2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

HLS parameter estimation for multi-input multi-output systems

Ping Yuan, Feng Ding and Peter X. Liu

Abstract— **In order to reduce computational burden of identification methods for multivariable systems, a hierarchical least squares (HLS) algorithm is developed. The basic idea is to use the hierarchical identification principle to decompose the identification model of the multivariable system into several submodels with smaller dimensions and fewer variables, and then to identify the parameter vector of each submodel. The analysis indicates that the parameter estimation error given by the proposed algorithm converges to zero under the persistent excitation. Also, the algorithm has much less computational efforts than the recursive least squares algorithm and is easy to implement on computer. Finally, we test the proposed algorithm by an example.**

Index Terms: Least squares identification; parameter estimation; convergence properties; hierarchical identification principle; multivariable systems.

I. Problem formulation

Consider a multivariable system described by the following state-space model,

$$\begin{aligned} x(t+1) &= Ax(t)+Bu(t), && (1)\\ y(t) &= Cx(t)+Du(t), && (2)\end{aligned}$$

where $x(t)\in\mathbb{R}^n$ is the state vector, $u(t)\in\mathbb{R}^r$ the system input vector, $y(t)\in\mathbb{R}^m$ the system output vector, (A,B,C,D) the system matrices of appropriate sizes. Although many identification algorithms can be used to estimate the parameters of this state space model [13], [10], the transfer matrix representation with the input-output relationship is very useful in practice [5]. Thus the state-space model is transformed into the transfer matrix model to be identified. The details are as follows. Let z^{-1} be the unit delay operator: $zx(t)=x(t+1)$. From (1)-(2), we can get

$$\begin{aligned} y(t) &= [C(zI-A)^{-1}B+D]u(t)\\ &= \left[\frac{C\,\mathrm{adj}[zI-A]B}{\det[zI-A]}+D\right]u(t)\\ &= \left[\frac{z^{-n}C\,\mathrm{adj}[zI-A]B}{z^{-n}\det[zI-A]}+D\right]u(t)\\ &=: \frac{Q(z)}{\alpha(z)}u(t), && (3)\end{aligned}$$

This research was supported by the National Natural Science Foundation of China (No. 60574051) and the Natural Science Foundation of Jiangsu Province, China (BK2007017) and by Program for Innovative Research Team of Jiangnan University.

P. Yuan and F. Ding (Corresponding author) are with the Control Science and Engineering Research Center, Jiangnan University, Wuxi 214122, P.R. China. Email: nyp506@163.com (Y. Ping), fding@jiangnan.edu.cn (F. Ding).

P.X. Liu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada K1S 5B6; Email: xpliu@sce.carleton.ca

where $\alpha(z)$ is the characteristic polynomial in z^{-1} of the system (of degree n), $Q(z)$ is the polynomial matrix in z^{-1}, and they can be represented as

$$\begin{aligned} \alpha(z) &= z^{-n}\det[zI-A]\\ &= 1+\alpha_1 z^{-1}+\alpha_2 z^{-2}+\cdots+\alpha_n z^{-n}\in\mathbb{R}^1,\\ Q(z) &= z^{-n}C\,\mathrm{adj}[zI-A]B+z^{-n}\det[zI-A]D\\ &= Q_0+Q_1z^{-1}+\cdots+Q_nz^{-n}\in\mathbb{R}^{m\times r},\\ &\quad Q_i\in\mathbb{R}^{m\times r},\ i=0,1,\cdots,n.\end{aligned}$$

Equation (3) can also be written as

$$y(t)+\sum_{i=1}^{n}\alpha_i y(t-i)=\sum_{i=0}^{n}Q_i u(t-i).$$

Define the parameter matrix θ, parameter vector α, information vector $\varphi(t)$ and information matrix $\psi(t)$ as

$$\begin{aligned} \theta^{\mathrm{T}} &= [Q_0,Q_1,\cdots,Q_n]\in\mathbb{R}^{m\times n_0},\ n_0:=(n+1)r,\\ \alpha &= \begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_n\end{bmatrix}\in\mathbb{R}^n,\quad \varphi(t)=\begin{bmatrix}u(t)\\ u(t-1)\\ \vdots\\ u(t-n)\end{bmatrix}\in\mathbb{R}^{n_0},\\ \psi(t) &= [y(t-1),y(t-2),\cdots,y(t-n)]\in\mathbb{R}^{m\times n}.\end{aligned}$$

Taking into account that there exist disturbances in the physical systems and introducing a noise vector $v(t)$ in (3) yield

$$y(t)+\psi(t)\alpha=\theta^{\mathrm{T}}\varphi(t)+v(t). \qquad (4)$$

Because the system in (4) contains both a parameter matrix θ and a parameter vector α, Model (4) is transformed by re-parameterization into a form which can be identified by the standard recursive least squares (RLS) algorithm.

