,t-1}, X_{K,t-2}, \dots, X_{K,t-M}) + \varepsilon_t \quad \dots\dots\dots(3.72)$$

このGMDHに基づく階層モデルを図示したものがFig.3.6である。1変数の場合と同様に、 ε_t は残差項であり、全ての入力変数と無相関な量である。したがって、システムの付加雑音 u_t が白色雑音のときだけ、正しい同定が可能となる。

逆に、付加雑音 u_t が入力変数と何等かの相関を持つ場合は、出力 y_t を入力にもどしたフィードバック型のモデルを考える必要がある。

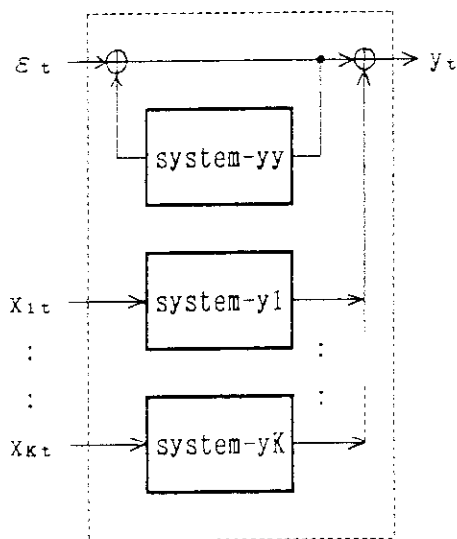


Fig.3.7 Multivariate GMDH model.

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-M}, X_{1,t-1}, X_{1,t-2}, \dots, X_{1,t-M}, \dots, X_{K,t-1}, X_{K,t-2}, \dots, X_{K,t-M}) + \varepsilon_t \quad \dots\dots\dots(3.73)$$

これは、Fig.3.7に示す様な多変数モデルである。サブシステム yy が、Fig.3.5における y_t の付加雑音 u_t を白色化する。したがって、ARモデルと同様に、 u_t は入力変数と相関があっても雑音源 ε_t が白色雑音であれば同定が可能となる。

ただし、GMDHに基づく多変数階層モデルのフィッティング・アルゴリズムは、構造上の制約から、1回の同定

計算で多入力多出力モデルを一度に同定することはできず、変数の個数分だけ(3.73)式を作っ
て同定を繰返し、これを後で結合する必要がある。このモデルの詳細は第5章で述べる。

フィードバック型とオープン・ループ型のフィッティング・アルゴリズムの違いは、データ
行列Zの変数の1つに出力ベクトルYの変数と同じものを入れるか否かだけである。したがっ
て、以下の節ではオープン・ループ型のK入力1出力階層モデルのアルゴリズムを例にとって
説明する。

3.9.3 多変数データ行列の生成

(3.71)式で与えられた各変数のN個のオリジナル・データから次のK個の行列Z_i(i=1,..,K)を
作る。

$$Z_i = \begin{pmatrix} X_{i,L} & X_{i,L-1} & \dots & X_{i,1} \\ \vdots & \vdots & & \vdots \\ X_{i,L+M} & X_{i,L+M-1} & \dots & X_{i,M+1} \end{pmatrix} \quad (i=1,\dots,K) \quad \dots\dots\dots(3.74)$$

各変数のデータ行列Z_iの性質については、次の事が言える。

- 1) 各列データは、ある時刻に対応した1からMの遅れ時間データの組であるから、統計的な
平均操作に関しては互に独立なデータの組とみなせる。したがって、列の入替え操作が
可能。
- 2) 基礎関数の2つの入力データ点は、2つの異なる遅れ時間点、すなわち2つの異なる行
データに対応し、フィッティング演算は任意の2つの行データに対して行われる。した
がって、行データにその変数名、遅れ時間点の情報が付されていれば、行データは行列
のどの行にあっても良い。

そこで、多変数階層モデルの同定の場合には、これらK個のデータ行列をZ_iを縦に並べて1つ
にまとめた、(KM x L)行列Zを使用することが可能である。

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_K \end{pmatrix} \quad \dots\dots\dots(3.75)$$

3.9.4 多変数データ行列Zの行に関する前処理

(1) 相関関数の推定

3.3節で述べた1変数のデータ行列Zの前処理と同様に、各入力変数の中で弱い相関関係に
ある時間遅れ点を棄却するために、ここではまず、各入力変数の相関関数を計算しておく。

出力変数 y_t に対する入力変数 $\{x_{1t}, \dots, x_{kt}\}$ の各相関関数は、1変数の場合と同様に、(3.7)式で定義でき、(3.10)式または(3.11)式を用いて計算できる。

$$\{ R_{YX_{1m}}, R_{YX_{2m}}, \dots, R_{YX_{km}} \} \dots\dots\dots(3.76)$$

$$(m=1, \dots, M)$$

また、各相関関数の規格化は、(3.15)式により計算できる。

$$\{ R'_{YX_{1m}}, R'_{YX_{2m}}, \dots, R'_{YX_{km}} \} \dots\dots\dots(3.77)$$

$$(m=1, \dots, M)$$

(2) 弱い相関を持つ時間遅れ点の棄却

先に求めた規格化された相関関数を用いて、変数毎に遅れ時間区間 $[1, M]$ の範囲から弱い相関関係にある遅れ点のデータを棄却する処理を行う。棄却判定に際して、1変数の時と同様にあるしきい値を設けておく。

(I-2) 棄却すべき相関の度合い $R_{i\max}$ を解析前の事前情報や経験から変数毎に決定し、これを棄却判定のしきい値 θ_{yir} とする。

$$\theta_{yir} \leftarrow R_{i\max} \dots\dots\dots(3.78)$$

$$0 < R_{i\max} < 1 \quad (i=1, \dots, K)$$

ここで、注意すべきは、 $R_{i\max}$ の値である。相互相関関数は、入出力変数が同一の強い雑音源で駆動されている場合を除けば、通常は比較的小さい値を示すのが普通であり、また変数の種類によって大きさの異なるのが普通である。したがって、しきい値に各変数とも一律のある大きさをもった値を採用した場合には、システムに対する寄与が無視できない変数であっても、その変数のデータを全て棄却してしまう場合が在り得る。

このため、しきい値の設定に際しては、少なくとも第1層の組合せにおいて弱い相関ではあっても代表的な成分が残る様に設定する必要がある。

さて、(3.78)式で与えた各変数のしきい値により、多変数データ行列 Z の行の判別を行う。判別は、該当する変数の行毎に行う。

$$| R'_{y1, m} | \geq \theta_{y1r} \dots\dots\dots(3.79)$$

$$(m=1, \dots, M, i=1, \dots, K)$$

各変数毎に行った棄却判定の結果、変数毎に $\{M'(1), \dots, M'(K)\}$ 行の時間遅れ点が残ったとしよう。

$$\{ \Gamma_{11} \Gamma_{12} \dots \Gamma_{1M'(1)} \}, \dots, \{ \Gamma_{K1} \Gamma_{K2} \dots \Gamma_{KM'(K)} \} \dots\dots\dots(3.80)$$

変数毎に、データ行列 Z_i 中の弱い相関行のデータを削除した行列を Z'_i と置くと、(3.74) 式の変数データ行列 Z は、次の行列 Z' となる。

$$Z' = \begin{pmatrix} Z'_1 \\ Z'_2 \\ \vdots \\ Z'_K \end{pmatrix} \quad \dots\dots\dots(3.81)$$

ここで、(3.80) 式の行番号をまとめて、

$$\{ r_1 \ r_2 \ \dots \ r_{M'} \} \quad \dots\dots\dots(3.82)$$

ただし、 $M' = M'(1) + \dots + M'(K)$

という様に、1 変数の場合と同じ記号で表せば、前処理後の行数 Z' はサイズ $(M' \times L)$ の行列となり、第 1 層のデータ行列の分割以降の処理は 1 変数のアルゴリズムが全く同様に適用できる。

4. 基礎関数の係数推定と統計的関数

本章では、基礎関数にKolmogorov-Gaborの多項式を採用したGMDHにおいて、基礎関数の係数の推定に使われる統計的関数について、若干の考察をする。

4.1 高次の相関関数とその性質

本章の議論を容易にするために、ここで、変数 x の観測データ $\{x_t; t=1, \dots, N\}$ に対する平均(1次モーメント)および2次から4次までの自己相関関数の推定式を定義しておこう。

$$\mu_t = \frac{1}{N} \sum_{t=1}^N x_t \quad \dots\dots\dots(4.1)$$

$$R_t(m1) = \frac{1}{N} \sum_{t=1}^N x_t x_{t-m1} \quad \dots\dots\dots(4.2)$$

$$R_t(m1, m2) = \frac{1}{N} \sum_{t=1}^N x_t x_{t-m1} x_{t-m2} \quad \dots\dots\dots(4.3)$$

$$R_t(m1, m2, m3) = \frac{1}{N} \sum_{t=1}^N x_t x_{t-m1} x_{t-m2} x_{t-m3} \quad \dots\dots\dots(4.4)$$

ここで、 $m1, \dots, m4$ は、時刻 t に対する任意の時間遅れ点を表し、最大遅れ時間点を M とすれば、

$$0 \leq m1, m2, m3 \leq M \quad \dots\dots\dots(4.5)$$

を満たす値をとるものとする。

これらの推定式は、サンプル数 N を無限大にしたときの極限值が真の値を与えるが、 N が有限の場合は推定誤差を持ち、さらに、 N が一定の場合でも異なる時刻(時間区間)のデータを用いた推定では、推定値は偏り誤差を持つ。この意味から(4.1)~(4.4)式の各関数には時刻 t の添え字が付されている。

次に、2つの出力変数 y_t と入力変数 x_t 間の2次、3次の相互相関関数を同様に定義しておこう。観測データを $\{y_t, x_t; t=1, \dots, N\}$ とすれば、各推定式は次の様に定義される。

$$R_{yx_t}(m_1) = \frac{1}{N} \sum_{t=1}^N y_t x_{t-m_1} \dots\dots\dots(4.6)$$

$$R_{yxx_t}(m_1, m_2) = \frac{1}{N} \sum_{t=1}^N y_t x_{t-m_1} x_{t-m_2} \dots\dots\dots(4.7)$$

さて、各相関関数は時間遅れ変数の値に応じて対称性を持つ。以下では、サンプル数Nが無
限大の場合を考えよう。2次の相関関数は、時間遅れ $m_1=0$ 軸に対して2通りの対称性のあるこ
とが知られている。

$$R(m_1) = R(-m_1) \dots\dots\dots(4.5)$$

また、3次の相関関数は、時間遅れ点 $m_1=0, m_2=0, m_1+m_2=0$ の3つの軸に対して対称関係にある
ことが知られている³⁷⁾。

$$\begin{aligned} R(m_1, m_2) &= R(m_2, m_1) \\ &= R(-m_1, m_2 - m_1) = \dots \\ &= R(-m_2, -m_2 + m_1) = \dots \dots\dots(4.6) \end{aligned}$$

すなわち、2つの時間遅れ m_1, m_2 が交換可能で、しかも時間遅れ点の基準の取り方に3通りの
方法があるため、6通りの対称な関数が存在する。

同様に、4次の相関関数は、 $m_1=0, m_2=0, m_3=0, m_1+m_2+m_3=0$ の4つの軸に対して対称である。

$$\begin{aligned} R(m_1, m_2, m_3) &= R(m_2, m_3, m_1) = R(m_3, m_1, m_2) \\ &= R(m_1, m_3, m_2) = R(m_2, m_1, m_3) = R(m_3, m_2, m_1) \\ &= R(-m_1, m_2 - m_1, m_3 - m_1) = \dots \\ &= R(-m_2, -m_2 + m_1, m_3 - m_2) = \dots \\ &= R(-m_3, -m_3 + m_1, -m_3 + m_2) = \dots \dots\dots(4.7) \end{aligned}$$

すなわち、3つの遅れ点 m_1, m_2, m_3 が全て交換可能(6通り)で、しかも時間遅れ点の基準の取り
方が4通りあるから合計24通りの対称な関数が存在する。

4.2 正規方程式の統計的意味

GMDHに基づく階層モデルの同定で仮定される基礎関数の関数形は様々であるが、本論文では
Kolmogorov-Gaborの多項式を採用し、第3章ではこの基礎関数の係数の解を与える正規方程式
(3.44)式を導出した。ここでは、この方程式を用いた解の推定計算で、どの様な統計的関数が

使われているかを考察をする。

(3.44)式の正規方程式は、係数ベクトルA(3.45)式、出力データ点と入力データ点の積からなるベクトルB、入力データ点の積からなるデータ行列Cから構成され、この中でベクトルB、行列Cに種々の統計的関数が用いられている。

この正規方程式は、第1層において、時刻tに対する2つの任意の遅れ時間点i,jにおける入力の観測データ点 x_{t-i}, x_{t-j} で記述される基礎関数 G_{ij} (3.41)式を、出力点 y_t に対してフィッティングした場合に導出されたものである。

$$y_t \sim G_{ij}^{(1)}(x_{t-i}, x_{t-j}) \dots\dots\dots(4.8)$$

このため、ベクトルB、行列Cに現れる統計的関数も時刻tおよび2つの遅れ時間点t-i,t-jの各データ点で定義される関数となる。

第2層以降では、入力データ点 $\{x_{t-i}, x_{t-j}\}$ を中間変数 $\{z_i, z_j\}$ に置き換えた基礎関数が出力点 y_t にフィッティングされる。

$$y_t \sim G_{ij}^{(k)}(z_i^{(k)}, z_j^{(k)}) \dots\dots\dots(4.9)$$

この場合も正規方程式は第1層と同様に適用できるが、各中間データ点は時刻の概念が複雑となるため、ベクトルB、行列Cに現れる統計的関数の解釈は困難になる。

以下では、第1層に現れる統計的関数について検討してみよう。

いま、基礎関数に関する3つの観測データ点 $\{y_t, x_{t-m_1}, x_{t-m_2}\}$ が与えられているとする。ここで、2つの入力データ点の時刻の表現が簡単になる様に次の様に定義しておこう。

$$\begin{aligned} t_1 &= t - m_1 && \dots\dots\dots(4.11) \\ t_2 &= t - m_2 \end{aligned}$$

