Automatica 71 (2016) 308–313

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Technical communique

Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model ☆

Yanjiao Wang, Feng Ding

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form
22 March 2016
Accepted 6 May 2016
Available online 10 June 2016

Keywords:
Recursive identification
Least squares
Auxiliary model
Hierarchical identification
MIMO system

ABSTRACT

This communique uses the auxiliary model method to study the identification problem of a multiple-input multiple-output (MIMO) system. For such a MIMO system whose outputs are contaminated by an ARMA noise process (i.e., correlated noise), an auxiliary model based recursive least squares parameter estimation algorithm is presented through filtering input–output data. The proposed algorithm has higher estimation accuracy than the existing multivariable identification algorithm. The simulation example is given.

1. Introduction

Multiple-input multiple-output (MIMO) systems widely exist in industrial processes (Zhang, Shi, & Saadat, 2011). For example, Wang, Ding, and Zhu (2013) designed a multivariable controller for linear time-invariant MIMO systems through optimizing controller parameters. Mercèrea and Bako (2011) studied the parameterization and identification method for a MIMO canonical state-space model from data directly. Also, a hierarchical gradient-based identification algorithm was proposed for multivariable discrete-time systems (Ding & Chen, 2005).

The filtering technique has attracted much attention due to its potential for solving many problems in signal processing and analysis (Qaisar, Fesquet, & Renaudin, 2014), communication and system identification. Recently, a filtering based least squares algorithm has been developed for an equation-error MIMO system whose disturbance is an autoregressive (AR) process (Wang et al., 2013); a filtering based hierarchical stochastic gradient algorithm and two filtering based hierarchical iterative algorithms have been presented for multivariable systems (Wang & Ding, 2016a,b). On the basis of these work, this communique investigates novel parameter estimation methods for an output-error MIMO systems whose disturbance is an ARMA noise using the auxiliary model and data filtering methods. The main contributions of this work are as follows.

- This communique derives a filtering based auxiliary model recursive least squares (AM-RLS) algorithm for output-error MIMO systems with ARMA noise through filtering input–output data.
- Compared with the AM-RLS algorithm, the proposed filtering based AM-RLS algorithm has higher estimation accuracy because it uses the filtered input–output data and uses the outputs of the auxiliary model to replace the unknown variables.
- This work is based on the output-error MIMO model with ARMA noise and differs from the previous work with AR noise in Wang et al. (2013), because estimating the parameters of an AR process is a linear problem, while estimating the parameters of an ARMA process is a nonlinear one Li, Zhu, and Dickinson (1989).
- The convergence of the proposed algorithm is analyzed using the stochastic process theory.

The communique is organized as follows. Section 2 derives a filtering based auxiliary model recursive least squares (F-AM-RLS) identification algorithm for MIMO systems and analyzes its convergence. Section 3 compares the proposed algorithm with the existing algorithm. Section 4 gives a simulation example. Finally, the concluding remarks are given in Section 5.

☆ This work was supported in part by the National Natural Science Foundation of China (No. 61273194). The material in this paper was partially presented at the 9th International Symposium on Advanced Control of Chemical Processes, June 7–10, 2015, Whistler, British Columbia, Canada. This paper was recommended for publication in revised form by Associate Editor A. Pedro Aguiar under the direction of Editor André L. Tits.

E-mail addresses: yjwang12@126.com (Y. Wang), fding@jiangnan.edu.cn (F. Ding).

http://dx.doi.org/10.1016/j.automatica.2016.05.024
0005-1098/

2. The F-AM-RLS algorithm

Consider the following MIMO system with ARMA noise,

$$\boldsymbol{y}(t)=\boldsymbol{x}(t)+\boldsymbol{w}(t), \tag{1}$$

$$\boldsymbol{x}(t)=\boldsymbol{G}(z)\boldsymbol{u}(t), \tag{2}$$

where $\boldsymbol{y}(t)\in\mathbb{R}^m$ is the output vector, $\boldsymbol{u}(t)\in\mathbb{R}^r$ is the input vector, $\boldsymbol{w}(t):=H(z)\boldsymbol{v}(t)\in\mathbb{R}^m$ is an ARMA noise process, $\boldsymbol{v}(t)\in\mathbb{R}^m$ is a white noise vector with zero mean, the rational fractions $\boldsymbol{G}(z):=\boldsymbol{\beta}(z)/\alpha(z)$ and $H(z):=d(z)/c(z)$ are the transfer matrix/function, z^{-1} is a unit backward shift operator, and