Let us introduce some notation first. The symbol I_m is an $m\times m$ identity matrix; the norm of the matrix X is defined by $\|X\|^2=\mathrm{tr}[XX^{\mathrm{T}}]$; $\otimes$ denotes the Kronecker product or direct product, if $A=[a_{ij}]\in\mathbb{R}^{m\times n}$, $B=[b_{ij}]\in\mathbb{R}^{p\times q}$, then $A\otimes B=[a_{ij}B]\in\mathbb{R}^{mp\times nq}$ (in general, $A\otimes B\neq B\otimes A$); $\mathrm{col}[X]$ denotes the vector formed by the column of the matrix X, if $X=[x_1,x_2,\cdots,x_n]\in\mathbb{R}^{m\times n}$, $x_i\in\mathbb{R}^m$, $i=1,2,\cdots,n$, then $\mathrm{col}[X]=[x_1^{\mathrm{T}},x_2^{\mathrm{T}},\cdots,x_n^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{R}^{mn}$.

Define the parameter vector ϑ and the information matrix $\Phi(t)$ as

$$\begin{aligned} \vartheta &= \begin{bmatrix}\alpha\\ \mathrm{col}[\theta^{\mathrm{T}}]\end{bmatrix}\in\mathbb{R}^{n+mn_0},\\ \Phi(t) &= [-\psi(t),\varphi^{\mathrm{T}}(t)\otimes I_m]\in\mathbb{R}^{m\times(n+mn_0)}.\end{aligned}$$

Based on the above definition, the system in (4) can be written in the form:

$$y(t) = \Phi(t)\vartheta + v(t). \tag{5}$$

For Model (5), Sen and Sinha used a matrix pseudo-inverse approach [8] to give the following recursive least squares algorithm to obtain the estimate $\hat{\vartheta}(t)$ of ϑ,

$$\begin{aligned} \hat{\vartheta}(t) &= \hat{\vartheta}(t-1) + P(t)\Phi^{\mathrm{T}}(t)[y(t) - \Phi(t)\hat{\vartheta}(t-1)], &(6)\\ P^{-1}(t) &= P^{-1}(t-1) + \Phi^{\mathrm{T}}(t)\Phi(t). &(7)\end{aligned}$$

Applying the matrix inversion formula

$$(A+BC)^{-1} = A^{-1} - A^{-1}B(I+CA^{-1}B)^{-1}CA^{-1}$$

to (7) gives

$$\begin{aligned} P(t) &= P(t-1) - P(t-1)\Phi^{\mathrm{T}}(t) \times \\ &\quad [I_m + \Phi(t)P(t-1)\Phi^{\mathrm{T}}(t)]^{-1}\Phi(t)P(t-1). \end{aligned}$$

However, the Sen and Sinha's algorithm has large computational load since the information matrix $\Phi(t) \in \mathbb{R}^{m\times(mn_0+n)}$ in the estimation algorithm has very large sizes especially for large m, r or n, which leads to compute the covariance matrix $P(t) \in \mathbb{R}^{(mn_0+n)\times(mn_0+n)}$ of large dimensions at each step, see Table I. Therefore, the objective of this paper is, by means of the hierarchical identification principle, to propose new and computationally efficient identification algorithm to estimate the unknown parameter vector ϑ and to study convergence properties of the algorithm involved.

II. The hierarchical least squares algorithm

This section presents the hierarchical identification algorithm with less computational efforts to estimate ϑ by using the hierarchical identification principle. The hierarchical identification principle [3], [4] is to decompose the system in (5) into several fictitious subsystems with smaller dimension and fewer variables, and then the parameter vector of each subsystem is identified, respectively[12]. However, because there exists associated items between two subsystems, i.e., the ith subsystem contains the unknown parameter vector ϑ_j $(j \neq i)$ of other subsystems, so it makes iterative calculations difficulty. In order to settle this problem, when computing the estimates of ϑ_i in the ith subsystem at time t, we replace the unknown parameter vectors ϑ_j of other subsystems with their estimates $\hat{\vartheta}_j(t-1)$ at time $(t-1)$. Now we will derive hierarchical least squares algorithm according to the hierarchical identification principle.