また、後の表現の便宜をはかり、時刻 t_1 と t_2 の差 τ を次の様に定義しておく。

$$\begin{aligned} \tau &= t_1 - t_2 && \dots\dots\dots(4.12) \\ &= m_2 - m_1 \end{aligned}$$

ベクトルBは、(3.46)式では統計的平均操作の記号 $E(\cdot)$ を用いて定義されているが、これを前節で定義した出力 y_t の1次モーメントおよび出力と入力の2次、3次の相互相関関数で置き換えれば、次の様に表現できる。

$$B = (\mu_{y_t} \text{ Ryx}_t(m_1) \text{ Ryx}_t(m_2) \text{ Ryxx}_t(m_1, m_1) \text{ Ryxx}_t(m_1, m_2) \text{ Ryxx}_t(m_2, m_2))^T \dots\dots\dots(4.13)$$

上式の統計的関数はいずれも、時刻 t を基準に定義された推定式から成り立ち、各相関関数に現れる時間遅れ点は、 $m1$ または $m2$ である。

一方、(3.47)式の行列 C は、入力の1次モーメントと4次までの自己相関関数によって次の様に表現できる。

$$C = \begin{pmatrix} 1 & \mu_{t_1} & \mu_{t_2} & R_{t_1}(0) & R_{t_1}(\tau) & R_{t_2}(0) \\ & R_{t_1}(0) & R_{t_2}(-\tau) & R_{t_1}(0,0) & R_{t_1}(0,\tau) & R_{t_2}(0,-\tau) \\ & & R_{t_2}(0) & R_{t_1}(0,-\tau) & R_{t_1}(\tau,\tau) & R_{t_2}(0,0) \\ & & & R_{t_1}(0,0,0) & R_{t_1}(0,0,\tau) & R_{t_2}(0,-\tau,-\tau) \\ & * & & & R_{t_1}(0,\tau,\tau) & R_{t_2}(-\tau,-\tau,-\tau) \\ & & & & & R_{t_2}(0,0,0) \\ & & & & & \dots\dots\dots \end{pmatrix} \quad (4.14)$$

上式中に現れる統計的関数は、時刻 t_1 または t_2 を基準とした推定式であり、各相関関数の時間遅れ点は、0または τ だけである。(4.5)~(4.7)式に述べた対称関係は、異なる2つの時刻の基準 t_1, t_2 に関して、次の様になる。

$$R_{t_1}(\tau) = R_{t_2}(-\tau) \quad \dots\dots\dots(4.15)$$

$$\begin{aligned} R_{t_1}(0,\tau) &= R_{t_1}(\tau,0) = R_{t_2}(-\tau,-\tau) \\ R_{t_1}(\tau,\tau) &= R_{t_2}(0,-\tau) = R_{t_2}(-\tau,0) \end{aligned} \quad \dots\dots\dots(4.16)$$

$$\begin{aligned} R_{t_1}(0,0,\tau) &= R_{t_1}(0,\tau,0) = R_{t_1}(\tau,0,0) = R_{t_2}(-\tau,-\tau,-\tau) \\ R_{t_1}(0,\tau,\tau) &= R_{t_1}(\tau,0,\tau) = R_{t_1}(\tau,\tau,0) \\ &= R_{t_2}(0,-\tau,-\tau) = R_{t_2}(-\tau,0,-\tau) = R_{t_2}(-\tau,-\tau,0) \\ R_{t_1}(\tau,\tau,\tau) &= R_{t_2}(0,0,-\tau) = R_{t_2}(0,-\tau,0) = R_{t_2}(-\tau,0,0) \end{aligned} \quad \dots\dots\dots(4.17)$$

したがって、(4.13)式は数ある表現の中の1例にすぎない。いま、行列 C の時間遅れ τ を、上の対象関係を用いて、計算の都合の良い様に全て正の値に取った表現は次の様になる。

$$C = \begin{pmatrix} 1 & \mu_{t_1} & \mu_{t_2} & R_{t_1}(0) & R_{t_1}(\tau) & R_{t_2}(0) \\ & R_{t_1}(0) & R_{t_1}(\tau) & R_{t_1}(0,0) & R_{t_1}(0,\tau) & R_{t_1}(\tau,\tau) \\ & & R_{t_2}(0) & R_{t_1}(\tau,\tau) & R_{t_1}(\tau,\tau) & R_{t_2}(0,0) \\ & & & R_{t_1}(0,0,0) & R_{t_1}(0,0,\tau) & R_{t_1}(0,\tau,\tau) \\ & * & & & R_{t_1}(0,\tau,\tau) & R_{t_1}(0,0,\tau) \\ & & & & & R_{t_2}(0,0,0) \\ & & & & & \dots\dots\dots \end{pmatrix} \quad (4.18)$$

(4.13)式および(4.18)式は、サンプル数Nが十分に大きく時刻の依存性が無視できる場合には、関数の時刻に関する添え字 t, t_1, t_2 が取れ、それぞれ、次の様になる。

$$B = (\mu_y \quad R_{yx}(m1) \quad R_{yx}(m2) \quad R_{yxx}(m1, m1) \quad R_{yxx}(m1, m2) \quad R_{yxx}(m2, m2)) \dots\dots\dots(4.19)$$

$$C = \begin{pmatrix} 1 & \mu & \mu & R(0) & R(\tau) & R(0) \\ & R(0) & R(\tau) & R(0,0) & R(0, \tau) & R(\tau, \tau) \\ & & R(0) & R(\tau, \tau) & R(\tau, \tau) & R(0,0) \\ & & & R(0,0,0) & R(0,0, \tau) & R(0, \tau, \tau) \\ & * & & & R(0, \tau, \tau) & R(0,0, \tau) \\ & & & & & R(0,0,0) \end{pmatrix} \dots\dots\dots(4.20)$$

以上の例は、第1層のフィッティングで現れる相関関数であった。第2層以降のフィッティングに際しても、中間変数に対して同じ相関関数が使われる。ただし、遅れ時間の解釈は困難なものとなる。正規方程式の行列C(3.47)式を振り返って見ると、3次および4次の相関関数を生み出す要因は、基礎関数(3.41)式に含まれる2次の項であることが分かる。第2層では、基礎関数は1から4次の項を含むため、行列Cの要素は2次から8次の相関関数を組合せ加算した統計的関数になり、もはや記述は不可能となる。

以上をまとめると、次の事が言える。Kolmogorov-Gaborの多項式を基礎関数としたGMDHでフィッティングした場合、係数の最小2乗解は、

- 1) 1次の多項式の場合、入出力の情報は通常の(2次の)相関関数によって完全に表現される。
- 2) 2次の多項式の場合、出力に関する情報は3次までの相関関数によって、また入力に関する情報は4次までの相関関数によって表現される。
- 3) フィッティングに用いられる時間遅れ点の情報は、0, m1, m2, (m1-m2)の4点である。すなわち、基礎関数のフィッティングでは、出力 y_t に対する2つの入力点の遅れ時間以外に、これらの差の遅れ時間も考慮される。

5. GMDHとARモデルの比較

本章では、GMDHによる線形階層モデルと自己回帰モデル(AR)に関する若干の理論的な比較を試みる。

一般に、GMDHに基づく階層モデルは、システム構造との関係が明らかでなく、モデルの物理的な解釈が容易でないとされてきた。その中で、基礎関数に線形のものを採用した線形階層モデルは、第2章の基礎関数の議論から分かる様に、フィッティング過程は見掛け上階層構造を形成するが、最終的に得られるモデル自体は本質的に単層構造であり、ARモデルと良く似ている。一方、ARモデルは最も普及したモデリング手法の1つであり、その数学的記述とシステムの構造(物理モデル)との関係も明確である。

そこで、両モデルを比較することにより、GMDHに基づく階層モデルの物理的解釈を試みるのが本章のねらいである。

5.1 フィッティング構造

本節では、GMDHとARモデルの両フィッティング・アルゴリズムにおける情報取り込みの構造について比較する。議論を簡単にするために、ここでは両モデルとも1変数の場合を考える。いま、信号 x の観測データ $\{x_t\}$ が与えられているものとする。

1次元ARモデルは、モデルの次数を M 、係数を $\{a_m; m=1, \dots, M\}$ 、白色雑音からなる残差成分を ε_t とすれば次の様に表される。

$$x_t = \sum_{m=1}^M a_m x_{t-m} + \varepsilon_t \quad \dots\dots\dots(5.1)$$

右辺第1項は、 M 個の連続した遅れ時間点と係数の積の総和から成立ち、これは、 M 個のパラメータを持つ線形多項式 f と見なすことができる。

$$f(x_{t-1}, \dots, x_{t-M}; a_1, \dots, a_M) = \sum_{m=1}^M a_m x_{t-m} \quad \dots\dots\dots(5.2)$$

ARモデルの同定とは、観測データ $\{x_t\}$ に対して関数 f をあてはめて、次数 M と M 個のパラメータを決定することに他ならない。

$$x_t \sim f(x_{t-1}, \dots, x_{t-M}; a_1, \dots, a_M) \quad \dots\dots\dots(5.3)$$

モデル次数Mは未知であるため、フィッティング・アルゴリズムでは、モデル次数mと使用する遅れ時間データを1点ずつ増やししながら、うまく近似できるまで(評価関数が最小になるまで)関数型を拡張していく方法を取っている。

$$\begin{aligned}
 X_t &\sim f^{(1)}(X_{t-1}; a_1) \\
 &\sim f^{(2)}(X_{t-1}, X_{t-2}; a_1, a_2) \quad \dots\dots\dots(5.4) \\
 & \vdots \\
 &\sim f^{(m)}(X_{t-1}, \dots, X_{t-m}; a_1, \dots, a_m) \\
 & \vdots
 \end{aligned}$$

ここで、関数fの添え字(1)~(m)は次数を表す。

このフィッティング過程においては、2次以降の計算は前次数の計算結果が利用できるが、それだけでは不足で、新しい次数の計算には必ず新しいデータ点の情報を取り込む必要がある。すなわち、関数の拡張は観測データに関して階層的ではなく平面的に行うアルゴリズムとなっている。

一方、GMDHに基づく階層モデルは、2入力型の基礎関数Gを用いた場合、第1層では2つのデータ点で記述される3パラメータのモデルfが、まず1~Mの遅れ点の全組合せについて計算される。ついで、層を増やす毎に前層の結果を次々と基礎関数に代入して、構造を拡張していくアルゴリズムとなっている。第2層では4つのデータ点と9つのパラメータ、第3層では8つのデータ点と27のパラメータが同定される。

$$\begin{aligned}
 X_t &\sim G_1^{(1)}(X_{t1}, X_{t2}) \\
 &= f^{(1)}(X_{t1}, X_{t2}; C_1, \dots, C_3) \quad \dots\dots\dots(5.5a)
 \end{aligned}$$

$$\begin{aligned}
 &\sim G_1^{(2)}(Z_1^{(2)}, Z_2^{(2)}) \\
 &= G_1^{(2)}(G_1^{(1)}(X_{t1'}, X_{t2'}), G_2^{(1)}(X_{t3'}, X_{t4'})) \\
 &= f^{(2)}(X_{t1'}, X_{t2'}, \dots, X_{t4'}; C_1', \dots, C_9') \quad \dots\dots\dots(5.5b)
 \end{aligned}$$

$$\begin{aligned}
 &\sim G_1^{(3)}(Z_1^{(3)}, Z_2^{(3)}) \\
 &= G_1^{(3)}(G_1^{(2)}(Z_1^{(2)}, Z_2^{(2)}), G_2^{(2)}(Z_3^{(2)}, Z_4^{(2)})) \\
 &= G_1^{(3)}(G_1^{(2)}(G_1^{(1)}(X_{t1''}, X_{t2''}), G_2^{(1)}(X_{t3''}, X_{t4''})), \\
 &\quad G_2^{(2)}(G_3^{(1)}(X_{t5''}, X_{t6''}), G_4^{(1)}(X_{t7''}, X_{t8''}))) \\
 &= f^{(3)}(X_{t1''}, X_{t2''}, \dots, X_{t8''}; C_1'', \dots, C_{27}'') \quad \dots\dots\dots(5.5c)
 \end{aligned}$$

ここで、 $t-1 \geq t_1 > t_2 \geq t-M$
 $t-1 \geq t_1, t_3 > t_2, t_4 \geq t-M$
 $t-1 \geq t_1, t_3, \dots, t_7 > t_2, t_4, \dots, t_8 \geq t-M$

こうして、うまく近似できるまで(評価関数を最小にするまで)フィッティングが続けられる。階層モデルのフィッティング計算では、層を進める際に関数の拡張と新しいデータ点の取り込みが同時に行われるが、この新しいデータ点の取り込みは、第1層のみが直接的であり、第2層以降は前層の基礎関数を通じて間接的に行われる。

逆に言えば、この同定アルゴリズムでは、第1層においては遅れ時間点の全組合せについて基礎関数をフィッティングしなければならないし、第2層以降も層数が決定されるまで全ての候補についてフィッティングしなければならない。ただし、層を進める際に新たに取り込む時間遅れデータ点は、ARモデルの様に遅れ時間t-1から順次1点ずつ取り込むのではなく、常に[t-1, t-M]の範囲から寄与の大きい任意の点を取り込み、しかも重複した使用も許される。