$$\alpha(z):=1+\alpha_1 z^{-1}+\alpha_1 z^{-2}+\cdots+\alpha_n z^{-n},$$

$$\boldsymbol{\beta}(z):=\boldsymbol{\beta}_1 z^{-1}+\boldsymbol{\beta}_2 z^{-2}+\cdots+\boldsymbol{\beta}_n z^{-n},$$

$$c(z):=1+c_1 z^{-1}+c_2 z^{-2}+\cdots+c_{n_c} z^{-n_c},$$

$$d(z):=1+d_1 z^{-1}+d_2 z^{-2}+\cdots+d_{n_d} z^{-n_d}.$$

Without loss of generality, suppose that the structure parameters n, n_c and n_d are predetermined and $\boldsymbol{y}(t)=\boldsymbol{0}$, $\boldsymbol{u}(t)=\boldsymbol{0}$, $\boldsymbol{w}(t)=\boldsymbol{0}$ and $\boldsymbol{v}(t)=\boldsymbol{0}$ for $t\leqslant 0$. The available input–output data are $\{\boldsymbol{u}(t),\boldsymbol{y}(t)\}$.

Define the parameter vectors/matrices and the information vectors/matrices:

$$\boldsymbol{\theta}^{\mathrm{T}}:=[\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_n]\in\mathbb{R}^{m\times(nr)},$$

$$\boldsymbol{\alpha}:=[\alpha_1,\alpha_2,\ldots,\alpha_n]^{\mathrm{T}}\in\mathbb{R}^n,$$

$$\boldsymbol{\rho}:=[c_1,c_2,\ldots,c_{n_c},d_1,d_2,\ldots,d_{n_d}]^{\mathrm{T}}\in\mathbb{R}^{n_c+n_d},$$

$$\boldsymbol{\varphi}(t):=[\boldsymbol{u}^{\mathrm{T}}(t-1),\boldsymbol{u}^{\mathrm{T}}(t-2),\ldots,\boldsymbol{u}^{\mathrm{T}}(t-n)]^{\mathrm{T}}\in\mathbb{R}^{nr},$$

$$\boldsymbol{\zeta}(t):=[-\boldsymbol{x}(t-1),-\boldsymbol{x}(t-2),\ldots,-\boldsymbol{x}(t-n)]\in\mathbb{R}^{m\times n},$$

$$\boldsymbol{\xi}(t):=[-\boldsymbol{w}(t-1),-\boldsymbol{w}(t-2),\ldots,-\boldsymbol{w}(t-n_c),$$
$$\boldsymbol{v}(t-1),\boldsymbol{v}(t-2),\ldots,\boldsymbol{v}(t-n_d)]\in\mathbb{R}^{m\times(n_c+n_d)}.$$

Then, Eqs. (1)–(2) can be rewritten as

$$\boldsymbol{x}(t)=\boldsymbol{\zeta}(t)\boldsymbol{\alpha}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t), \tag{3}$$

$$\boldsymbol{w}(t)=\boldsymbol{\xi}(t)\boldsymbol{\rho}+\boldsymbol{v}(t), \tag{4}$$

$$\boldsymbol{y}(t)=\boldsymbol{\zeta}(t)\boldsymbol{\alpha}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t)+\boldsymbol{w}(t). \tag{5}$$

Remark 1. For the identification model in (3)–(5), the input–output data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$ are available. That is, only $\boldsymbol{y}(t)$ and $\boldsymbol{\varphi}(t)$ are available but $\boldsymbol{\zeta}(t)$ and $\boldsymbol{\xi}(t)$ are unavailable and unknown.

Remark 2. The difficulty of identification is that the information matrix $\boldsymbol{\zeta}(t)$ contains the unknown inner variable $\boldsymbol{x}(t-i)$. The solution here is to construct an auxiliary model using the measured data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, and to replace the unknown $\boldsymbol{x}(t-i)$ in the identification algorithm with the output $\boldsymbol{x}_{\mathrm{a}}(t-i)$ of the auxiliary model $\boldsymbol{G}_{\mathrm{a}}(z)$ in Fig. 1.

