第16卷 第4期 自动化学报 Vol. 16, No. 4
1990年7月 ACTA AUTOMATICA SINICA July, 1990

非线性系统辨识——GMDH的一种新算法及其应用

王秀峰　　刘　丹
（南开大学）

摘　　要

本文提出了用单向逐步回归方法确定 GMDH 中最佳部分多项式结构和进行参数估计,并且给出了递推算法. 这种算法避免了原算法[3]中的大量重复运算, 对多步预测模型更显出其优越性. 实际应用例子证明了该算法的有效性.

关键词——系统辨识,GMDH,参数估计,非线性系统,预测.

一、引　　言

GMDH (Group Method of Data Handling)[1] 是辨识非线性系统很有效的方法. 它的最大优点是不需要输入、输出变量之间的任何先验信息, 可以自己选择模型的结构(非线性阶次). 最基本的 GMDH 方法需要如下启发式条件:

1) 预先固定部分多项式结构(二次完全多项式);

2) 将原始数据分为拟合组和检验组;

3) 在每一选择层固定中间变量个数.

在基本的 GMDH 中,为了选出中间变量需要大量的重复计算,而得到的最终模型却很少是最佳组合;另外,辨识结果也强烈地依赖于启发式条件. 因此,Tamura 和 Kondo[2] 提出了启发自由的 GMDH,简记为 HFGMDH,在实际应用中得到了满意结果.

在时间序列的建模与预测领域中, 由于 HFGMDH 在参数估计和中间变量选择上,其准则本质上是基于一步超前预测误差,又由于被辨识的最终模型固有的非线性性质,所以多步超前预测不够理想. 针对这个问题, Gallo 等[3]提出了一个在所考虑的整个预测域内加权平均预测误差平方和准则,记为 MWSS,用来协调一步和多步预测之间的性能,改善了多步预测结果. 但它在预测域内每一步都需要大量的重复性运算, 对于较大的预测域其计算量是不可想象的.

本文给出的单向逐步回归方法是针对上述问题提出的. 该法采用递推算法, 不仅可以得到最佳部分多项式,而且避免了重复计算.

本文于1988年1月22日收到.

二、GMDH 方法及几种准则简介

一非线性系统，$x_1, x_2, \cdots, x_n$ 是输入；y 是输出，其关系为

$$y = f(x_1, x_2, \cdots, x_n). \tag{1}$$

函数 f 的离散 Volterra 级数展开式——Kolmogorov-Gabor 多项式为

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots. \tag{2}$$

它广泛地用作非线性模型的完全描述.

GMDH 是通过多层筛选的办法求得模型(2)的近似描述．基本的 GMDH 方法示意图如图 1 所示．后来对 GMDH 方法的各种改进(例如 Tamura 和 Kondo, 1980; Gallo 等, 1985)，都是集中讨论每一层如何产生部分多项式；如何确定部分多项式的结构；按什么准则选择中间变量．虽然不同的准则可能得到不同的最终模型，但目前普遍采用启发自由、选择最佳部分多项式结构的方法.