5.2 基礎関数の係数とARモデルの係数

ここでは、線形GMDHの基礎関数の係数とARモデルの係数との関係を見てみよう。GMDHに基づく階層モデルとしては、Kolmogorov-Gaborの1次多項式を基礎関数とし、出力変数が入力変数と同じ回帰型のデータ{x_t}に対するフィッティングを考える。

このモデルの第1層におけるフィッティングは次の様に書ける。

$$\begin{aligned}
 x_t &\sim G(x_{t-m_1}, x_{t-m_2}; c_1, c_2, c_3) \\
 &= c_1 + c_2 x_{t-m_1} + c_3 x_{t-m_2} \dots\dots\dots(5.6)
 \end{aligned}$$

正規方程式(3.44)式は、(4.20)式の表現を用いると、具体的には次の連立方程式となる。

$$\begin{aligned}
 \mu &= c_1 + (c_2 + c_3)\mu \dots\dots\dots(5.7) \\
 R(m_1) &= c_1 \mu + c_2 R(0) + c_3 R(\tau) \\
 R(m_2) &= c_1 \mu + c_2 R(\tau) + c_3 R(0)
 \end{aligned}$$

これを解くと、下記の3つの係数が得られる。

$$\begin{aligned}
 c_1 &= \left(1 - \frac{R(m_1) + R(m_2)}{R(0) + R(\tau)} \right) \mu \dots\dots\dots(5.8) \\
 c_2 &= \frac{(R(m_1) - \mu^2)(R(0) - \mu^2) - (R(m_2) - \mu^2)(R(\tau) - \mu^2)}{(R(0) - \mu^2)^2 - (R(\tau) - \mu^2)^2} \\
 c_3 &= \frac{(R(m_2) - \mu^2)(R(0) - \mu^2) - (R(m_1) - \mu^2)(R(\tau) - \mu^2)}{(R(0) - \mu^2)^2 - (R(\tau) - \mu^2)^2}
 \end{aligned}$$

ARモデルのフィッティングの際には、あらかじめデータの平均値を0とする前処理が良く用いられている。そこで、平均値を0としてみると、(5.8)式は次の様になる。

$$\begin{aligned}
 c_1 &= 0 && \dots\dots\dots(5.9) \\
 c_2 &= \frac{R(m_1)R(0) - R(m_2)R(\tau)}{R(0)^2 - R(\tau)^2} \\
 c_3 &= \frac{R(m_2)R(0) - R(m_1)R(\tau)}{R(0)^2 - R(\tau)^2}
 \end{aligned}$$

一方、ARモデルの場合、Yule-Walker方程式を解くことにより係数が求まるが、次数が2次の時には2つの係数 $a_{2,2}, a_{2,1}$ は、次の様な解になる。

$$\begin{aligned}
 a_{2,2} &= \frac{R(0)R(2) - R(1)^2}{R(0)^2 - R(1)^2} && \dots\dots\dots(5.10) \\
 a_{2,1} &= \frac{R(0)R(1) - R(2)R(1)}{R(0)^2 - R(1)^2}
 \end{aligned}$$

ここで、比較のために、GMDHの係数の解(5.9)式において2つの遅れ時間を $m_1=1, m_2=2$ とすれば、その差の遅れ時間は $\tau=1$ となり、(5.9)式は次の様に上式と全く一致するのが分かる。

$$\begin{aligned}
 c_2 &= \frac{R(1)R(0) - R(2)R(1)}{R(0)^2 - R(1)^2} && \dots\dots\dots(5.11) \\
 c_3 &= \frac{R(2)R(0) - R(1)R(1)}{R(0)^2 - R(1)^2}
 \end{aligned}$$

すなわち、線形GMDHの第1層における基礎関数の係数解は、ARモデルと全く同じ構造を持っているのが分かった。これは、Kolmogorov-Gaborの多項式(2.6)式において、高次の項を断ち切った場合にARモデルの表現を与える事を考えると、当然の帰結と言えよう。

両者の異なる点は、ARモデルの2次の場合には取り込まれる遅れ時間点は1,2に限定されているのに対して、GMDHの場合には(5.9)式における遅れ時間 m_1, m_2 が範囲 $[1, M]$ 内で、任意に選択できる点である。

5.3 GMDHの完全表現とARモデルの関係

前節では第1層の基礎関数の係数とARモデルの係数を比較したが、ここでは基礎関数を階層的に配置して得られる線形GMDHの完全表現とARモデルの係数の関係を見てみよう。このため、まず基礎関数(5.6)式を一般化して完全表現を作る。

完全表現の最終層数をKとすれば、第k層の基礎関数の数(または中間データ点の数)Nd_kと最下層(第1層)に入力される時間遅れデータ点の総数Mは、それぞれ次の通りになる。

$$\begin{aligned} Nd_k &= 2^{K-k+1} \\ M &= 2^{K+1} \end{aligned} \quad \dots\dots\dots(5.12)$$

いま、第k層のn番目の基礎関数の出力値(中間データ点の値)をz(k,n)、また第1層のn番目の基礎関数に入力される2つの遅れ時間データ点を便宜的にx(0,2n-1),x(0,2n)と記せば、第k層および第1層のn番目の基礎関数G_n^(k)は、それぞれ次の様に書ける。

$$\begin{aligned} z(k,n) &= G_n^{(k)}(z(k-1,2n-1), z(k-1,2n)) \quad \dots\dots\dots(5.13a) \\ &\quad (n=1, \dots, Nd_k, k=K, \dots, 2) \end{aligned}$$

$$z(1,n) = G_n^{(k)}(x(0,2n-1), x(0,2n)) \quad \dots\dots\dots(5.13b)$$

ここで、第1層の入力データ点x(0,2n-1),x(0,2n)は、それぞれ時刻t-t(2n-1),t-t(2n)に対応する遅れ時間データ点である。遅れ時間t(2n-1),t(2n)は基礎関数の番号nに対応して1からMまで機械的に数え上げたもので、実際には重複して何回も同じ遅れ時間が用いられる可能性がある。

さて、(5.13)式の表現を用いると、出力値x_tに対するフィッティングは最終層(第K層)の基礎関数によって次の様に書ける。

$$\begin{aligned} x_t &\sim z(K,1) \\ &= G_1^{(K)}(z(K-1,1), z(K-1,2)) \quad \dots\dots\dots(5.14) \end{aligned}$$

ここで、基礎関数に具体的な形を与えよう。データの平均値が0の場合を考えると、基礎関数G_n^(k)は2つの係数c₂,c₃だけで表される。ここでは、第k層のn番目の基礎関数の係数c₂,c₃をそれぞれc(k,2n-1),c(k,2n)と記すことにする。このとき、第k層のn番目の基礎関数の出力値(中間データ)は、下記の様に表現できる。

$$\begin{aligned} z(k,n) &= c(k,2n-1)z(k-1,2n-1) + c(k,2n)z(k-1,2n) \quad \dots\dots\dots(5.15) \\ &\quad (n=1, \dots, Nd_k, k=2, \dots, K) \end{aligned}$$

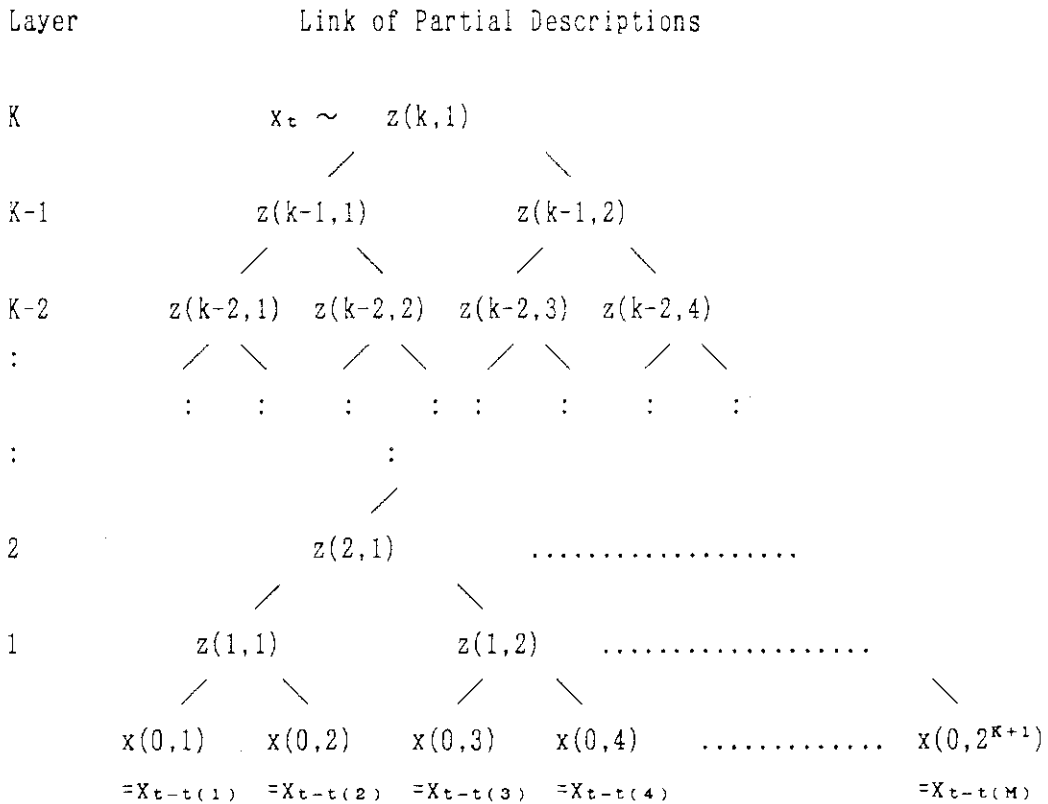


Fig. 5.1 Example of a complete description

線形GMDHの完全表現は見かけ上階層モデルとなっているが、(5.15)式からもわかる様に、非線形要素が無いために中間データは第1層に入力した時間遅れデータ点の線形結合で表わされる。各層毎の中間データの結合関係をFig.5.1に示す。完全表現の第1層にM個の時間遅れデータ点 $\{x_{t-t(n)}; n=1, \dots, M\}$ が入力され、順次結合を繰返して、最終層Kの出力値 $z(K,1)$ が作られる。

ここで、最下層の1番目の入力データ $x(0,1)$ が、最終層の基礎関数の値 $z(K,1)$ にどのように寄与しているかを見てみよう。Fig.5.1においては、データ $x(0,1)$ は、常に一番左側の結合経路を通して、最終層に達していることがわかる。一番左側の経路の各層における基礎関数は、(5.15)式では $n=1$ の場合に相当する。したがって、完全表現に対する1番目のデータ $x(0,1)$ の寄与の大きさ Γ_1 は、次の様に与えられる

$$\Gamma_1 = x(0,1) \prod_{k=1}^K c(k,1) \quad \dots\dots\dots (5.16)$$

ここで、係数 $c(k,1)$ の各層に渡る積を1つにまとめて $a(1)$ と記せば、上式は次の様に書き換えられる。

$$\begin{aligned} \Gamma_1 &= a(1)x(0,1) && \dots\dots\dots(5.17) \\ &= a(1)x_{t-t(1)} \end{aligned}$$

ここで、

$$a(1) = \prod_{k=1}^K c(k,1) \quad \dots\dots\dots(5.18)$$

他の入力データ点 $x(0,2), \dots$ についても同様の表現を得ることができる。いま、 m 番目の遅れ時間データ点の完全表現に対する寄与を Γ_m 、係数積を $a(m)$ とすると、次の一般的な表現が得られる。

$$\begin{aligned} \Gamma_m &= a(m)x(0,m) && \dots\dots\dots(5.19) \\ &= a(m)x_{t-t(m)} \end{aligned}$$

ここで、

$$a(m) = \prod_{k=1}^K c(k, 1 + \text{int}(\frac{m}{2^k})) \quad \dots\dots\dots(5.20)$$

ここで、 int は小数点以下切捨てるの整数の意味である。これらの表現を用いれば、(5.14)式の完全表現は次の様に見えることができる。

$$x_t \sim \sum_{m=1}^M \Gamma_m = \sum_{m=1}^M a(m)x_{t-t(m)} \quad \dots\dots\dots(5.21)$$

近似誤差を ε_t とおけば、上式は次の様になる。

$$x_t = \sum_{m=1}^M a(m) x_{t-t(m)} + \varepsilon_t \quad \dots\dots\dots(5.22)$$

さて、 M 個の時間遅れデータ点 $\{x_{t-t(m)}; m=1, \dots, M\}$ を時刻の順に並べ換え、また係数 $\{a(m)\}$ もこれに従って並べ換える。 M 個の遅れ時間 $\{t(m)\}$ の中で最大のものを M' と記せば、使用されている個々の時間遅れデータ点は $\{x_{t-1}, \dots, x_{t-M'}\}$ のいずれかに該当する。ここで、これらに対応する M' 個の係数 $\{a'(1), \dots, a'(M')\}$ を考えれば、時刻 $(t-i)$ に対応するデータ点 $x_{t-t(m)}$ の係数 $a(m)$ は、次の様に書ける。

$$a'(i) = a(m) \quad \dots\dots\dots(5.23a)$$

もし、ある1つの時刻に対応するデータ点が複数個重複して存在する場合には、これらに掛るもとの係数 $a(m)$ の和を計算して1まとめにする。例えば、時刻 $(t-i)$ に対して d 個のデータ点 $\{x_{t-t(m1)}, x_{t-t(m2)}, \dots, x_{t-t(md)}\}$ が存在している時、新しい係数 $a'(i)$ は次の様に書き換える。

$$a'(i) = a(m1)+a(m2)+ \dots +a(md) \quad \dots\dots\dots(5.23b)$$

さらに、データ $\{x_{t-1}, \dots, x_{t-m}\}$ の中で使用されていない点がある場合には、対応する係数は存在しないが、GMDHの場合にはその時間遅れ点はモデルに対する寄与が無い(小さい)として棄却されたことを意味している。したがって、それに対応した係数は、0と置けば良い。

$$a'(i) = 0 \quad \dots\dots\dots(5.23c)$$

このとき、(5.22)式は次式で表され、線形GMDHによる階層モデルの表現形式は、(5.1)式のARモデルと良い一致を見る。

$$x_t = \sum_{m=1}^{M'} a'(m) x_{t-m} + \varepsilon_t \quad \dots\dots\dots(5.24)$$