Remark 3. From (5), we can see that the output $\boldsymbol{y}(t)$ contains the correlated noise $\boldsymbol{w}(t)$, which results in biased estimates. In this work, we use the filtering technique and combine the auxiliary model for investigating a novel identification method. The details are as follows.

We use a linear filter $L(z):=H^{-1}(z)$ to filter the input–output data $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$, and $\boldsymbol{\zeta}(t)$ and $\boldsymbol{\varphi}(t)$, leading to the filtered variables $\boldsymbol{y}_{\mathrm{f}}(t):=L(z)\boldsymbol{y}(t)\in\mathbb{R}^m$, $\boldsymbol{u}_{\mathrm{f}}(t):=L(z)\boldsymbol{u}(t)\in\mathbb{R}^r$, $\boldsymbol{\zeta}_{\mathrm{f}}(t):=L(z)\boldsymbol{\zeta}(t)\in\mathbb{R}^{m\times n}$ and $\boldsymbol{\varphi}_{\mathrm{f}}(t):=L(z)\boldsymbol{\varphi}(t)\in\mathbb{R}^{nr}$. Because $H(z)$ is to be identified and unknown, the proposed algorithm has to be implemented through the recursive/iterative scheme using the estimate of $L(z)$. Multiplying both sides of (5) by $L(z)$, we obtain a filtered identification model,

$$\boldsymbol{y}_{\mathrm{f}}(t)=\boldsymbol{\zeta}_{\mathrm{f}}(t)\boldsymbol{\alpha}+\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}_{\mathrm{f}}(t)+\boldsymbol{v}(t). \tag{6}$$

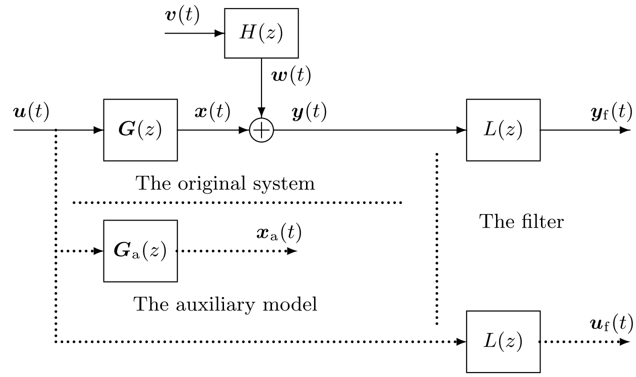

Fig. 1. The MIMO system with an auxiliary model.

Remark 4. Because the filtered output $\boldsymbol{y}_{\mathrm{f}}(t)$ in (6) involves only the white noise $\boldsymbol{v}(t)$, we will derive a new algorithm for estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ using the filtered input–output data $\boldsymbol{u}_{\mathrm{f}}(t)$ and $\boldsymbol{y}_{\mathrm{f}}(t)$ (i.e., $\boldsymbol{\zeta}_{\mathrm{f}}(t)$ and $\boldsymbol{\varphi}_{\mathrm{f}}(t)$)—see Fig. 1.

The filtered identification model in (6) contains both a parameter vector $\boldsymbol{\alpha}$ and a parameter matrix $\boldsymbol{\theta}$. One solution is to derive the estimation algorithm of the parameter vector $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\theta}$, and to derive the estimation algorithm of the parameter matrix $\boldsymbol{\theta}$ for fixed $\boldsymbol{\alpha}$ using the hierarchical identification principle in Ding and Chen (2005).