The information matrix $\Phi(t)$ in (5) is decomposed into N information sub-matrices with dimensions $m \times n_i$, and the parameter vector ϑ into N parameter sub-vectors with dimensions n_i, i.e.,

$$\begin{aligned} \Phi(t) &= [\Phi_1(t), \Phi_2(t), \cdots, \Phi_N(t)] \in \mathbb{R}^{m\times(n+mn_0)}, \\ \vartheta &= \begin{bmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_N \end{bmatrix} \in \mathbb{R}^{n+mn_0}, \end{aligned}$$

$$\Phi_i(t) \in \mathbb{R}^{m\times n_i}, \quad \vartheta_i \in \mathbb{R}^{n_i}, \quad n_1 + n_2 + \cdots + n_N = n + mn_0.$$

Thus Equation (5) is decomposed into N submodels which can be expressed as

$$y_i(t) = \Phi_i(t)\vartheta_i + v(t), \ i = 1, 2, \cdots, N, \tag{8}$$

where

$$\begin{aligned} y_i(t) &:= y(t) - \eta_i(t), &(9)\\ \eta_i(t) &:= \sum_{j=1, j\neq i}^{N} \Phi_j(t)\vartheta_j, \ i = 1, 2, \cdots, N. &(10)\end{aligned}$$

$\eta_i(t)$ is called the associated items between the subsystems, which contains the unknown parameter vector ϑ_j $(j \neq i)$ of other subsystems, $y_i(t)$ is a function of the associated variables $\eta_i(t)$.

Let $\hat{\vartheta}_i(t)$ denote the estimate of ϑ_i at time t. According to the least squares principle, one can get the following recursive least squares algorithm for estimating ϑ_i in (8),

$$\begin{aligned} \hat{\vartheta}_i(t) &= \hat{\vartheta}_i(t-1) + P(t)\Phi^{\mathrm{T}}(t) \times \\ &\quad [y_i(t) - \Phi_i(t)\hat{\vartheta}_i(t-1)], &(11)\\ P^{-1}(t) &= P^{-1}(t-1) + \Phi^{\mathrm{T}}(t)\Phi(t). &(12)\end{aligned}$$

Substituting (9)-(10) into (11), we have

$$\begin{aligned} \hat{\vartheta}_i(t) &= \hat{\vartheta}_i(t-1) + P(t)\Phi^{\mathrm{T}}(t) \times \\ &\quad [y(t) - \eta_i(t) - \Phi_i(t)\hat{\vartheta}_i(t-1)] \\ &= \hat{\vartheta}_i(t-1) + P(t)\Phi^{\mathrm{T}}(t) \times \\ &\quad \left[y(t) - \sum_{j=1, j\neq i}^{N} \Phi_j(t)\vartheta_j - \Phi_i(t)\hat{\vartheta}_i(t-1) \right]. \end{aligned} \tag{13}$$

Difficulty arises in that the expression on the right-hand side of (13) contains the unknown parameter vectors ϑ_j, $j = 1, 2, \cdots, i-1, i+1, \cdots, N$; so the estimate $\hat{\vartheta}_i(t)$ is impossible to compute by (13). In order to compute the estimate $\hat{\vartheta}_i(t)$, the approach here is based on the hierarchical identification principle: the unknown variables ϑ_j $(j \neq i)$ in (13) are replaced with their estimates $\hat{\vartheta}_j(t-1)$ at time $(t-1)$, then we obtain the hierarchical least squares identification algorithm as follows:

$$\begin{aligned} \hat{\vartheta}_i(t) &= \hat{\vartheta}_i(t-1) + P(t)\Phi^{\mathrm{T}}(t) \times \\ &\quad \left[y(t) - \sum_{j=1, j\neq i}^{N} \Phi_j(t)\hat{\vartheta}_j(t-1) - \Phi_i(t)\hat{\vartheta}_i(t-1) \right] \\ &= \hat{\vartheta}_i(t-1) + P(t)\Phi^{\mathrm{T}}(t)[y(t) - \Phi(t)\hat{\vartheta}(t-1)], \\ &\quad i = 1, 2, \cdots, N, \end{aligned} \tag{14}$$

where

$$\hat{\vartheta}(t) = \begin{bmatrix} \hat{\vartheta}_1(t) \\ \hat{\vartheta}_2(t) \\ \vdots \\ \hat{\vartheta}_N(t) \end{bmatrix} \in \mathbb{R}^{n+mn_0} \text{ denotes the estimate of } \vartheta.$$

Equations (12) and (14) form the hierarchical least squares identification algorithm for estimating the parameter vector

ϑ. To initialize the algorithm, we generally take $P_i(0) = p_0 I_{n_i}$ with p_0 normally a large positive number (e.g., $p_0 = 10^6$) and $\hat{\vartheta}_i(0) = \mathbf{1}_{n_i}/p_0$, $i = 1,2,\cdots,N$ [7]. This HLS algorithm differs not only from the ones in [1], [2] with two identification submodels: one parameter vector and one parameter matrix, but also from the ones in [4], [3] which decompose the parameter vector and information vector into parameter sub-vectors and information sub-vectors, and the parameter matrix and information vector into parameter sub-matrices and information sub-vectors, respectively, but the approach decomposes the parameter vector and information matrix into parameter sub-vectors and information sub-matrices, So the algorithm is easy to implement on computer and the contribution lies in reducing computational burden of identification methods.