5.4 多変数フィードバック・システムの同定

本節では、フィードバック・システムをARモデルによって同定した場合に、ARモデルの係数がシステムの構造とどの様に関係付けられるかを復習し、次いでGMDHによって同定した場合、階層モデルのパラメータがシステムの構造をどの様に反映するかを考察する。

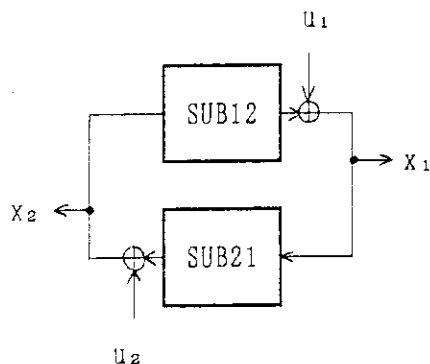


Fig.5.2 2-D feedback system

5.4.1 フィードバック・システム

まず、Fig.5.2に示す様な2つの変数 $\{x_1, x_2\}$ から成る2変数フィードバック・システムを考えよう。各変数には観測雑音または内部雑音 $\{u_1, u_2\}$ が付加されている。各変数間には、2つのサブ・システム $\{SUB12, SUB21\}$ がある。このシステムは線形で、各サブ・システムのインパルス応答関数 $\{h_{12,m}, h_{21,m} ; m=0, \dots, M\}$ が与えられているとすると、このシステムは下記の2つの連立方程式によって記述できる。

$$x_{1,t} = \sum_{m=1}^M h_{12,m} x_{2,t-m} + u_{1,t} \quad \dots\dots\dots(5.25a)$$

$$x_{2,t} = \sum_{m=1}^M h_{21,m} x_{1,t-m} + u_{2,t} \quad \dots\dots\dots(5.25b)$$

同様に、K個の変数 $\{x_i; i=1, \dots, K\}$ から成るK変数フィードバック・システムは、次のK元連立方程式で記述される。

$$x_{i,t} = \sum_{j=1}^K \sum_{m=1}^M h_{ij,m} x_{j,t-m} + u_{i,t} \quad \dots\dots\dots(5.26)$$

(i=1, \dots, K, j=1, \dots, K, j \neq i)

以上の様に、フィードバック・システムの数学的な記述は、1つ1つの方程式を取り出すとオープン・ループの多入力1出力システムという意味しか持たないが、出力変数が他の方程式では入力になるため、連立方程式という形を通じてフィードバック・システムを表現する様になっている。

5.4.2 ARモデルによる表現

次に、ARモデルを用いて上の多変数フィードバック・システムを同定する問題を考えよう。多変数ARモデルは、各出力変数毎に表現した下記のK元連立方程式で表される。

$$x_{i,t} = \sum_{j=1}^K \sum_{m=1}^M a_{ij,m} x_{j,t-m} + \varepsilon_{i,t} \quad \dots\dots\dots(5.27)$$

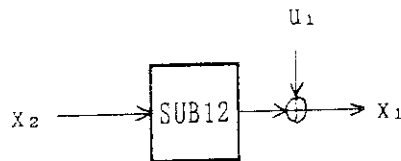
2変数フィードバック・システム(5.25)式に対応する2変数のARモデルは、次式で与えられる。

$$x_{1,t} = \sum_{m=1}^M a_{11,m} x_{1,t-m} + \sum_{m=1}^M a_{12,m} x_{2,t-m} + \varepsilon_{1,t} \quad \dots\dots\dots(5.28a)$$

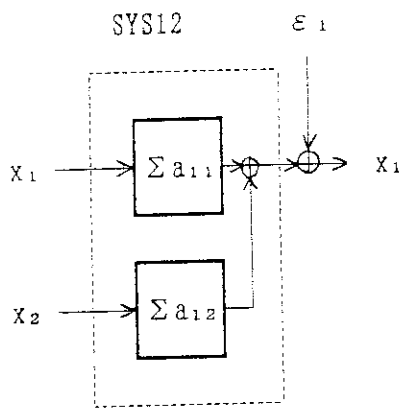
$$x_{2,t} = \sum_{m=1}^M a_{21,m} x_{1,t-m} + \sum_{m=1}^M a_{22,m} x_{2,t-m} + \varepsilon_{2,t} \quad \dots\dots\dots(5.28b)$$

ARモデルは、フィードバック・システムの数学的な記述と同様に、各出力変数毎に定義されたオープン・ループの多入力1出力システムの記述から成立っている。ただし、フィードバック・システムの記述の場合はシステムの一部を直接的に抜き出したために、各方程式を単独に取り扱おうと意味が違ってきてしまうのに対して、AR表現では、各方程式の内部にフィードバック・ループの意味を取込むことができ、単独で取り扱うことができる。

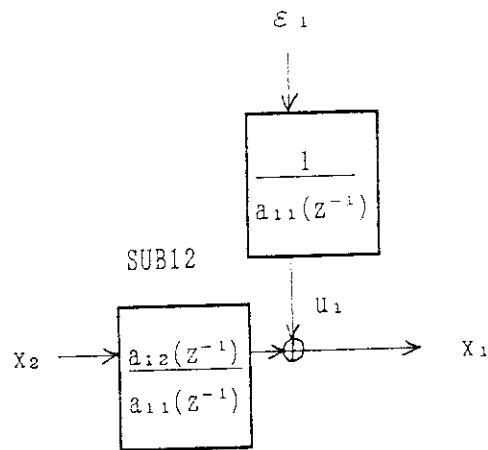
例えば、Fig.5.2のシステムにおける出力変数 x_1 についてのサブ・システムSUB12は、フィードバック・システムの数学的な記述では、Fig.5.3(a)の様に1入力1出力システムとなるのに対して、ARモデルでは、Fig.5.3(b)の様に出力変数を含む2入力1出力システムとなっている。Fig.5.3(c)は、(b)のフィードバック・システム中における等価表現である。同定に際して(a)の式を直接用いると、内部雑音 u_1 がフィードバック・ループで一巡する効果が表現できない。これに対して、(b)では、出力変数と同じものを入力し、白色過程 Σa_{11} を通じて正味の一巡成分を計算し、これを他の入力成分に付加している形を取っているのである。



(a) Mathematical expression



(b) AR model



(c) Equivalent expression

Fig.5.3 Block-diagrams of the sub-system with the output variable x_1 .

ARモデルによる同定は、アルゴリズムの都合で、全ての変数に関する係数列 $\{a_{1j}, a_{2j}\}$ と白色雑音源 $\{\varepsilon_1\}$ の分散値を、一度に一括して求めるのが一般的になっているが、上で議論した様に、(5.27)式は単独に用いてもさしつかえない。

多変数ARモデルの係数と分散値は、観測データから相関関数を計算し、次のYule-Walker方程式を解くことによって得られる。

$$R \cdot A = \Sigma \quad \dots\dots\dots(5.29)$$

ただし、

$$R = \begin{pmatrix} R_0 & R_1 & \dots & R_M \\ R_{-1} & R_0 & \dots & R_{M-1} \\ \vdots & \vdots & & \vdots \\ R_{-M} & R_{-M+1} & \dots & R_0 \end{pmatrix}, \quad A = \begin{pmatrix} -I \\ -{}^t A_1 \\ \vdots \\ -{}^t A_M \end{pmatrix}, \quad \Sigma = \begin{pmatrix} {}^t \Sigma_0 \\ O \\ \vdots \\ O \end{pmatrix} \quad \dots\dots\dots(5.30)$$

$$R_m = \begin{pmatrix} R_{11,m} & \dots & R_{1k,m} \\ \vdots & & \vdots \\ R_{k1,m} & \dots & R_{kk,m} \end{pmatrix}, \quad A_m = \begin{pmatrix} a_{11,m} & \dots & a_{1k,m} \\ \vdots & & \vdots \\ a_{k1,m} & \dots & a_{kk,m} \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} \sigma^2_{11} & 0 \\ & \diagdown \\ 0 & \sigma^2_{kk} \end{pmatrix} \quad \dots\dots\dots(5.31)$$

I は単位行列、O はゼロ行列

ここで、k番目の変数 x_k を出力とするサブシステムだけについて同定する場合を考えると、必要な係数と分散は、変数 x_k が出力となるもの、すなわち(5.31)式の係数、分散は第k行のみとなり、それらの解を得るためには、相関関数も第k行だけが必要ということになる。

したがって、 x_k のサブシステムの表現は、 R_m 、 A_m 、 Σ_0 の行列を下記のベクトルに修正したYule-Walker方程式によって与えられる。

$$\begin{aligned} R_m &= \{ R_{k1,m} \dots R_{kk,m} \} \\ A_m &= \{ a_{k1,m} \dots a_{kk,m} \} \\ \Sigma_0 &= \{ \sigma^2_{11} \ 0 \ \dots \ 0 \} \end{aligned} \quad \dots\dots\dots(5.33)$$

実際に、観測データに基づき相関行列を推定して、上の方程式を解く場合には、一般に各雑音源の十分な白色化が為されないため、上式の分散ベクトルの各要素は値を持つ。このため、分散ベクトルは、以下の様に定義しておく。

$$\Sigma'_0 = \{ \sigma^2_{k1} \ \dots \ \sigma^2_{kk} \} \quad \dots\dots\dots(5.34)$$

各要素が値を持つ場合は、各雑音源 $\{\varepsilon_i\}$ が互に独立でなくなり、付加雑音 u_i は、 ε_i だけでなく全ての雑音源 $\{\varepsilon_i\}$ からの寄与の和で表わされることになる。このため、Fig.5.3(b)とFig.5.3(c)の等価性も成立たなくなり、システムの雑音源やサブシステムの伝達関数の推定といった、システムの内部構造に立入った解析は不可能になる。しかし、この場合、同定の目的を出力変数 $\{x_i\}$ と付加雑音 $\{u_i\}$ から成るフィードバック・システムに関するオーバーオール解析だけに限定するならば、同定されたモデルは支障なく使用できる。

5.4.3 GMDHによる表現

GMDHによる多変数階層モデルで仮定されるシステムは、多入力1出力システムである。出力変数が入力変数と別であれば、オープン・ループ・システムの同定となるし、出力変数が入力変数の1つと同じに取れば、フィードバック・システムの同定となる。

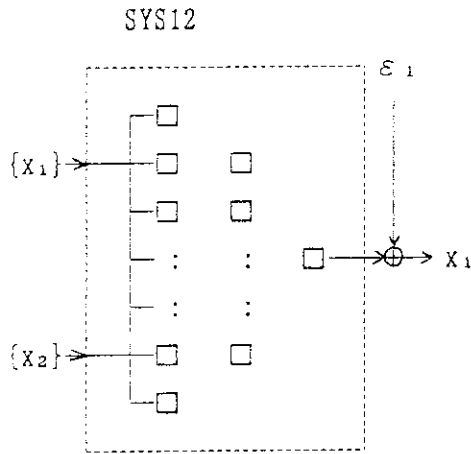


Fig.5.4 GMDH model of the subsystem-12 in the 2-D feedback system.

例えば、Fig.5.2の2変数フィードバック・システムを同定する場合、階層モデルの入力変数として $\{x_1, x_2\}$ を、出力変数に入力変数のうちの1つ $\{x_1\}$ を取れば、Fig.5.4に示すシステムが仮定されることとなり、ARモデルのサブシステムの表現(Fig.5.3.(b))と少なくとも外見上は全く同じになる。

すなわち、GMDHによる階層モデリングは、ARモデルのある1つの出力変数に対する記述と同様に、それ自身にフィードバック経路に関する表現を含んでいることが分る。

内部構造に関しては、ARモデルの場合、各係数は(5.28)式のように入力変数毎に明確に分離して定義されているため、インパルス応答関数として取り扱うことが可能で、強いては、雑音源 ε_t の独立性が満足できれば、各インパルス応答関数の表現(Fig.5.3(b))からサブシステムの伝達関数の表現(Fig.5.3(c))を得ることも可能であった。

これに対して、GMDHの階層モデルの場合、どの基礎関数 $\{x_1\}$ の白色過程でどの基礎関数 $\{x_2\}$ から $\{x_1\}$ へのインパルス応答に対応しているのかは、内部構造が複雑なために一見して分からない。しかし、線形GMDHにおいては、5.3節に述べた様に、各階層に配置される基礎関数の係数は各時間遅れデータ点毎にまとめることが可能であり、多変数モデルでは係数を変数毎にまとめることも可能である。したがって、線形GMDHによる階層モデルは、ARモデルと同様に、Fig.5.3(b)のインパルス応答関数による表現やFig.5.3(c)の伝達関数への変換も可能である。この場合、雑音源 ε_1 はARモデルと同様に白色雑音が仮定される。