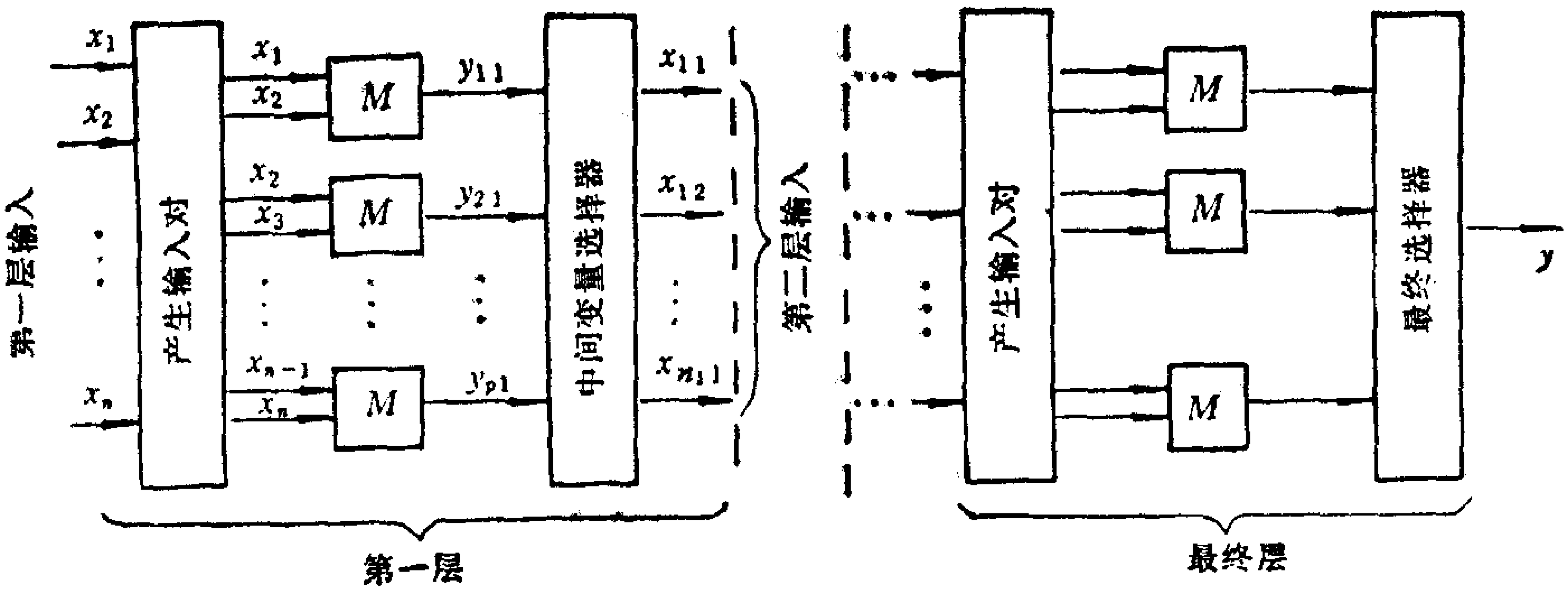

图 1　基本 GMDH 示意图

M 是部分多项式；y_{ij} 是由部分多项式模型得到的输出；x_{ij} 是中间变量.

确定部分多项式结构及选择中间变量常用的几种准则如下：

C1：预测误差平方和准则 PESS

$$\text{PESS} = \sum_{t=1}^{N} [y(t) - \hat{y}^*(t)]^2. \tag{3}$$

其中

$$\hat{y}^*(t) = a_{0t} + a_{1t}x_i(t) + a_{2t}x_j(t) + a_{3t}x_i(t)x_j(t) + a_{4t}x_i^2(t) + a_{5t}x_j^2(t)$$
$$(t = 1, 2, \cdots, N)$$

是通过除第 t 次量测外，所有数据用多元回归得到的部分多项式，对第 t 次输出 $y(t)$的估计值．N是数据长度.

(3) 式可以减化为

$$\text{PESS} = \sum_{t=1}^{N} \left(\frac{y(t) - \hat{y}(t)}{1 - \boldsymbol{x}_t^T (X^T X)^{-1} \boldsymbol{x}_t}\right)^2. \tag{4}$$

其中 $\hat{y}(t)$ 是用所有数据通过多元回归得到的对第 t 次输出的估计值.

$$\boldsymbol{x}_t^T = [1, x_i(t), x_j(t), x_i(t)x_j(t), x_i^2(t), x_j^2(t)],\quad t = 1, 2, \cdots, N,$$

$$X^T = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]. \tag{5}$$

将 GMDH 方法用于单变量时间序列分析时,只要令序列 $y(t)$ 的采样滞后 $y(t-i)$ 作为第一层的输入 x_i,则(4)式可写为

$$\text{PESS} = \sum_{t=1}^{N}\left[\frac{y(t) - \hat{y}(t)}{1 - \boldsymbol{y}_t^T(Y^TY)^{-1}\boldsymbol{y}_t}\right]^2. \tag{6}$$

其中

$$\boldsymbol{y}_t^T = [1, y(t-i), y(t-j), y(t-i)y(t-j), y^2(t-i), y^2(t-j)];$$

$$Y^T = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N].$$

C2: AIC 准则[4]

$$\text{AIC} = N\ln s_k^2 + 2k + c, \tag{7}$$

$$s_k^2 = \frac{1}{N}\sum_{t=1}^{N}[y(t) - \hat{y}(t)]^2. \tag{8}$$

其中 c 是常量; k 是独立可调参数的个数.