The computational burden of the hierarchical least squares algorithm and the recursive least squares algorithm are listed in Table I, where the numbers of multiplications and additions are for each step, and the numbers in the brackets in Table I denote the recorded numbers for a 5-input, 5-output and 5th-order system at each step. Here, $mn_0 + n = 155$, so the parameter matrix $\Phi(t) \in \mathbb{R}^{5\times 155}$ and the parameter vector $\vartheta \in \mathbb{R}^{155}$, we take $N = 31$, i.e., System (5) is decomposed into $N = 31$ submodels and $n_i = 5$, $i = 1,2,\cdots,31$. From Table I, it is clear that the hierarchical least squares algorithm is computationally more efficient than the RLS algorithms [1], [2].

TABLE I
COMPARISON OF COMPUTATIONAL EFFICIENCY OF RLS AND HLS

Algorithm	Number of multiplications	Number of additions
RLS	$2m(mn_0+n)^2$ $+2m(m+1)(mn_0+n)$ [249550]	$2m(mn_0+n+m)\times$ $(mn_0+n)-m^2+m$ [247980]
HLS	$\sum_{i=1}^{N}[2mn_i^2+2m(m+1)n_i]$ [17050]	$\sum_{i=1}^{N}[2m(n_i+m)n_i$ $-m^2+m]$ [14880]

III. CONVERGENCE OF HIERARCHICAL LEAST SQUARES IDENTIFICATION

In order to study the convergence properties of the hierarchical least squares algorithm, the following assumptions and preliminary facts are required.

We assume that $\{v(t)\}$ are martingale difference vector sequences defined on a probability space $(\Omega, \mathscr{F}, P)$, where $\{\mathscr{F}_t\}$ is generated by $\{v(t)\}$ and including time t, i.e., $\mathscr{F}_t = \sigma(y(t), u(t), y(t-1), \cdots, u(0))$. The sequences $\{v(t)\}$ satisfy the noise assumptions [6]:

$$(A1) \qquad \mathrm{E}[v(t)|\mathscr{F}_{t-1}] = \mathbf{0}, \text{ a.s.}$$

$$(A2) \qquad \mathrm{E}[\|v(t)\|^2|\mathscr{F}_{t-1}] = \sigma_v^2(t) \le \sigma_v^2 < \infty, \text{ a.s.}$$

$$(A3) \qquad \limsup_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} \|v(i)\|^2 \le \sigma_v^2 < \infty, \text{ a.s.}$$

Lemma 1: If the symmetric matrix $A = [A_{ij}] \in \mathbb{R}^{b\times b}$, $A_{ij} = A_{ji}^{\mathrm{T}} \in \mathbb{R}^{n_i\times n_j}$, satisfies $c_1 I \le A \le c_2 I$, where c_1 and c_2 are positive constants, $n_1 + n_2 + \cdots + n_N = b$, then

$$c_1 I \le \mathrm{diag}[A_{ii},\ i = 1,2,\cdots,N] \le c_2 I,\ c_1 I \le A_{ii} \le c_2 I.$$

Proof The proof is easy and omitted here.

Lemma 2: For the system in (5), if there exist an integer $p \ge b := mn_0 + n$ and constants c_3 and c_4 such that the following strong persistent excitation condition holds,

$$(A4) \qquad c_3 I \le \frac{1}{p}\sum_{j=0}^{p-1} \Phi^{\mathrm{T}}(t-j)\Phi(t-j) \le c_4 I, \text{ a.s.}$$

According to Lemma 1 and (A4) we have

$$c_3 I \le \frac{1}{p}\sum_{j=0}^{p-1} \Phi_i^{\mathrm{T}}(t-j)\Phi_i(t-j) \le c_4 I, \text{ a.s.}$$

Then the covariance matrix $P_i(t)$ satisfies

$$(t-p)c_3 I \le P_i^{-1}(t) \le [c_4(t+p) + 1/p_0]I.$$

Proof From (12) we have

$$\begin{aligned} P_i^{-1}(t) &= P_i^{-1}(t-1) + \Phi_i^{\mathrm{T}}(t)\Phi_i(t) \\ &= P_i^{-1}(t-p) + \sum_{k=0}^{p-1} \Phi_i^{\mathrm{T}}(t-k)\Phi_i(t-k). \end{aligned}$$