Let $\hat{\boldsymbol{\alpha}}(t)$, $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{\rho}}(t)$ be the estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ at time t, respectively. Based on the identification model in (6), for fixed $\boldsymbol{\theta}$ (that is $\boldsymbol{\theta}$ is regarded as being known in this step), defining and minimizing a quadratic criterion function, we can obtain the following recursive least squares relation for $\hat{\boldsymbol{\alpha}}(t)$ (Goodwin & Sin, 1984; Ljung, 1999):

$$\hat{\boldsymbol{\alpha}}(t)=\hat{\boldsymbol{\alpha}}(t-1)+\boldsymbol{P}_{\alpha}(t)\boldsymbol{\zeta}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{e}_{\alpha}(t), \tag{7}$$

$$\boldsymbol{e}_{\alpha}(t)=\boldsymbol{y}_{\mathrm{f}}(t)-\boldsymbol{\zeta}_{\mathrm{f}}(t)\hat{\boldsymbol{\alpha}}(t-1)-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}_{\mathrm{f}}(t), \tag{8}$$

$$\boldsymbol{P}_{\alpha}^{-1}(t)=\boldsymbol{P}_{\alpha}^{-1}(t-1)+\boldsymbol{\zeta}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{\zeta}_{\mathrm{f}}(t). \tag{9}$$

Similarly, for fixed $\boldsymbol{\alpha}$, defining and minimizing a quadratic criterion function, we can obtain the following recursive least squares relation for $\hat{\boldsymbol{\theta}}(t)$:

$$\hat{\boldsymbol{\theta}}(t)=\hat{\boldsymbol{\theta}}(t-1)+\boldsymbol{P}_{\theta}(t)\boldsymbol{\varphi}_{\mathrm{f}}(t)\boldsymbol{e}_{\theta}^{\mathrm{T}}(t), \tag{10}$$

$$\boldsymbol{e}_{\theta}(t)=\boldsymbol{y}_{\mathrm{f}}(t)-\boldsymbol{\zeta}_{\mathrm{f}}(t)\boldsymbol{\alpha}-\hat{\boldsymbol{\theta}}^{\mathrm{T}}(t-1)\boldsymbol{\varphi}_{\mathrm{f}}(t), \tag{11}$$

$$\boldsymbol{P}_{\theta}^{-1}(t)=\boldsymbol{P}_{\theta}^{-1}(t-1)+\boldsymbol{\varphi}_{\mathrm{f}}(t)\boldsymbol{\varphi}_{\mathrm{f}}^{\mathrm{T}}(t). \tag{12}$$

Based on the identification model in (4), we can obtain the recursive relation for $\hat{\boldsymbol{\rho}}(t)$:

$$\begin{aligned}\hat{\boldsymbol{\rho}}(t)&=\hat{\boldsymbol{\rho}}(t-1)+\boldsymbol{P}_{\rho}(t)\boldsymbol{\xi}^{\mathrm{T}}(t)[\boldsymbol{w}(t)-\boldsymbol{\xi}(t)\hat{\boldsymbol{\rho}}(t-1)]\\&=\hat{\boldsymbol{\rho}}(t-1)+\boldsymbol{P}_{\rho}(t)\boldsymbol{\xi}^{\mathrm{T}}(t)\\&\quad\times[\boldsymbol{y}(t)-\boldsymbol{\zeta}(t)\boldsymbol{\alpha}-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\varphi}(t)-\boldsymbol{\xi}(t)\hat{\boldsymbol{\rho}}(t-1)],\end{aligned} \tag{13}$$

$$\boldsymbol{P}_{\rho}^{-1}(t)=\boldsymbol{P}_{\rho}^{-1}(t-1)+\boldsymbol{\xi}^{\mathrm{T}}(t)\boldsymbol{\xi}(t). \tag{14}$$

$\boldsymbol{P}_{\alpha}(t)$, $\boldsymbol{P}_{\theta}(t)$ and $\boldsymbol{P}_{\rho}(t)$ are the covariance matrices. However, Eqs. (7)–(14) cannot be implemented because their right-hand sides contain the unknown variables $\boldsymbol{y}_{\mathrm{f}}(t)$, $\boldsymbol{\zeta}_{\mathrm{f}}(t)$, $\boldsymbol{\varphi}_{\mathrm{f}}(t)$, $\boldsymbol{w}(t)$, $\boldsymbol{\xi}(t)$, $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$.