C3: 预测域内加权误差平方和准则(Gallo 等)

$$\text{MWSS} = \sum_{i=1}^{N}\sum_{\lambda=1}^{L} W(\lambda)e_i^2(\lambda). \tag{9}$$

其中 $W(\lambda)$ 是对超前 λ 步误差的加权函数($\lambda \geqslant 1$,整数); $e_i(\lambda)$ 是在时刻 i 超前 λ 步的预测误差.

MWSS 也可作为整个多层筛选过程终止的判据. 用 MWSS(k,j) 表示在第 k 层对第 j 个部分多项式模型计算出的 MWSS 值,令

$$\overline{\text{MWSS}}(k) = \min\text{MWSS}(k,j).$$

当某个 $\hat{k}$,使得

$$\frac{\overline{\text{MWSS}}(\hat{k}) - \overline{\text{MWSS}}(\hat{k}-1)}{\overline{\text{MWSS}}(\hat{k})} \leqslant \varepsilon \tag{10}$$

时,则筛选过程在第 $\hat{k}$ 层停止. ε 由经验给定.

C1,C2 只给出固定预测步数较好的性能,C3 可以给出整个预测域内任意步数较好的性能. 但它对每一步数都要估计部分多项式参数及预测误差,反复计算将按指数增长.为了避免大量的重复计算,本文提出如下单向逐步回归算法.

三、单向逐步回归算法

从上面所介绍的基本 GMDH 以及各种修正方案，不难看出无论采用何种准则 (C1, C2, C3)，主要计算量都集中在每一层对所有可能的输入对确定部分多项式的最佳结构、参数估计以及计算预测误差. 本文给出的单向逐步回归递推算法可以同时得到部分多项式结构、参数估计及预测误差.

所谓单向逐步回归是指对每个变量逐步挑选，如果此变量对预测误差有明显改善，则被选出，否则不选，一旦被选中则不再被剔除.

设输入变量为 $x_1, x_2, \cdots, x_n$，输出变量为 y，任意两个输入变量对 x_i, x_j，其完全二次多项式为

$$\varphi = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2. \tag{11}$$

我们的目的是在含这两个变量的二次多项式中确定一个最佳结构，使之包括较少的项而输出误差又较小．基本的 GMDH 方法是固定多项式的结构为(11)式，这不仅计算量大而且多余的变量将会带来更大误差．因此，后来的修正方案都是企图去掉多余的项，以保证必要的简单结构．本文提出的单向逐步回归既能得到较好结构又有计算简单的优点．

引理． 设矩阵 $X = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_n)$，$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 是线性无关向量组，$\boldsymbol{y}$ 是与 $\boldsymbol{x}_i$ 同维向量．令

$$X_1 = (X\ \boldsymbol{y}), \quad X_2 = (\boldsymbol{y}\ X),$$
$$\Phi_1 = X_1^T X_1, \quad \Phi_2 = X_2^T X_2, \quad \Phi = X^T X.$$

则有

$$\Phi_1^{-1} = \begin{bmatrix} \Phi^{-1} + \boldsymbol{\theta}\boldsymbol{\theta}^T/\sigma & -\boldsymbol{\theta}/\sigma \\ -\boldsymbol{\theta}^T/\sigma & 1/\sigma \end{bmatrix}, \tag{12}$$

$$\Phi_2^{-1} = \begin{bmatrix} 1/\sigma & -\boldsymbol{\theta}^T/\sigma \\ -\boldsymbol{\theta}/\sigma & \Phi^{-1} + \boldsymbol{\theta}\boldsymbol{\theta}^T/\sigma \end{bmatrix}. \tag{13}$$