Using Condition (A4) gives

$$P_i^{-1}(t-p) + pc_3 I \le P_i^{-1}(t) \le P_i^{-1}(t-p) + pc_4 I.$$

Let $t = pj + k$, $0 \le k < p$, so we have

$$\begin{aligned} P_i^{-1}(t = pj+k) &\le P_i^{-1}[p(j-1)+k] + pc_4 I \\ &\le P_i^{-1}[p(j-2)+k] + 2pc_4 I \\ &\le P_i^{-1}(k) + jpc_4 I \\ &\le P_i^{-1}(0) + (j+1)pc_4 I \\ &\le P_i^{-1}(0) + (t+p)c_4 I \\ &= [c_4(t+p) + 1/p_0]I. \end{aligned}$$

A similar derivation leads to

$$\begin{aligned} P_i^{-1}(t = pj+k) &\ge P_i^{-1}(t) + jpc_3 I \ge jpc_3 I \\ &\ge (t-p)c_3 I,\ t > p. \end{aligned}$$

This completes the proof of lemma 2. □

Lemma 3: Assume that there exist functions $f(t) \ge 0$ and $g(t) \ge 0$ such that $\lim_{t\to\infty} f(t) = f_0$, and series $\sum_{t=1}^{\infty} g(t)$ is divergent and $\sum_{t=1}^{\infty} f(t)g(t)$ is convergent, i.e., $\sum_{t=1}^{\infty} g(t) = \infty$ and $\sum_{t=1}^{\infty} f(t)g(t) < \infty$, then $f_0 = 0$.

Proof the proof is straightforward and is omitted here.

Theorem 1: For the system in (5) and the algorithm in (12) and (14), assume that (A1)-(A4) hold, then the parameter estimation error vector converges to zero, i.e., $\hat{\vartheta}_i(t) \to \vartheta_i$, or $\hat{\vartheta}(t) \to \vartheta$ as $t \to \infty$.

Proof Here, we apply martingale convergence theorem, which is the main tools of analyzing the convergence of recursive identification algorithms [5], [11] , to prove the theorem above. For convergence analysis, the HLS algorithm

in (12) and (14) needs to be transformed into an equivalent form. From Lemma 2, we have

$$\frac{I}{c_4(t+p)+1/p_0} \le P_i(t) \le \frac{I}{c_3(t-p)},\ t>p.$$

Thus, the algorithms (14) and (12) can be simplified as

$$\begin{aligned}\hat{\vartheta}_i(t) &= \hat{\vartheta}_i(t-1)+\bar{P}_i(t)\Phi_i^{\mathrm{T}}(t)\times \\ &\quad [y(t)-\Phi(t)\hat{\vartheta}(t-1)], \qquad (15)\\ \bar{P}_i(t) &= \begin{cases} I, & t\le p,\\ \frac{I}{\gamma t+c}, & t>p,\ c>0,\\ & 0<c_3\le\gamma\le c_4<\infty. \end{cases} \qquad (16)\end{aligned}$$

The following is to prove $\lim_{t\to\infty}\hat{\vartheta}_i(t)\to\vartheta_i$, or $\hat{\vartheta}(t)\to\vartheta$ for any $\gamma\in[c_3,c_4]$. Defined the parameter estimation error vector,

$$\tilde{\vartheta}_i(t) := \hat{\vartheta}_i(t)-\vartheta_i, \qquad (17)$$

and

$$\begin{aligned}\tilde{y}(t) &:= \Phi(t)\tilde{\vartheta}(t-1)=\Phi(t)\hat{\vartheta}(t-1)-\Phi(t)\vartheta\\ &= \sum_{i=1}^{n}\Phi_i(t)\tilde{\vartheta}_i(t-1). \qquad (18)\end{aligned}$$

Substituting (15) into (17) and using (16), (5) and (18), we have

$$\tilde{\vartheta}_i(t) = \tilde{\vartheta}_i(t-1)+\frac{\Phi_i^{\mathrm{T}}(t)}{\gamma t+c}[-\tilde{y}(t)+v(t)]. \qquad (19)$$