ここで、GMDHによる階層モデルにおいて、非線形の基礎関数を用いた場合について考えてみよう。例えば、基礎関数にKolmogorov-Gaborの2次の多項式を用いたものでは、第1層の基礎関数は2つの入力変数 $\{x_1, x_2\}$ の時間遅れデータ点のいずれかが2つを入力データとするので、その係数は変数 $\{x_1\}$ あるいは $\{x_2\}$ の2乗までの項に掛るものの他に、2つの変数の積 $\{x_1 x_2\}$ に掛るものが現れる。これを階層構造によって結合した第K層の完全表現の係数は、原理的には各変数のデータ点のべき乗の項 $\{x_1^n, x_2^n\}$ ($n=1, \dots, K$)に掛るものと2つの変数の高次項の積 $\{x_1^n x_2^{K-n}\}$ ($n=0, \dots, K$)に掛るものの3種に分類される。

$$\begin{aligned}
 x_{1,t} = & \sum_{m=1}^M \sum_{n=1}^K a_{11}(m,n) (x_{1,t-m})^n + \sum_{m=1}^M \sum_{n=1}^K a_{12}(m,n) (x_{2,t-m})^n \\
 & + \sum_{m=1}^M \sum_{n=1}^{K-1} a_{112}(m,n) (x_{1,t-m})^n (x_{2,t-j})^{K-n} \dots\dots\dots (5.35)
 \end{aligned}$$

もし、同定の結果、第3番目の2つの変数にまたがる高次積が完全表現に含まれなかった場合には、第1番目の変数{x₁}の時間遅れデータ点のべき乗に掛る係数は、変数{x₁}の非線形白色過程{h₁₁}を意味し、また変数{x₂}の時間遅れデータ点のべき乗に掛る係数は、変数{x₂}から変数{x₁}への非線形インパルス応答{h₁₂}を意味するものと解釈することができよう。この場合には、同定したモデルの内部構造は変数毎に分離できる。

しかし、もし第3番目の2つの変数にまたがる高次積が完全表現に含まれた形で同定された場合には、モデルの解釈は容易ではない。

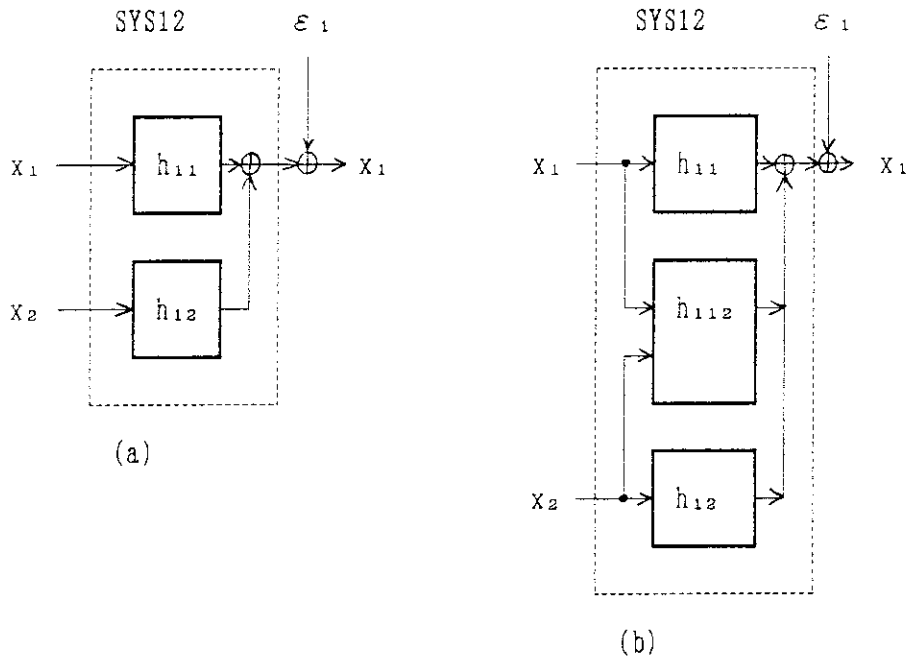


Fig.5.5 Non-linear GMDH model

5.5 スペクトル表現

ARモデルによる信号{x_t}のパワー・スペクトルS_x(f)は、次式で与えられる。

$$S_x(f) = \frac{\Delta t \sigma^2}{|1 - a(f)|^2} \dots\dots\dots (5.36)$$

ここで、

$$a(f) = \sum_{m=1}^M a_m \exp(-j2\pi f m) \quad \dots\dots\dots(5.37)$$

$$(f=0, \frac{1}{2\Delta t F}, \frac{2}{2\Delta t F}, \dots, \frac{F}{2\Delta t F})$$

Δt は信号 $\{x_t\}$ のサンプリング時間間隔、 σ^2 は残差系列(予測誤差)の分散値、分母の $a(f)$ はAR係数 $\{a_m\}$ のフーリエ変換で、また、 F は周波数領域の分割数を定める整数である。この場合、それぞれの係数 $a(m)$ は時間的に Δt 毎に連続で、(5.26)式の関数は連続である。

一方、GMDHの係数は一定時間間隔 Δt 毎には並んでいないが、係数を(5.21)式の様に変換することにより、ARモデルと同様の演算が可能である。

6. フィッティングの実例

本章では、幾つかのシミュレーション・データを用いて、GMDHに基づく階層モデルとARモデルのフィッティング比較を行う。フィッティングの結果は、階層数、モデルに取り込まれた時間遅れ点とモデル次数、係数の値、予測誤差等について考察された。

6.1 フィッティング・データ

フィッティングに用いたデータは、1変数および2変数のシミュレーション・データである。以下にその概要を説明する。

6.1.1 1変数シミュレーション・データ

1変数シミュレーション・データは、1変数ARモデルの各種フィッティング・アルゴリズムの同定比較用に作成されたもの^{38,39)}で、正弦波または正弦波に白色雑音を重畳したアナログ信号(6.1式)をサンプルした3種類の時系列データcase-I~IIIを用いた。

$$x(t) = A \sin(2\pi ft) + Bw(t) \quad \dots\dots\dots(6.1)$$

アナログ信号のパラメータA, B, f、サンプリング・パラメータ Δt , NをTable 6.1-1に示す。Case-Iは正弦波のみ、Case-IIは白色雑音を付加したもの、Case-IIIはそれを10倍早くサンプリングしたものとなっている。白色雑音(DC~600Hz)の最大振幅は、正弦波の振幅値の1/10程度になっている。

Table 6.1-1 Specifications of univariate simulation data.

	A	f	B	Δt	N	$n(1/2\pi f)$
case-I	2.5	1.5 Hz	0	0.061 sec	200	11
case-II	2.5	1.5 Hz	1	0.061 sec	4000	11
case-III	2.5	1.5 Hz	1	0.0061 sec	4000	111

6.1.2 2変数シミュレーション・データ

2変数シミュレーション・データは、多変数ARモデルのフィッティング性能を評価するために作成されたもの⁴⁰⁾で、ARモデルのシステム構造をアナログ的に模擬した2変数フィードバック・システムの出力信号からサンプルした時系列データである。システムは2つの白色雑音源と4つの1次または2次遅れ伝達関数から成立ち、インパルス応答関数を{ h_{ji} }と表せば、2つの出力変数は次式で表される。

$$x_j(t) = \sum_{i=1}^2 \int_0^{+\infty} h_{ji}(\tau_i)x_i(t-\tau_i)dt + w_j(t) \quad \dots\dots\dots (6.2)$$

(j=1,2)

4つのサブシステム{ $h_{11}, h_{12}, h_{21}, h_{22}$ }のゲインはほぼ1であり、また遮断周波数 f_c はTable 6.1-2に示す値に設定されている。{ h_{12} }だけが2次遅れである。この2つの信号{ x_1, x_2 }をTable 6.1-2に示す時間間隔でサンプルしたものがフィッティング試験に用いられた。サンプリングに当っては、各信号とも遮断周波数40Hz(115dB/oct)のローパス・フィルタに通されている。

Table 6.1-2 Specifications of 2-dimensional simulation data.

	fc				Δt	N
	h_{11}	h_{12}	h_{21}	h_{22}		
MAR007	10Hz	20Hz	15Hz	30Hz	0.0125 sec	4096

6.2 モデルのフィッティング

GMDHに基づく階層モデルについては、前節までに説明したアルゴリズムによってフィッティングを行った。基礎関数はKolmogorov-Gaborの多項式のうち、1次および2次の2通りを採用した。

ARモデルは、1変数モデルについてはMarpleの最小2乗法に基づくアルゴリズム³⁹⁾を用い、また2変数については赤池らのTIMSACと同じ多変数Yuie-Walker法を用いた。

GMDHに基づく階層モデルの階層数の判定は予測誤差の2乗平均値に基づく判定により行っているが、さらにARモデルとの比較の意味で、モデルのAICを計算した。

AICは尤度関数 $L(\theta)$ が最大、すなわち、確率密度関数が最も確からしい形になる様にパラメータを決定するための評価基準である。モデルの近似尤度は、一般に

$$L(y;n,k,z_1,\dots,z_k,\sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^{-n}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y - \sum_{k=1}^K z_k x_{k,i}\right)^2\right) \dots(6.3)$$

対数尤度は、

$$l(y;n,k,z_1,\dots,z_k,\sigma^2) = -\frac{1}{2} (n \log(2\pi\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^n \left(y - \sum_{k=1}^K z_k x_{k,i}\right)^2) \dots(6.4)$$

ここで、

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \left(y - \sum_{k=1}^K z_k x_{k,i}\right)^2 \dots\dots\dots(6.5)$$

とおけば、

$$\begin{aligned} l(y;n,k) &= -\frac{1}{2} (n \log(2\pi\sigma^2) + n) \\ &= -\frac{n}{2} (1 + \log 2\pi) - \frac{1}{2} n \log \sigma^2 \dots\dots\dots(6.6) \end{aligned}$$

$$\begin{aligned} AIC &= -2(\text{尤度}) + 2(\text{パラメータ数}) \\ &= n \log \sigma^2 + 2(k+1) + n(1 + \log 2\pi) \dots\dots\dots(6.7) \end{aligned}$$

実際の計算には、GMDHに基づく階層モデルのフィッティングでは、N個のデータをNt個のトレーニング用データとNc個のチェック用データに分割して、完全表現を構成するNd個の基礎関数のp個の係数(全部でpNd個)と完全表現の予測誤差の2乗平均値を決定するため、パラメータの数は(pNd+1)となる。しかし、線形GMDHはARモデルと等価であり、係数を遅れ時間点でまとめてしまえば、パラメータの数は(M'+1)である。これをAICの計算に用いた。

6.3 1点予測

フィッティングによってモデルのパラメータが得られると、これを用いて過去から現在までの観測データに基づいて1点未来の予測ができる。観測値と予測値の誤差、すなわち予測誤差は、フィッティングされたモデルの善し悪しを判別する重要な指標である。そこで、本章ではこれを観測データに対して計算し、グラフにプロットし、直感的に判断するようにした。

ARモデルの1点予測は、モデルの次数をM、係数を{a_m; m=1,...,M}とするとき、次式で与えられる。

$$\tilde{X}_t = \sum_{m=1}^M a_m X_{t-m} \quad \dots\dots\dots(6.8)$$

一方、GMDHによる階層モデルの1点予測は、GMDHの完全表現を用いて計算される。完全表現は、階層構造を持つため1度に表現しようとする多量複雑になるが、第k層のi-番目の基礎関数を $G_i^{(k)}$ 、その入力となる2つの中間変数を $Z_{i1}^{(k)}$ 、 $Z_{i2}^{(k)}$ とし、最大層数をKとすると、次式で与えられる。

$$\begin{aligned} \tilde{X}_t &= G_1^{(K)}(Z_{11}^{(K)}, Z_{12}^{(K)}) \\ &= G_1^{(K)}(G_1^{(K-1)}(Z_{11}^{(K-1)}, Z_{12}^{(K-1)}), G_2^{(K-1)}(Z_{21}^{(K-1)}, Z_{22}^{(K-1)})) \\ &\vdots \end{aligned} \quad \dots\dots\dots(6.9)$$

ここで、 $k=K, K-1, \dots, 2$ のときは、

$$\begin{aligned} Z_{i,1}^{(k)} &= G_{2i-1}^{(k-1)}(Z_{2i-1,1}^{(k-1)}, Z_{2i-1,2}^{(k-1)}) \\ Z_{i,2}^{(k)} &= G_{2i}^{(k-1)}(Z_{2i,1}^{(k-1)}, Z_{2i,2}^{(k-1)}) \\ i &= 1, \dots, 2^{k-k} \end{aligned} \quad \dots\dots\dots(6.10)$$

また、 $k=1$ のときは、

$$\begin{aligned} Z_{i,1}^{(1)} &= X_{t-m(2i-1)} \\ Z_{i,2}^{(1)} &= X_{t-m(2i)} \end{aligned} \quad \dots\dots\dots(6.11)$$

ただし、 $m(2i-1), m(2i)$ はそれぞれ任意の時間遅れ点を表す

両モデルの予測誤差(残差)は、ともに、観測値と計算値との差

$$\varepsilon_t = X_t - \tilde{X}_t \quad \dots\dots\dots(6.12)$$

で計算される。

6.4 1変数モデルのフィッティング

GMDHによる階層モデルおよび通常のARモデルの両フィッティングは、Case-I ~ IIIのデータに対して、まず最大遅れ時間 $M_{max}=32$ に設定して行い、さらにCase-IIIについては $M_{max}=64$ の場合も行った。したがって、全部で4通りのフィッティングを行った。Table 6.2に、これらのフィッティングで得られた階層モデルの層数Kと最大遅れ点 M' 、ARモデルの次数M、ならびに両モデルの残差共分散 σ^2 の値を示す。

残差共分散(予測誤差の2乗平均値) σ^2 の値に注目すると、Case-IについてはGMDHモデルはARモデルに比べ小さく、約半分の値となっている。一方Case-II ~ IIIについてはGMDHモデルはARに比べて若干大きい。モデル次数Mまたは M' を見ると、Case-IIは両モデルとも似通った値を

持っているが、Case-IあるいはIIIは異なる。これらの理由を調べるために、さらに詳細に検討してみよう。

Table 6.2 Model order and innovation variance.