其中 $\boldsymbol{\theta}$ 是在最小二乘意义下 $\boldsymbol{y}$ 由 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 线性表示的组合系数；σ 是误差向量的模的平方．

证明．(只证(12)式．(13)式同理可证)

由于矩阵 X 是列线性无关的，所以 Φ^{-1} 存在，由矩阵分块运算法则[5]有

$$\Phi_1^{-1} = \begin{bmatrix} \Phi^{-1} + \Phi^{-1}X^T\boldsymbol{y}p^{-1}\boldsymbol{y}^T X\Phi^{-1} & -\Phi^{-1}X^T\boldsymbol{y}p^{-1} \\ -p^{-1}\boldsymbol{y}^T X\Phi^{-1} & p^{-1} \end{bmatrix}. \tag{14}$$

其中 $p = \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{y}^T X\Phi^{-1}X^T\boldsymbol{y}$.

令

$$\boldsymbol{\theta} = \Phi^{-1}X^T\boldsymbol{y}. \tag{15}$$

注意到 X 和 Φ 的定义，$\boldsymbol{\theta}$ 就是方程 $\boldsymbol{y} = X\boldsymbol{\theta}$ 的最小二乘解，即 $\boldsymbol{y}$ 由 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 线性表示的组合系数．

由于 $(I - X\Phi^{-1}X^T)^T(I - X\Phi^{-1}X^T) = I - X\Phi^{-1}X^T$,

又

$$(I - X\Phi^{-1}X^T)\boldsymbol{y} = \boldsymbol{y} - X\Phi^{-1}X^T y = \boldsymbol{y} - X\boldsymbol{\theta} \triangleq \boldsymbol{e},$$

所以

$$\begin{aligned} p &= \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{y}^T X\Phi^{-1}X^T\boldsymbol{y} = \boldsymbol{y}^T(I - X\Phi^{-1}X^T)\boldsymbol{y} \\ &= [(I - X\Phi^{-1}X^T)\boldsymbol{y}]^T[(I - X\Phi^{-1}X)\boldsymbol{y}] = \boldsymbol{e}^T\boldsymbol{e}. \end{aligned} \tag{16}$$

将式(15)，(16)代入(14)即得(12)式．

证毕．

有了上述引理，我们可以得到单向逐步回归算法． 假定取得了 N 组数据：$\{y_k, x_{1k}, \cdots, x_{nk}; k = 1, \cdots, N\}$．令

$$\boldsymbol{z}_0=(1,1,\cdots,1)^T,\qquad \boldsymbol{z}_1=(x_{i1},x_{i2},\cdots,x_{iN})^T,$$
$$\boldsymbol{z}_2=(x_{j1},x_{j2},\cdots,x_{jN})^T,\quad \boldsymbol{z}_3=(x_{i1}x_{j1},x_{i2}x_{j2},\cdots,x_{iN}x_{jN})^T,$$
$$\boldsymbol{z}_4=(x_{i1}^2,x_{i2}^2,\cdots,x_{iN}^2)^T,\quad \boldsymbol{z}_5=(x_{j1}^2,x_{j2}^2,\cdots,x_{jN}^2)^T,$$
$$\boldsymbol{y}=(y_1,y_2,\cdots,y_N)^T.$$

显然,若有

$$y=a_0+a_1x_i+a_2x_j+a_3x_ix_j+a_4x_i^2+a_5x_j^2,$$

则有
$$\boldsymbol{y}=a_0\boldsymbol{z}_0+a_1\boldsymbol{z}_1+a_2\boldsymbol{z}_2+a_3\boldsymbol{z}_3+a_4\boldsymbol{z}_4+a_5\boldsymbol{z}_5.$$

因此,若按如下顺序选择向量

$$\boldsymbol{y},\boldsymbol{z}_0,\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3,\boldsymbol{z}_4,\boldsymbol{z}_5,\tag{17}$$

如果 $\boldsymbol{z}_i$ 加入后,对 $\boldsymbol{y}$ 的预测有明显的改善,则被选入,否则不被选入.