Taking the norm of both sides of (19) and the summation from $i=1$ to $i=N$ is

$$\begin{aligned}\|\tilde{\vartheta}(t)\|^2 &= \sum_{i=1}^{N}\|\tilde{\vartheta}_i(t-1)+\frac{\Phi_i^{\mathrm{T}}(t)}{\gamma t+c}[-\tilde{y}(t)+v(t)]\|^2\\ &= \|\tilde{\vartheta}(t-1)\|^2-\left[\frac{2}{\gamma t+c}-\frac{\|\Phi(t)\|^2}{(\gamma t+c)^2}\right]\|\tilde{y}(t)\|^2\\ &\quad +\frac{2}{\gamma t+c}\left[1-\frac{\|\Phi(t)\|^2}{\gamma t+c}\right]\tilde{y}^{\mathrm{T}}(t)v(t)\\ &\quad +\frac{\|\Phi(t)\|^2}{(\gamma t+c)^2}\|v(t)\|^2.\end{aligned}$$

Taking the trace to Condition (A4) yields $\|\Phi(t)\|^2\le c_4 p$. For large t, $\gamma t+c-\|\Phi(t)\|^2>0$, we have

$$\begin{aligned}\|\tilde{\vartheta}(t)\|^2 &\le \|\tilde{\vartheta}(t-1)\|^2-\frac{1}{\gamma t+c}\|\tilde{y}(t)\|^2+\frac{2}{\gamma t+c}\times\\ &\quad \left[1-\frac{\|\Phi(t)\|^2}{\gamma t+c}\right]\tilde{y}^{\mathrm{T}}(t)v(t)+\frac{\|\Phi(t)\|^2}{(\gamma t+c)^2}\|v(t)\|^2. \qquad (20)\end{aligned}$$

Using Conditions (A1)-(A3), since $\tilde{\vartheta}(t-1)$, $\tilde{\Phi}(t)$, $\tilde{y}(t)$ are uncorrelated with $v(t)$ and are $\mathscr{F}_{t-1}$ measurable, taking the conditional expectation of both sides of (20) with respect to $\mathscr{F}_{t-1}$ gives

$$\begin{aligned}\mathrm{E}[\|\tilde{\vartheta}(t)\|^2|\mathscr{F}_{t-1}] &\le \|\tilde{\vartheta}(t-1)\|^2-\frac{1}{\gamma t+c}\|\tilde{y}(t)\|^2\\ &\quad +\frac{c_4 p}{(\gamma t+c)^2}\sigma_v^2. \qquad (21)\end{aligned}$$

According to the martingale convergence theorem [6], [9], we can draw that $\|\tilde{\vartheta}(t)\|^2$ almost surely converges to a finite random variable, say C_0, i.e.,

$$\|\tilde{\vartheta}(t)\|^2\to C_0<\infty,\quad \text{a.s.} \qquad (22)$$

and

$$\lim_{t\to\infty}\sum_{j=1}^{t}\frac{1}{\gamma j+c}\|\tilde{y}(j)\|^2<\infty,\quad \text{a.s.} \qquad (23)$$

Further, from Equation (19) we have

$$\begin{aligned}\tilde{\vartheta}_i(t+j) &= \tilde{\vartheta}_i(t-1)+\sum_{k=0}^{j}\frac{\Phi_i^{\mathrm{T}}(t+k)}{\gamma t+c}\times\\ &\quad [-\tilde{y}(t+k)+v(t+k)],\ i=1,2,\cdots,n,\end{aligned}$$

which can be written a more compact form,

$$\begin{aligned}\tilde{\vartheta}(t+j) &= \tilde{\vartheta}(t-1)+\sum_{k=0}^{j}\frac{\Phi^{\mathrm{T}}(t+k)}{\gamma t+c}\times\\ &\quad [-\tilde{y}(t+k)+v(t+k)]\\ &=: \tilde{\vartheta}(t-1)+\sum_{k=0}^{j}\Delta\tilde{\theta}(t+k), \qquad (24)\end{aligned}$$

where

$$\Delta\tilde{\theta}(t) := \frac{\Phi^{\mathrm{T}}(t)}{\gamma t+c}[-\tilde{y}(t)+v(t)]. \qquad (25)$$

From (18) we have

$$\tilde{y}(t+j)=\Phi(t+j)\tilde{\vartheta}(t+j-1).$$

Substituting (24) into the above equation gives

$$\Phi(t+j)\tilde{\vartheta}(t-1)=\tilde{y}(t+j)-\Phi(t+j)\sum_{k=0}^{j-1}\Delta\tilde{\theta}(t+k).$$

Taking the norm $\|*\|$ and summing from $j=0$ to $j=p-1$ gives

$$\begin{aligned}&\tilde{\vartheta}^{\mathrm{T}}(t-1)\left[\sum_{j=0}^{p-1}\Phi^{\mathrm{T}}(t+j)\Phi(t+j)\right]\tilde{\vartheta}(t-1)\\ &=\sum_{j=0}^{p-1}\left\|\tilde{y}(t+j)-\Phi(t+j)\sum_{k=0}^{j-1}\Delta\tilde{\theta}(t+k)\right\|^2.\end{aligned}$$