	Mmax	GMDH			AR	
		K	M'	σ^2	M	σ^2
case-I	32	1	27	3.6E-4	8	6.1E-4
case-II	32	5	27	9.1E-2	29	8.8E-2
case-III	32	5	32	9.6E-2	32	8.5E-2
	64	5	63	8.6E-2	47	8.1E-2

Table.6.3-1~6.3-4には、各ケースのGMDHモデルのパラメータを層毎にまとめたもので、基礎関数の最大数Nmaxと採用された遅れ時間点に対応する基礎関数の数Nd、残差共分散 σ^2 、モデルのAIC値、最大遅れの設定値Mmaxと実際に採用された遅れ点の最大値M'、そして採用された時間遅れ点が表示されている。

(1) Case-Iの結果

Case-I (Table 6.3-1)では、各層毎のモデルに採用された時間遅れ点は1~27に分布しているが、データ中の正弦波の周期が11点であることを考えると、時間遅れ点はいずれも正弦波の+側頂点付近(1,11,22)、あるいは-側頂点付近(6,16,26)のものが採用されているのが分かる。

階層構造は、予測誤差の判定によって第5層までフィッティングされたが、パラメータ数と共分散値から計算されたAICによれば、第1層の結果が最適モデルとなっている。第1層のモデルは、わずか2点を用いたモデルであるが、遅れ時間は正弦波の両頂点(11,27)を的確に採用し、共分散値も十分小さくなっている。

Table 6.3-1 Model parameters obtained for Case-I

K	Nd/Nmax	σ^2	AIC	M'/Max	Time-lag-points
1	2/2	3.6427E-5	-2017.6	27/32	11,27
2	4/4	3.4435E-5	-2004.2	27/32	1,10,11,27
3	5/8	3.4319E-5	-1992.5	27/32	6,10,11,22,27
4	7/16	3.3302E-5	-1973.8	27/32	6,10,11,16,20,26,27
5	7/32	3.3046E-5	-1975.3	31/32	6,10,11,16,20,26,27

ARモデルのフィッティングにおいては、最大遅れ点 $M_{max}=32$ の設定に対して、16次以降のフィッティング計算は数値上の条件悪化から失敗に終わった。得られた1~16次の範囲内でAIC最小モデル($M=8$)は、階層モデルに比べてパラメータの多い割に共分散値は2倍も大きく、階層モデルより劣る。

Fig.6.1にデータの時刻歴波形グラフと、それに対応する両モデルの予測値系列、予測誤差系列を示す。

(2) Case-IIの結果

Case-II (Table 6.3-2)では、データが正弦波に白色雑音を重畳したものであるため、モデル構造も複雑化しAICによる最適モデルも第5層となった。採用された時間遅れ点はCase-Iとほとんど似通ったものとなった。階層を進める程近似の度合いは良くなる。これはS/Nの関係であり、採用する時間遅れ点が多くなる程白色雑音の影響が低減されることを示している。

Fig.6.2にデータの時刻歴波形グラフと、それに対応する両モデルの予測値系列、予測誤差系列を示す。両モデルともあまり差は無い。予測誤差は白色雑音というよりは少し系統的な変動が混じった感じだが、これはサンプリング時間間隔が粗い(1周期がほぼ11点)ためであろう。

Table 6.3-2 Model parameters obtained for Case-II

K	Nd/Nmax	σ^2	AIC	M'/Max	Time-lag-points
1	2/2	1.1624E-1	-2193.4	17/32	16,17
2	4/4	9.8685E-2	-2352.1	27/32	16,17,22,27
3	6/8	9.3026E-2	-2403.6	27/32	10,12,16,17,22,27
4	8/16	9.1721E-2	-2409.2	27/32	5,6,10,12,16,17,22,27
5	9/32	9.1096E-2	<u>-2411.8</u>	27/32	5,6,7,10,12,16,17,22,27

(3) Case-IIIの結果

Case-III (Table 6.3-3, Table 6.3-4)は、Case-IIのデータを10倍細かくサンプルしたもので、予測誤差の変動も白色に近いものとなっている。 $M_{max}=32$ では両者とも次数は32となった。正弦波の周期は111(>32)点のため、常識的には無理なフィッティングである。 $M_{max}=64$ とした場合には、一応ARモデルのMAICE次数は47となったが、これでもまだ足りない。GMDHによる階層モデルではいきなり50~60点付近の遅れ時間を採用している。おそらく $M_{max}=120$ 位にとればもっと的確なモデルができたであろう。

Fig.6.3にデータの時刻歴波形グラフと、それに対応する両モデルの予測値系列、予測誤差系列を示す。注意すべきは予測波形の頂点付近の傾向である。ARモデルによるものでは白色雑音成分がまだ残っているのが見えるが、GMDHではそれが少し低減され正弦波に近づいているのが分かる。予測誤差の点から見るとGMDHの方が若干大きい共分散値を持つのは、ARモデルの方

が重畳する白色雑音成分もモデルに取り入れているからである。

Table 6.3-3 Model parameters obtained for Case-III (Max ≤ 32)

K	Nd/Nmax	σ^2	AIC	M'/Max	Time-lag-points
1	2/2	1.5277E-1	-1914.2	32/32	2,32
2	3/4	1.1594E-1	-2191.9	32/32	1,2,32
3	5/8	1.0133E-1	-2320.8	32/32	1,2,3,6,32
4	7/16	9.7334E-2	-2353.2	32/32	1,2,3,4,5,6,32
5	8/32	9.5844E-2	<u>-2364.6</u>	32/32	1,2,3,4,5,6,23,32

Table 6.3-4 Model parameters obtained for Case-III (Max ≤ 64)

K	Nd/Nmax	σ^2	AIC	M'/Max	Time-lag-points
1	2/2	1.2155E-1	-2147.8	54/64	2,54
2	4/4	9.8199E-2	-2357.2	59/64	51,54,58,59
3	8/8	8.8874E-2	<u>-2441.3</u>	61/64	48,51,54,55,56,58,59,61
4	11/16	8.6765E-2	-2430.3	63/64	48,51,53,54,55,56,57,58,59,61,63
5	12/32	8.6482E-2	<u>-2451.2</u>	63/64	48,50,51,53~59,61,63

(4) ARモデル換算係数

また、Table 6.4-1~6.4-4には、各ケースのGMDHモデルの階層数毎の完全表現をAR係数に換算した結果が示されている。

Case-Iでは、第1層~第4層までは全ての係数は1.0以下であるが、第5層では1.0以上の値を持つ様になる。反射係数をM'=27の係数と考えれば、これは常に1.0以下の値を取っている。

Case-IIでは、Case-Iと同様に第1層~第4層までは全ての係数の値は1.0以下であり、第5層の係数は1.0を大きく越えたものが目立つ。反射係数は第1層ではM'=17、第2層以降M'=27とすれば第4層までは1.0以下であるが、第5層で1.0を大きく越える。これはモデルが不適格であると考えられることもできる。

Case-III (Mmax=32)では、反射係数(M'=32)は第2層で1.0を越え、以降増大する傾向にある。もう一方のフィッティング(Mmax=64)の結果は、反射係数はM'=54,59,63であり第5層まで1.0を越えることは無かった。

以上の結果から、最大遅れ時間Mmaxの初期設定は重要であることが分かる。また、階層数が5以上では階層演算の精度が極端に悪化していくことが伺われる。

Table 6.4-1 Equivalent AR coefficients obtained for Case-I

M'	K= 1	2	3	4	5
0	0.06822	-0.00000	-0.00000	-0.00000	-0.00000
1		0.04807			
6			-0.00857	-0.98991	3.2789
10		0.51269	0.05267	-0.10676	-0.91273
11	-0.30599	-0.02394	0.04987	0.08647	0.51794
16				0.67957	6.1793
20				0.32453	2.4134
22			-0.10071		
26				0.71185	4.4938
27	0.69828	0.64365	0.71591	0.14651	-0.23373

Table 6.4-2 Equivalent AR coefficients obtained for Case-II

M'	K= 1	2	3	4	5
0	0.21699	-0.00047	-0.00011	-0.00002	-0.00001
5				-0.01361	12.711
6				-0.44516	-16.161
7					23.658
10			0.24838	0.30074	-15.155
12			-0.34240	0.28218	18.023
16	-0.38355	-0.20750	0.03321	-0.27777	10.297
17	-0.64092	-0.29724	-0.25606	-0.41738	0.34167
22		-0.27822	-0.13250	-0.01194	-8.4194
27		0.22337	-0.04908	-0.17190	-19.930

Table 6.4-3 Equivalent AR coefficients obtained for Case-III ($M_{\max} \leq 32$)

M'	K= 1	2	3	4	5
0	0.01084	-0.00248	-0.00116	-0.00031	-0.00002
1		-0.01999	-0.01033	-0.02520	0.19066
2	-0.08636	-0.04216	-0.01900	-0.02867	0.19875
3			-0.01774	-0.00619	0.35164
4				-0.00842	0.46204
5				-0.14808	0.61535
6			-0.13982	-0.17037	0.68377
23					3.1682
32	0.94985	1.0000	1.1557	1.4621	20.546

Table 6.4-4 Equivalent AR coefficients obtained for Case-III ($M_{\max} \leq 64$)

M'	K= 1	2	3	4	5
0	0.09172	-0.00045	0.00019	-0.00005	-0.00001
2	-0.49991				
48			-0.27581	-0.29384	-0.76755
50					-0.36453
51		-0.26002	-0.05548	-0.01338	-0.38790
53				-0.25785	-0.30481
54	0.49532	-0.23139	-0.11242	-0.11331	-0.45457
55			-0.09804	-0.21978	-0.30732
56			-0.15526	-0.27812	-0.18152
57				-0.32476	-0.05283
58		-0.26807	-0.13838	-0.08007	-0.33253
59		-0.25110	-0.07320	-0.01755	-0.34660
61				-0.25834	-0.16950
63			-0.25539	-0.01265	-0.19604

6.5 2変数モデルのフィッティング結果

2変数シミュレーション・データに対して、階層モデルは最大遅れ時間をMmax=64に設定し、 $\{x_2 \rightarrow x_1\}$ の方向についてだけフィッティングを行った。多変数ARモデルの場合は、1度の計算で $\{x_2 \rightarrow x_1\}$ と $\{x_1 \rightarrow x_2\}$ の両方向のサブシステムがフィッティングできるが、ここではその片側だけを比較する。

多変数ARモデルでは2つの雑音源相互間の相関を記述する残差共分散 σ_{12}^2 も算出されるが、多変数階層モデルはアルゴリズム上 σ_{11}^2 と σ_{12}^2 は分離できず、これを1まとめた σ_1^2 で記述される。本来2つの雑音源が独立ならば、 $\sigma_{12}^2=0$ となり、 σ_{11}^2 と σ_1^2 は同じ意味を持つはずである。

Fig. 6.4に出力変数 $\{x_1\}$ のデータの時刻歴波形グラフと、それに対応する階層モデルの予測値系列、予測誤差系列を示す。ARモデルによる結果もこれとほとんど同じである。

Table 6.5に、得られた階層モデルの層数Kと最大遅れ点M'、ARモデルの次数M、ならびに両モデルの残差共分散 σ^2 の値を示す。さて、結果を見ると、残差共分散の値は階層モデルの方が約2桁程小さいことが分かる。ただし、ARモデルの次数は2であるのに対して階層モデルの層数は5と大きい。そこで、階層モデルの結果をもう少し詳細に見てみよう。Table 6.6に階層モデルのパラメータを層毎に示す。

Table 6.5 Model order and innovation variance.

Data file	Mmax	2-D GMDH			MAR			
		K	M'	σ_1^2	M	σ_{11}^2	σ_{22}^2	σ_{12}^2
MAR007	64	5	8	5.18E-6	2	2.92E-4	6.90E-4	2.51E-4

Table 6.6 Model parameters obtained for 2-dimensional data

K	Nd/Nmax	σ^2	AIC	M'/Max	Time-lag-points	
					X ₁	X ₂
1	2/2	5.5735E-6	-23221.2	2/64	-	1,2
2	4/4	5.1992E-6	-23350.7	2/64	1,2	1,2
3	5/8	5.1910E-6	-23351.7	4/64	1,2,4	1,2
4	6/16	5.1849E-6	-23351.9	4/64	1,2,3,4	1,2
5	7/32	5.1825E-6	-23350.8	8/64	1,2,3,4,8	1,2

はじめに気付く事は、第1層から第5層まで共通して、取り込んだデータ点がいずれも時間遅れが小さい(1,2点付近)に集中している事である。これはもともとのフィードバック・システムの雑音源が白色雑音であり、サブシステムが1次あるいは2次遅れ系から構成されているために、信号自体が強い白色性を持っているからであると考えられる。

また別の特徴として、第1層のモデルは、変数 $\{x_2\}$ の方からだけデータ点が取り込まれており、 $\{x_1\}$ からの寄与が全く無い単なる1入力1出力モデルとなっている事に気付く。層数を増やしていくと、第2層以降 $\{x_1\}$ からの寄与も現れ、フィードバック系を表現する様になる。第5層まで進めた場合、 $\{x_1\}$ からのデータ点取り込みは次第に遠くの遅れ時間点に向っていくのに対して、 $\{x_2\}$ からの寄与は始めの2点から増えない事が分かる。この事は、システムの構造に関係しているものと思われるが、はっきりした事は分からない。

AICを計算した結果、5つのモデルの中では第4層のものが最小となった。Table 6.7に各層毎の換算AR係数を示しておく。第4層の結果は、 $\{x_1\}$ に直接結び付いた雑音源の有色過程に関しては4次、変数間のサブシステム $\{x_2 \rightarrow x_1\}$ に関しては2次のARモデルが得られたことになる。