记
$$R_0=[\boldsymbol{y}],$$
$$R_1=[\boldsymbol{y}\boldsymbol{z}_0],$$
$$\vdots$$

一般地, $R_{i+1}=[R_i\boldsymbol{z}]$. $\boldsymbol{z}$ 是 $\boldsymbol{z}_0,\cdots,\boldsymbol{z}_5$ 中的某一个.

记
$$\Phi_i=R_i^TR_i,\ \Phi_{i+1}=R_{i+1}^TR_{i+1}=\begin{bmatrix}R_i^TR_i & R_i^T\boldsymbol{z}\\ \boldsymbol{z}^TR_i & \boldsymbol{z}^T\boldsymbol{z}\end{bmatrix}.\tag{18}$$

若 Φ_i^{-1} 已知,则易算出

$$\boldsymbol{\theta}=\Phi_i^{-1}R_i^T\boldsymbol{z},\quad \boldsymbol{e}=\boldsymbol{z}-R_i\boldsymbol{\theta}$$

以及 $\sigma=\boldsymbol{e}^T\boldsymbol{e}$. 利用引理可得 Φ_{i+1}^{-1}. 另一方面,我们也可将 R_{i+1} 重新分块为 $R_{i+1}=[\boldsymbol{y}\ \tilde{R}_i]$,其中 $\tilde{R}_i=[\boldsymbol{z}_0\cdots\boldsymbol{z}]$, 这时可利用(13)式求得关于 $\boldsymbol{y}$ 的组合系数 $\hat{\boldsymbol{\beta}}$ 及 $\tilde{\Phi}_i^{-1}$, 这里 $\tilde{\Phi}_i=\tilde{R}_i^T\tilde{R}_i$. 从而可得误差向量 $\boldsymbol{\xi}=\boldsymbol{y}-\tilde{R}_i\hat{\boldsymbol{\beta}}$. $\tilde{\Phi}_i^{-1}$ 相当于(4)式中的 $(X^TX)^{-1}$. 令 $\boldsymbol{x}_k$ 表示 $\tilde{R}_i$ 的第 k 行,因此又可得到

$$\text{PESS}=\sum_{k=1}^{N}\left[\frac{\boldsymbol{\xi}(k)}{1-\boldsymbol{x}_k^T\tilde{\Phi}_i^{-1}\boldsymbol{x}_k}\right]^2.\tag{19}$$

如果 $\boldsymbol{z}$ 对于 PESS 有明显改善,则选入,从 R_{i+1} 开始,否则从 R_i 开始,按(17)式的顺序再选其它向量,直至选到 $\boldsymbol{z}_5$ 为止.

另外,由于 $\tilde{\Phi}_i^{-1}$ 已知,因此对任意向量 $\tilde{\boldsymbol{y}}$,能很方便地算出最小二乘估计

$$\hat{\boldsymbol{\beta}}=\tilde{\Phi}_i^{-1}\tilde{R}_i^T\tilde{\boldsymbol{y}}.\tag{20}$$

及其误差向量

$$\boldsymbol{\xi}=\tilde{\boldsymbol{y}}-\tilde{R}_i\hat{\boldsymbol{\beta}}.\tag{21}$$

所以对于准则 C3,应用上述算法,只要对 $\lambda=1$ 运算一次,用(20),(21)式就可很方便地得到 λ 的其它值所对应的参数值及相应的 PESS. 避免了大量重复计算.

对于每一输入对,单向逐步回归计算步骤如下:

1) 构造向量 $\boldsymbol{y},\boldsymbol{z}_0,\boldsymbol{z}_1,\cdots,\boldsymbol{z}_5$;

2) 置 $i=0,R_0=[\boldsymbol{y}],\Phi_0^{-1}=1/\boldsymbol{y}^T\boldsymbol{y},k=0$;

3) 取向量 $\boldsymbol{z}_k$;

4) 构造矩阵 $R_{i+1}=[R_i\boldsymbol{z}_k]$;