Using Condition (A4), we have

$$\begin{aligned}\|\tilde{\vartheta}(t-1)\|^2 &\le \frac{2}{c_3 p}\sum_{j=0}^{p-1}\|\tilde{y}(t+j)\|^2\\ &\quad +\frac{2c_4}{c_3}\sum_{j=0}^{p-1}\sum_{k=0}^{j-1}\left\|\Delta\tilde{\theta}(t+k)\right\|^2.\end{aligned}$$

Dividing by $\gamma l+c$ and the summing from $l=l_0$ to $l=t$ yield

$$\begin{aligned}\sum_{l=l_0}^{t}\frac{\|\tilde{\vartheta}(l-1)\|^2}{\gamma l+c} &\le \frac{2}{c_3 p}\sum_{j=0}^{p-1}\sum_{l=l_0}^{t}\frac{\|\tilde{y}(t+j)\|^2}{\gamma l+c}+\frac{2c_4}{c_3}\times\\ &\quad \sum_{j=0}^{p-1}\sum_{k=0}^{j-1}\sum_{l=l_0}^{t}\frac{\|\Delta\tilde{\theta}(t+k)\|^2}{\gamma l+c}. \qquad (26)\end{aligned}$$

Also, we have

$$\begin{aligned}
&\lim_{t\to\infty}\sum_{l=l_0}^{t}\frac{\|\Delta\tilde{\theta}(l+k)\|^2}{\gamma l+c}\\
&\quad=\sum_{l=l_0}^{t}\frac{1}{\gamma l+c}\|\frac{\Phi^{\mathrm{T}}(l+k)}{\gamma(l+k)+c}\times\\
&\qquad[-\tilde{y}(l+k)+v(l+k)]\|^2\\
&\quad\le\sum_{l=l_0}^{t}\frac{2\|\tilde{y}(l+k)\|^2}{\gamma(l+k)+c}\\
&\qquad+\sum_{l=l_0}^{t}\frac{2\|v(l+k)\|^2}{(\gamma l+c)[\gamma(l+k)+c]}<\infty,\ \text{a.s.}\quad(27)
\end{aligned}$$

Using (23) and (27), from (26) we have

$$\lim_{t\to\infty}\sum_{l=l_0}^{t}\frac{\|\tilde{\vartheta}(l-1)\|^2}{\gamma l+c}<\infty,\ \text{a.s.}\qquad(28)$$

Since $\|\tilde{\vartheta}(t)\|^2$ almost surely converges to a random variable C_0, using Lemma 3 leads to $C_0\to 0$ as $t\to\infty$. That means that the estimate $\hat{\vartheta}(t)$ converges to the true parameter vector ϑ. This completes the proof of Theorem 1. □

IV. Simulation tests

Consider a 2-input, 2-output stochastic system,

$$\begin{aligned}
&\begin{bmatrix}y_1(t)\\y_2(t)\end{bmatrix}-\alpha_1\begin{bmatrix}y_1(t-1)\\y_2(t-1)\end{bmatrix}-\alpha_2\begin{bmatrix}y_1(t-2)\\y_2(t-2)\end{bmatrix}\\
&\qquad=Q_1\begin{bmatrix}u_1(t-1)\\u_2(t-1)\end{bmatrix}+Q_2\begin{bmatrix}u_1(t-2)\\u_2(t-2)\end{bmatrix}+\begin{bmatrix}v_1(t)\\v_2(t)\end{bmatrix},
\end{aligned}$$

where $m=n=r=2$,

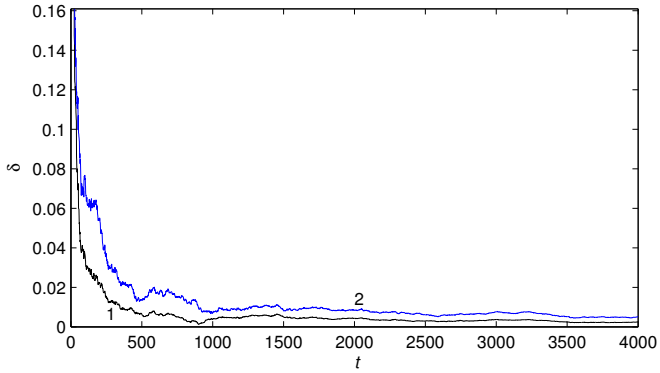
$$\begin{aligned}
\alpha &= \begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix}=\begin{bmatrix}0.5\\0.8\end{bmatrix},\\
\theta^{\mathrm{T}} &= Q=[Q_1,Q_2]=[Q(i,j)]=\begin{bmatrix}2.0&0.6&1.3&1.0\\1.0&1.2&1.5&2.0\end{bmatrix}.
\end{aligned}$$