Table 6.7 Equivalent AR coefficients obtained for 2-dimensional data

M'	K= 1	2	3	4	5
0	-0.00286	-0.00014	-0.00001	-0.00000	-0.00000
1		0.11900	0.19721	0.13254	-0.64665
2		0.42179	0.04923	0.72286	1.5906
x_1 3				0.44362	-0.43515
4			0.44709	-0.02974	-0.00099
8					-0.45198
x_2 1	0.29647	0.09035	0.07415	0.20752	0.55166
2	0.64622	0.45001	0.39465	0.32671	2.3706

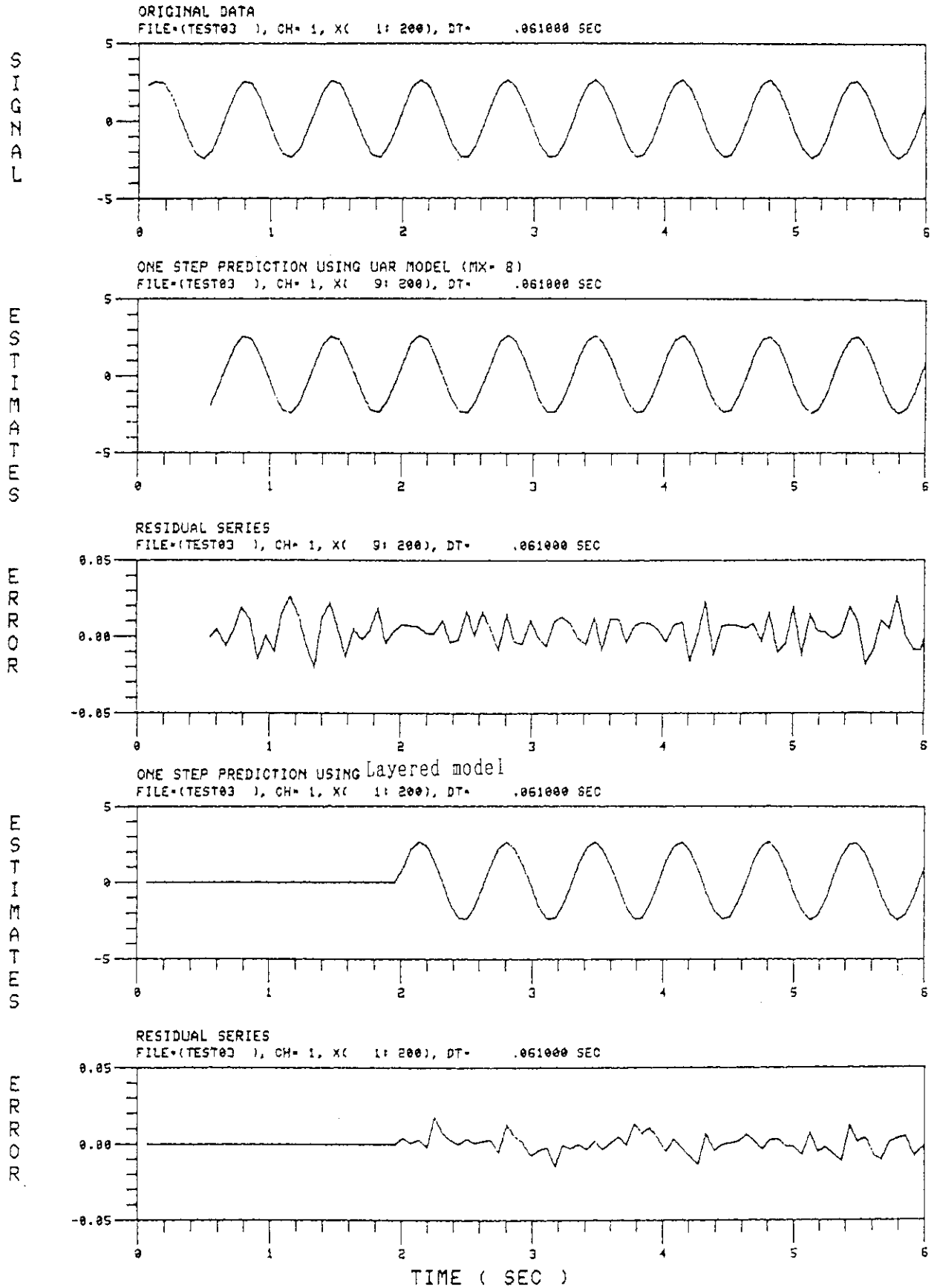


Fig.6.1 One step predictions and its errors obtained by UAR model and the layered model (Case-I)

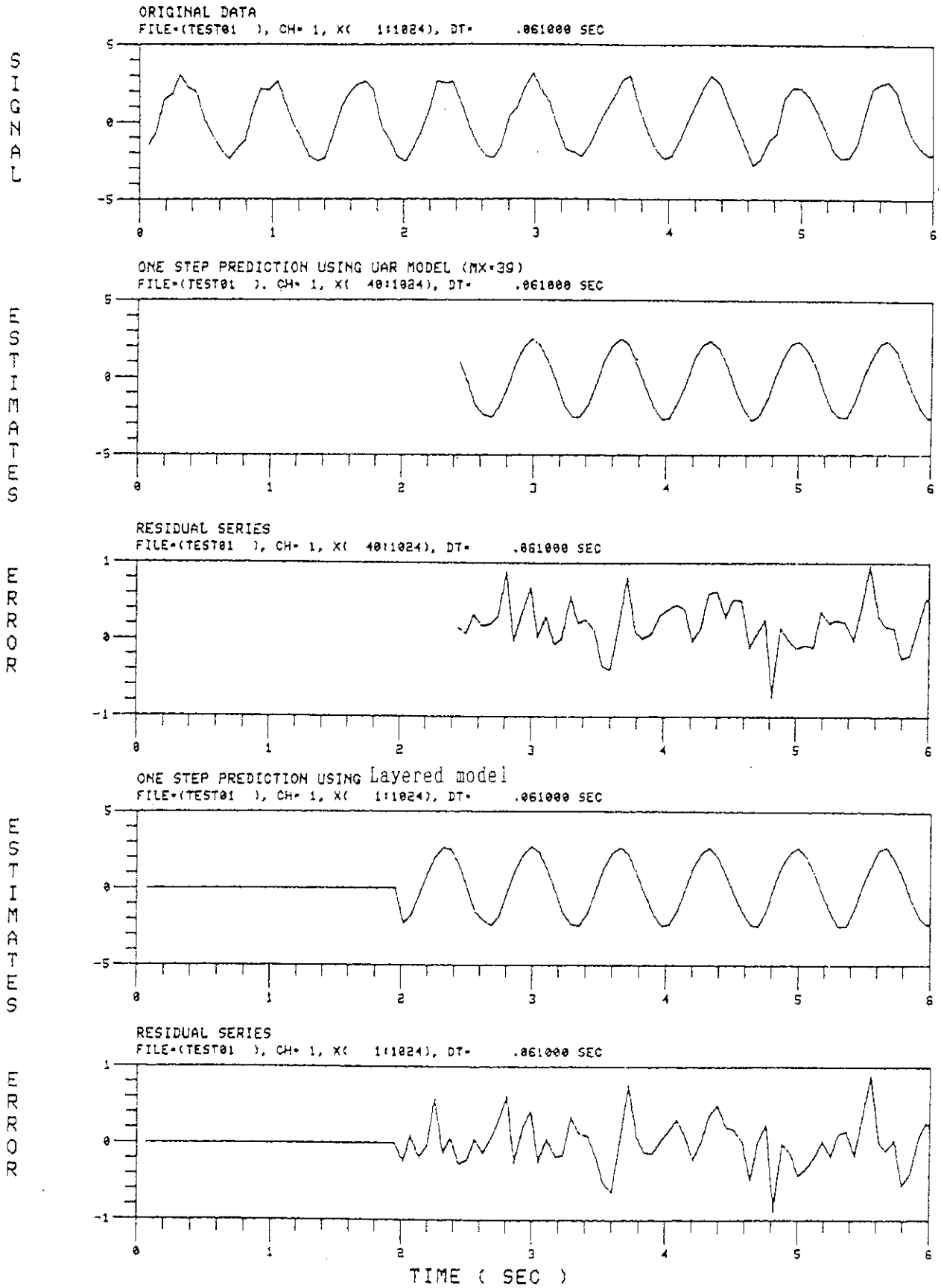


Fig.6.2 One step predictions and its errors obtained by UAR model and the layered model (Case-II)

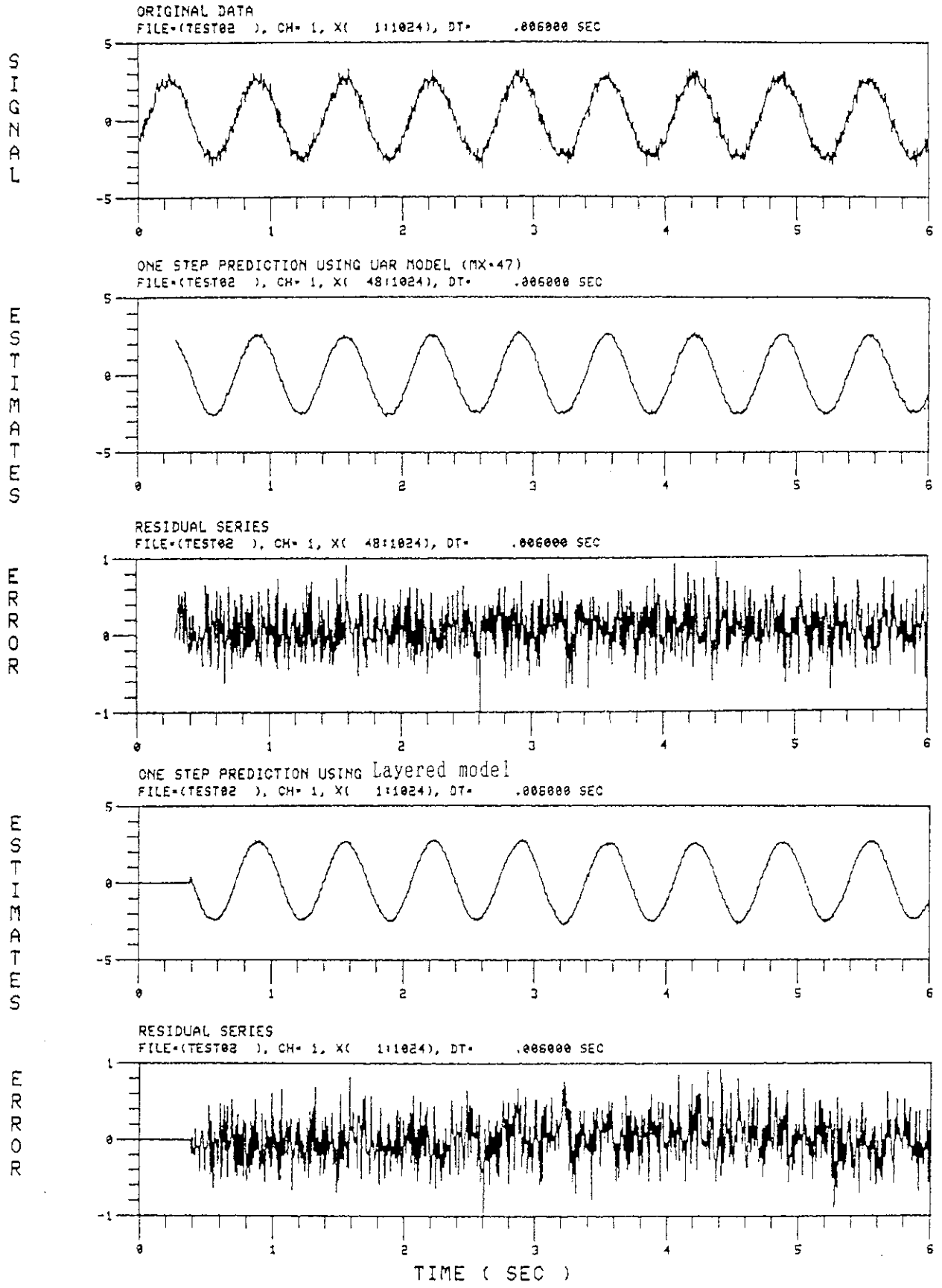


Fig.6.3 One step predictions and its errors obtained by UAR model and the layered model (Case-III)