5) 计算　$\boldsymbol{\theta} = \Phi_i^{-1}R_i^T\boldsymbol{z}_k$，$\boldsymbol{e} = \boldsymbol{z}_k - R_i\boldsymbol{\theta}$，$\sigma = \boldsymbol{e}^T\boldsymbol{e}$，

$$\Phi_{i+1}^{-1} = \begin{bmatrix} \Phi_i^{-1} + \boldsymbol{\theta}\boldsymbol{\theta}^T/\sigma & -\boldsymbol{\theta}/\sigma \\ -\boldsymbol{\theta}^T/\sigma & 1/\sigma \end{bmatrix};$$

6) 计算　$\hat{\boldsymbol{\beta}}$ 和 $\tilde{\Phi}_i^{-1}$(由(13)式)，

构造　$\tilde{R}_i$，并计算残差向量 $\boldsymbol{\xi}$.

$$\boldsymbol{\xi} = \boldsymbol{y} - \tilde{R}_i\hat{\boldsymbol{\beta}};$$

7)　计算 PESS；

8)　如果 $k = 5$ 转 10；

9)　如果 PESS 显著改善，则 $k+1 \Rightarrow k, i+1 \Rightarrow i$ 转 3. 否则 $k+1 \Rightarrow k$ 转 3；

10) 输出多项式结构及参数.

注 1. 如果用准则 C3，则需在第 7 步后增加计算

$$\hat{\boldsymbol{\beta}}_\tau = \tilde{\Phi}_i^{-1}\tilde{R}_i^T\boldsymbol{y}_\tau,\quad \boldsymbol{\xi}_\tau = \boldsymbol{y}_\tau - \tilde{R}_i\hat{\boldsymbol{\beta}}_\tau.\quad \tau = 1, \cdots, l.$$

注 2. 设 PESS1，PESS2 分别为增加变量前后两次的 PESS，如果

$$\frac{\text{PESS1} - \text{PESS2}}{\text{PESS1}} > \varepsilon,$$

则说明新加的变量对 PESS 有明显改善，ε 是正数，可经验确定.

四、应 用 实 例

作者将上述算法应用于太阳黑子活动情况的预测，得到了比较满意的结果. 下面是用 1770 年到 1869 年的 100 个数据，考虑的两种情形：

1) 假定　$y(t) = f(y(t-1), y(t-2), y(t-3))$，

得到的最终模型为

$$y(t) = 5.48 + 0.524x_2 + 0.004x_2^2,$$
$$x_2 = -3.63 + 1.84y(t-1) - 0.007y(t-1)y(t-3) - 0.005y^2(t-1) - 0.002y^2(t-3).$$

实测值与估计值的图形如图 2 所示.