Here $\{u_1(t)\}$ and $\{u_2(t)\}$ were taken as two persistent excitation signal sequences with zero mean and unit variances, $\{v_1(t)\}$ and $\{v_2(t)\}$ as two white noise sequences with zero mean and variances $\sigma^2(1)$ and $\sigma^2(2)$. The system contains $mn_0+r=10$ parameters and is decomposed into $N=2$ subsystem (thus $n_1=n_2=5$). The proposed algorithm is applied to estimate the parameters, the parameter estimation error $\delta:=\|\hat{\vartheta}(t)-\vartheta\|/\|\vartheta\|$ versus t is shown in Figure 1. When the noise variances are $\sigma^2(1)=0.20^2$ and $\sigma^2(2)=0.40^2$, the noise-to-signal ratios of two output channels are $\delta_{\mathrm{ns}}(1)=8.71\%$ and $\delta_{\mathrm{ns}}(2)=15.66\%$; when $\sigma^2(1)=0.60^2$ and $\sigma^2(2)=0.80^2$, $\delta_{\mathrm{ns}}(1)=26.13\%$ and $\delta_{\mathrm{ns}}(2)=31.32\%$.

From Figure 1, it is clear that the estimation errors become generally small as the data length increasing, and the estimation accuracy become high for smaller noise variances.

V. Conclusion

This paper presents the hierarchical least squares identification algorithm to reduce the computational burden for multivariable systems according to the hierarchical identification principle. Compared with the existing identification algorithms, the proposed algorithm has small computational load and is easy to implement on computer. The convergence analysis of the algorithm using the martingale convergence theorem is done only for its simplified version. Finally, we verify the theoretical findings through simulation.



1: $\sigma^2(1)=0.2^2$, $\sigma^2(2)=0.4^2$;
2: $\sigma^2(1)=0.6^2$, $\sigma^2(2)=0.8^2$.

Fig. 1. The HLS estimation errors δ versus t with the different $\sigma^2(i)$

References

[1] F. Ding and T. Chen, "Hierarchical gradient-based identification of multivariable discrete-time systems," *Automatica*, vol. 41, no. 2, pp. 315-325, 2005.

[2] F. Ding and T. Chen, "Hierarchical least squares identification methods for multivariable systems," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 397-402, 2005.

[3] F. Ding and T. Ding, "Hierarchical least squares identification and convergence for large-scale multivariable systems," *Tsinghua Science and Technology*, vol. 7, no. 3, pp. 299-304, 2002.

[4] F. Ding and J. Yang, "Hierarchical identification of large scale systems," *Acta Automatica Sinica*, vol. 25, no.5, pp. 647-654, 1999.

[5] F. Ding, D.Q. Zhu and B. Zou, "Convergence analysis of hierarchical least squares algorithms for MIMO systems," *Dynamics of Continuous Discrete and Impulsive Systems: Series A-Mathematical Analysis*, vol. 13, no. 1, pp. 23-31, 2006.

[6] G.C. Goodwin and K.S. Sin, *Adaptive filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[7] L. Ljung, *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, 1999.

[8] A. Sen and N.K. Sinha, "On-line estimation of the parameters of a multivariable system using matrix pseudoinverse," *International Journal of Systems Science*, vol. 7, no. 4, pp. 461C471, 1976.

[9] V. Solo, "The convergence of AML," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 958-962, 1979.

[10] V. Verdult and M. Verhaegen, "Subspace identification of multivariable linear parameter-varying systems," *Automatica*, vol. 38, no. 5, pp. 805-814, 2002.

[11] L. Wang, F. Ding and P.X. Liu, "Consistency of HLS estimation algorithms for MIMO ARX-like systems," *Applied Mathematics and Computation*, vol. 190, no. 2, pp. 1081-1093, 2007.

[12] P. Yuan and F. Ding, "Hierarchical Least Squares Parameter Estimation for Multivariable ARX-like Systems," *Science Techhnology and Engineering*, vol. 8 , no. 4, pp. 1005-1007, 2008.

[13] C.J. Zhang and R. Bitmead, "Subspace system identification for training-based MIMO channel estimation," *Automatica*, vol. 41, no. 9, pp. 1623-1632, 2005.