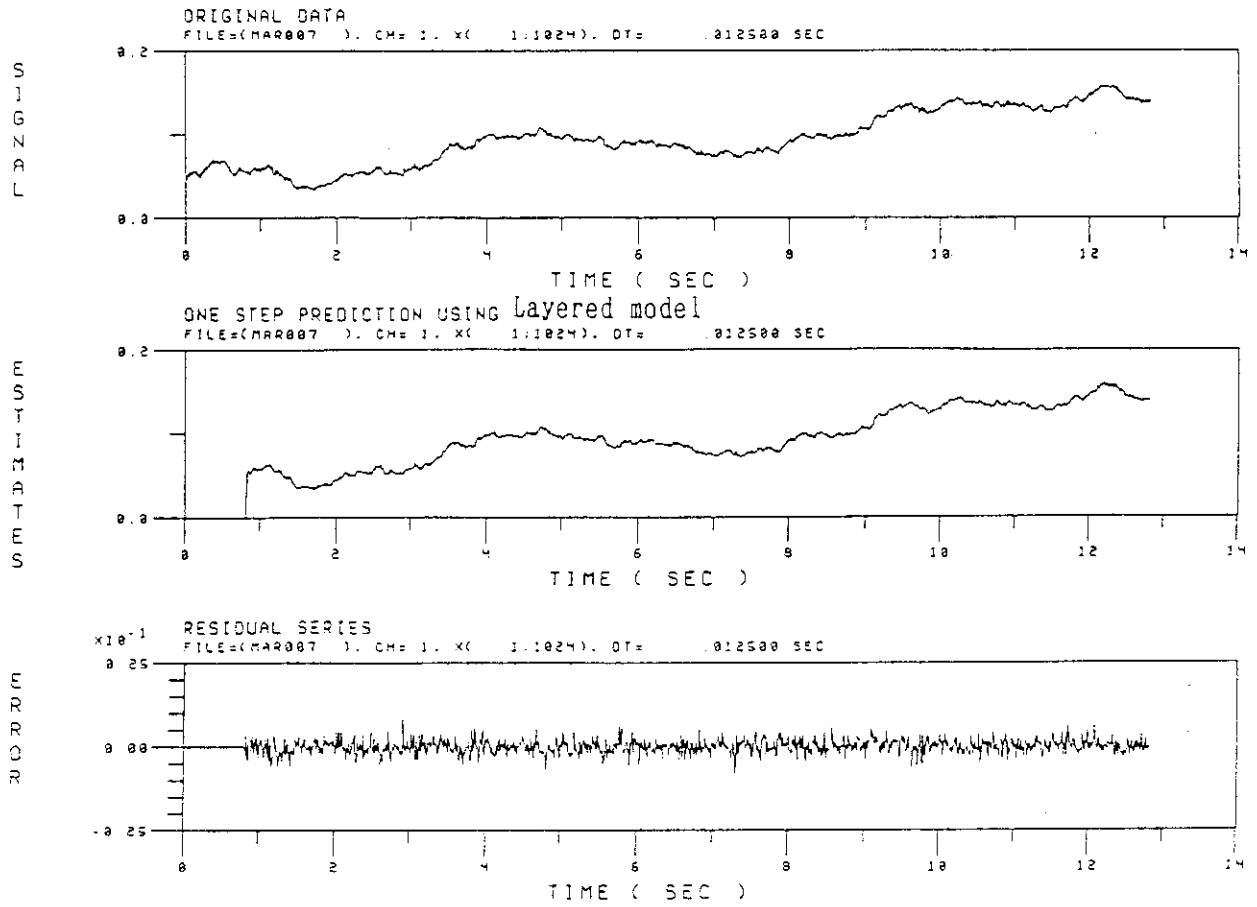


Fig.6.4 One step predictions and its errors obtained by the multivariate layered model (2-dimensional data)

7. おわりに

炉雑音信号の統計的解析に基づく原子炉システムの同定や異常予知診断の分野においては、非線形現象や非定常信号の解析が重要な課題になりつつあり、目下これらを一手に解決する有力な解析手法が待望されている。

本論文では、GMDHに基づく階層型モデリングをその最も有力な手法と位置付け、まず始めにGMDHの原理ならびにGMDHの基礎理論、すなわちシステムの並列化表現、階層構造、Kolmogorov-Gabor多項式に基づく基礎関数の導出、階層モデルのフィッティング方法、評価関数(発見的規範)について概説した。また、線形および非線形階層モデルのフィッティングを具体化するための計算アルゴリズムを処理段階毎に詳細に示した。アルゴリズムは1入力1出力オープン・ループ型モデルの他に、1変数型(自己回帰型)、多変数フィードバック型モデルへの拡張についても述べた。

次に、統計的信号処理の理論の立場から、Kolmogorov-Gabor型の基礎関数の係数を与える正規方程式を考察した。この結果、係数の推定式は高次の相関関数から成立っている事、すなわち、基礎関数が1次の場合には平均値(1次モーメント)と通常の相関関数(2次モーメント)から、また2次の場合にはそれに加えて2次および3次の高次相関関数(～4次モーメント)から成立っている事が明らかになった。

さらに、線形階層モデルとARモデルの関係を理論的に考察した。その結果、線形階層モデルはARモデルと等価な係数を持つこと、また白色雑音源を仮定することにより、ARモデルと同様に多変数フィードバック・システムの同定とその分解が可能であること等が明らかになった。

最後に、幾つかの線形シミュレーション・データを用いて、線形階層モデルのフィッティング試験を行い、係数値や予測誤差に関してARモデルと比較した。その結果、信号の構造が簡単な場合には、線形階層モデルはARモデルに比較してコンパクトでしかも十分実用的なモデルの得られることが確認できた。

参考文献

- 1) E.Turkcan, M.Kitamura ; Analysis of time-varying characteristics of signals from Borssele reactor, Progress in Nuclear Energy, Vol.15, pp.853-861,1985.
- 2) 林光二、他 ; 炉雑音解析手法を用いたNSRRの異常診断(I)(II), JAERI-M 84-056 (1984), JAERI-M 84-137 (1987).
- 3) K.Hayashi, Y.Shinohara, E.Turkcan ; A method of nonstationary noise analysis using instantaneous AR spectrum and its application to Borssele reactor noise

7. おわりに

炉雑音信号の統計的解析に基づく原子炉システムの同定や異常予知診断の分野においては、非線形現象や非定常信号の解析が重要な課題になりつつあり、目下これらを一手に解決する有力な解析手法が待望されている。

本論文では、GMDHに基づく階層型モデリングをその最も有力な手法と位置付け、まず始めにGMDHの原理ならびにGMDHの基礎理論、すなわちシステムの並列化表現、階層構造、Kolmogorov-Gabor多項式に基づく基礎関数の導出、階層モデルのフィッティング方法、評価関数(発見的規範)について概説した。また、線形および非線形階層モデルのフィッティングを具体化するための計算アルゴリズムを処理段階毎に詳細に示した。アルゴリズムは1入力1出力オープン・ループ型モデルの他に、1変数型(自己回帰型)、多変数フィードバック型モデルへの拡張についても述べた。

次に、統計的信号処理の理論の立場から、Kolmogorov-Gabor型の基礎関数の係数を与える正規方程式を考察した。この結果、係数の推定式は高次の相関関数から成立っている事、すなわち、基礎関数が1次の場合には平均値(1次モーメント)と通常の相関関数(2次モーメント)から、また2次の場合にはそれに加えて2次および3次の高次相関関数(～4次モーメント)から成立っている事が明らかになった。

さらに、線形階層モデルとARモデルの関係を理論的に考察した。その結果、線形階層モデルはARモデルと等価な係数を持つこと、また白色雑音源を仮定することにより、ARモデルと同様に多変数フィードバック・システムの同定とその分解が可能であること等が明らかになった。

最後に、幾つかの線形シミュレーション・データを用いて、線形階層モデルのフィッティング試験を行い、係数値や予測誤差に関してARモデルと比較した。その結果、信号の構造が簡単な場合には、線形階層モデルはARモデルに比較してコンパクトでしかも十分実用的なモデルの得られることが確認できた。

参考文献

- 1) E.Turkcan, M.Kitamura ; Analysis of time-varying characteristics of signals from Borssele reactor, Progress in Nuclear Energy, Vol.15,pp.853-861,1985.
- 2) 林光二、他 ; 炉雑音解析手法を用いたNSRRの異常診断(I)(II), JAERI-M 84-056 (1984), JAERI-M 84-137 (1987).
- 3) K.Hayashi, Y.Shinohara, E.Turkcan ; A method of nonstationary noise analysis using instantaneous AR spectrum and its application to Borssele reactor noise

- analysis, Progress in Nuclear Energy, Vol.21, pp.707-716, 1988.
- 4) Y.Kuroda, H.Suzuki ; Some aspects of reactor noise analysis from the Bayesian viewpoint, Progress in Nuclear Energy, Vol.21, pp.717-723, 1988.
 - 5) 林光二、黒田義輝 ; 高出力炉における固有雑音と人工雑音-その3次スペクトルによる解析、東海大学大学院工学研究科修士論文、1979.
 - 6) H.Konno, H.Soneda, T.Tanaka, K.Hayashi ; Analysis of pressure fluctuations in the Borssele nuclear reactor based on bispectrum and higher order moments, Annals of nuclear energy, Vol 15, No.10/11, pp.487-500, 1988.
 - 7) 計測自動制御学会編 ; 自動制御ハンドブック、(8.4.6 非線形システムの同定、p.582) オーム社、1983.
 - 8) A.G.Ivakhnenko ; Heuristic self-organization in problems of engineering cybernetics, Automatica, Vol.6, pp.207-219, 1970.
 - 9) A.G.Ivakhnenko ; Polynomial Theory of Complex Systems, IEEE Trans., System, Man and Cybernetics, pp.364-378, 1971.
 - 10) 池田三郎、榎木義一 ; 解説-GMDH(発見的自己組織化法)と複雑な系の同定・予測、計測と制御, Vol.14, No.2, p185, (1975).
 - 11) 池田三郎、井原二郎 ; 講義 GMDH(変数組合せ計算法)の基礎と応用-発見的自己組織化による複雑なシステムのモデリング-I~VIII, システムと制御, Vol.23, No.12~Vol.24, No.7, (1979, 1980).
 - 12) M.Minsky, S.Papert ; Perceptrons, The MIT Press, USA, 1969.
齋藤正男訳、「パーセプトロン-パターン認識理論への道」東大出版会、1971.
 - 13) 木村正行 ; 自己組織系構成論、情報科学講座 E.18.2, 共立出版、1973.
 - 14) 麻生英樹 ; ニューラルネットワーク情報処理 -コネクショニズム入門あるいは柔らかな記号に向けて-, 産業図書、1988.?
 - 15) 井原二郎 ; 改良GMDH - 動的世界人口モデルの場合 -, システムと制御, Vol.19, No.4, pp.201-210, 1975.
 - 16) 湊小太郎、桑原道義 ; GMDHの収束に関する2,3の性質、電気学会論文誌 C, Vol.96, No.3, pp.56-60., 1976.
 - 17) 池田三郎、落合美紀子、榎木義一 ; 逐次GMDHアルゴリズムとその河川流量予測への応用、計測自動制御学会論文集、Vol.12, No.2, pp.209-126, 1976.
 - 18) 田村担之、青木年永 ; 対話形GMDHによる高域パターンの同定、計測自動制御学会論文集、Vol.12, No.2, pp.69-75, 1976.
 - 19) 佐藤敬 ; 技術報告 - 空容器回収量予測へのGMDHの応用、システムと制御、Vol.20, No.7, pp.365-372, 1976.
 - 20) 田村担之、近藤正 ; 部分表現式の次数を自己選択する改良型GMDHとその広域大気汚染濃度パターンの同定への応用、計測自動制御学会論文集、Vol.13, No.4, pp.351-357, 1977.
 - 21) 井原二郎 ; 多層型GMDHアルゴリズムの安定化 - その2 : 単純平均を用いる安定化法II、電子技術総合研究所彙報、Vol.41, No.4, pp.289-303, 1977.

- 22) 遠藤明男 ; 改造GMDHによる非線形制御系の同定、計測自動制御学会論文集、Vol.14, No.2, pp.130-135, 1978.
- 23) 近藤正、田村担之 ; 情報量基準AICを用いて中間表現式を自己選択する改良形GMDH、計測自動制御学会論文集、Vol.15, No.4, pp.466-471, 1979.
- 24) 田村担之、近藤正 ; 改良形GMDHによる大気汚染のモデリングと短期予測、計測自動制御学会論文集、Vol.15, No.5, pp.622-627, 1979.
- 25) 斎藤肅、小林主一郎 ; 水需要のモデリングと予測—情報量基準AICを統一的規範とするGMDHの適用—、計測自動制御学会論文集、Vol.17, No.5, pp.537-544, 1981.
- 26) 西川緯一、野村淳二、他 ; IWSM法とGMDHを用いた販売予測手法、システムと制御、Vol.27, No.7, pp.475-479
- 27) 田中英夫、林薫、他 ; ファジィGMDHの定式化、システムと制御、Vol.30, No.9, pp.581-587, 1986.
- 28) 兼安昌美、大石知司、他 ; 集積センサとパターン認識による臭気同定—GMDHによる混合臭気定量モデル—、計測自動制御学会論文集、Vol.23, No.10, pp.1030-1037, 1987.
- 29) 林薫、田中英夫、他 ; ファジィGMDHによるダム貯水池水温の解析と予測、計測自動制御学会論文集、Vol.23, No.12, pp.1304-1311, 1987.
- 30) Y. Shinohara, A. Toraiishi ; Application of GMDH to Reactor Dynamics Identification, 1977.
- 31) 篠原慶邦 ; 計算機プログラム—GMDH—, 私信(1977)
- 32) 池田秀松、鈴木篤之、他 ; パルサーフィーダー型パルスカラムにおける軸方向ホールドアップ、(II)GMDHによるモデル化と分析、日本原子力学会誌、Vol.29, No.2, pp.141-150, 1987.
- 33) H.L. Van Trees ; Synthesis of Optimum Nonlinear Control Systems, MIT Press, USA, 1962.
- 34) R.J. Roy and J. Sherman ; A Learning Technique for Volterra Series Representation, IEEE Trans., Automatic Control, pp.761-764, 1967
- 35) 赤池弘次、中川東一郎 ; ダイナミックシステムの統計的解析と制御、サイエンス社, 1972.
- 36) 赤池弘次 ; 情報量基準AICとは何か、数理科学、No.153, pp.5-11, 1976.
- 37) M. Rosenblatt, J.W. Van Ness ; Estimation of the bispectrum, Annals of Mathematical Statistics, Vol.36, pp.1120-1136, 1965.
- 38) 林光二、尾熊律雄 ; Burgのアルゴリズムに基づくARモデルの同定とその原子炉異常診断への応用、JAERI-M 82-009, 1982.
- 39) 林光二 ; 最小2乗法に基づくARモデル・フィッティング・アルゴリズムの評価、JAERI-M 84-127, 1984.
- 40) K. Hayashi, Y. Shinohara and K. Nabeshima ; Study on the goodness of system identification using multivariate AR modeling, Progress in Nuclear Energy, Vol. 21, pp.717-726, 1988.