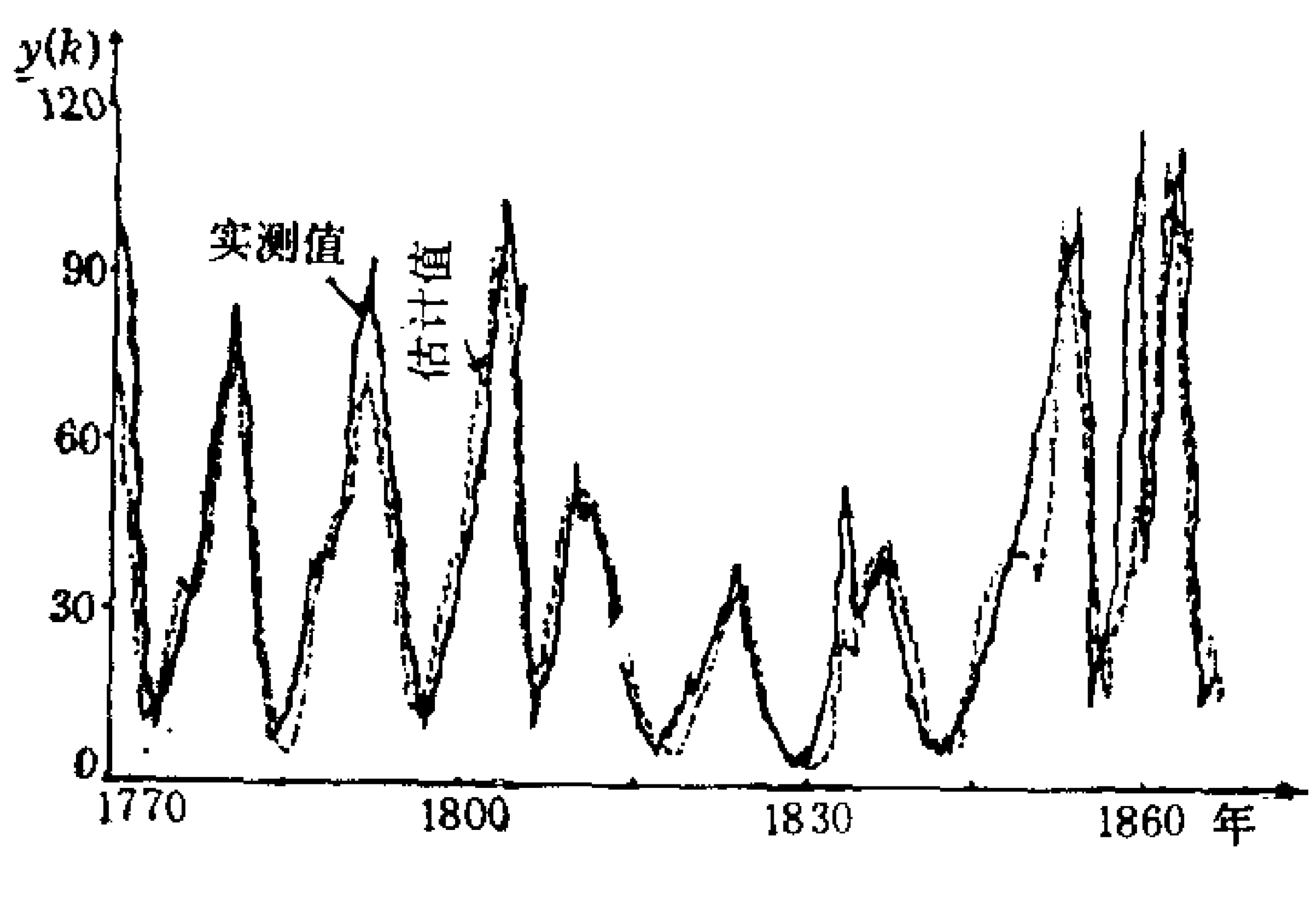

图 2　模型 1 的预测结果

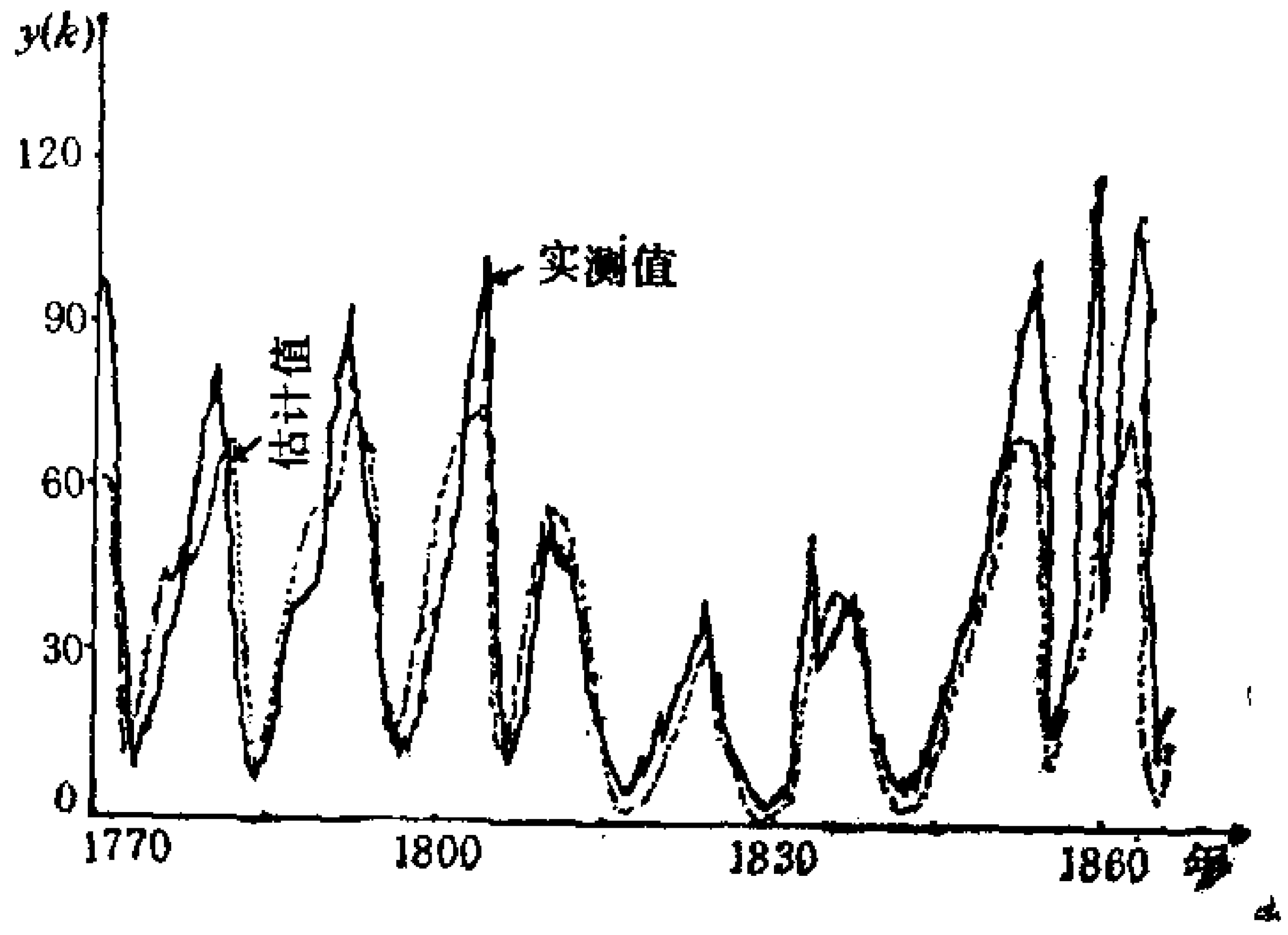

图 3　模型 2 的预测结果

2) 假定 $y(t)=f(y(t-1),y(t-2))$，
得到最终模型为

$$y(t)=1.75+1.48y(t-1)-0.006y^2(t-1).$$

实测值与估计值如图 3 所示.

结 束 语

本文提出的算法对 PESS 的计算非常方便. 对于准则 C3,不需重复 l 次辨识程序就可得到不同预测步数的多项式结构、参数及 MWSS,为准则 C3 的应用扫除了障碍.

参 考 文 献

[1] Ivakhnenko, A. G., Heuristic Self-Organisation in Problems of Engineering Cybernetics, *Automatica*, 6 (1970), 207—219.

[2] Tamura, H. and T. Kondo, Heuristics Free Group Method of Data Handling Algorithm of Generating Optimal Partial Polynomials With Application to Air Pollution Predication, *Int. J. Systems Sci.*, 11(1980), 1095—1111.

[3] Gallo, P., R. Genesio and A. Vicino, Nonlinear Structure Identification for Multistep Prediction Via a Modified GMDH Algorithm, Preprints 7th IFAC Symp. on Iden. and Syst. Par. Estimation, York, UK (1985), 759—764.

[4] Akaike, H., A New Look at the Statistical Model Identification, *IEEE. Trans.* AC-19(1974), 716—723.

[5] 〔日〕须田信英等著,曹长修译,自动控制中的矩阵理论,科学出版社,(1979),31—33.

AN ALGORITHM AND ITS APPLICATION FOR GMDH — NONLINEAR SYSTEMS IDENTIFICATION

WANG XIUFENG LIU DAN

(*Nankai University*)

ABSTRACT

In this paper a unidirectional stepwise regression algorithm is presented to determine the structure of the optimal partial polynomial and estimate parameters in GMDH. In order to avoid intensive iterative computation, a recursion algorithm is proposed. The algorithm is particularly suited to multistep predicted modelling. Application examples have shown its effectiveness.

Key words——Systems identification; GMDH; prediction; parameter estimation; nonlinear systems.