Since the filter $L(z)$ and the inner variable $\boldsymbol{x}(t)$ are unknown, so are the filtered variables $\boldsymbol{y}_{\mathrm{f}}(t)$, $\boldsymbol{\zeta}_{\mathrm{f}}(t)$ and $\boldsymbol{\varphi}_{\mathrm{f}}(t)$. To solve this problem, referring to Fig. 1, we use $\boldsymbol{G}_{\mathrm{a}}(z)$ to generate the estimate $\boldsymbol{x}_{\mathrm{a}}(t)$ of $\boldsymbol{x}(t)$ by means of the auxiliary model or reference model method, and use the estimate $\hat{L}(t,z)$ of $L(z)$ at time t to generate the estimates $\hat{\boldsymbol{y}}_{\mathrm{f}}(t)$, $\hat{\boldsymbol{\zeta}}_{\mathrm{f}}(t)$ and $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t)$ of $\boldsymbol{y}_{\mathrm{f}}(t)$, $\boldsymbol{\zeta}_{\mathrm{f}}(t)$ and $\boldsymbol{\varphi}_{\mathrm{f}}(t)$:

$$\hat{\boldsymbol{y}}_{\mathrm{f}}(t):=\hat{L}(t,z)\boldsymbol{y}(t),\qquad \hat{\boldsymbol{u}}_{\mathrm{f}}(t):=\hat{L}(t,z)\boldsymbol{u}(t), \tag{15}$$

$$\hat{\zeta}_{\rm f}(t) := \hat{L}(t,z)\zeta(t), \qquad \hat{\varphi}_{\rm f}(t) := \hat{L}(t,z)\varphi(t), \tag{16}$$

$$\hat{L}(t,z) := \frac{1+\hat{c}_1(t)z^{-1}+\hat{c}_2(t)z^{-2}+\cdots+\hat{c}_{n_c}(t)z^{-n_c}}{1+\hat{d}_1(t)z^{-1}+\hat{d}_2(t)z^{-2}+\cdots+\hat{d}_{n_d}(t)z^{-n_d}}.$$

Use the estimates $\hat{\boldsymbol{w}}(t)$ and $\hat{\boldsymbol{v}}(t)$ of $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ and the output $\boldsymbol{x}_{\rm a}(t)$ of the auxiliary model $\boldsymbol{G}_{\rm a}(z)$ to construct the estimates $\hat{\boldsymbol{\xi}}(t)$ and $\hat{\boldsymbol{\zeta}}(t)$ of $\boldsymbol{\xi}(t)$ and $\boldsymbol{\zeta}(t)$:

$$\hat{\boldsymbol{\xi}}(t) := [-\hat{\boldsymbol{w}}(t-1), -\hat{\boldsymbol{w}}(t-2), \ldots, -\hat{\boldsymbol{w}}(t-n_c), \hat{\boldsymbol{v}}(t-1), \hat{\boldsymbol{v}}(t-2), \ldots, \hat{\boldsymbol{v}}(t-n_d)], \tag{17}$$

$$\hat{\boldsymbol{\zeta}}(t) := [-\boldsymbol{x}_{\rm a}(t-1), -\boldsymbol{x}_{\rm a}(t-2), \ldots, -\boldsymbol{x}_{\rm a}(t-n)]. \tag{18}$$

According to (3), we use the estimates $\hat{\boldsymbol{\zeta}}(t)$, $\hat{\boldsymbol{\alpha}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ to define the auxiliary model or reference model:

$$\boldsymbol{x}_{\rm a}(t) = \hat{\boldsymbol{\zeta}}(t)\hat{\boldsymbol{\alpha}}(t) + \hat{\boldsymbol{\theta}}^{\rm T}(t)\boldsymbol{\varphi}(t). \tag{19}$$

According to (1), we can obtain the estimate of $\boldsymbol{w}(t)$:

$$\hat{\boldsymbol{w}}(t) := \boldsymbol{y}(t) - \boldsymbol{x}_{\rm a}(t). \tag{20}$$

According to (4), we can obtain the estimate of $\boldsymbol{v}(t)$:

$$\hat{\boldsymbol{v}}(t) := \hat{\boldsymbol{w}}(t) - \hat{\boldsymbol{\xi}}(t)\hat{\boldsymbol{\rho}}(t). \tag{21}$$

Therefore, according to the hierarchical identification principle, replacing the unknown variables $\boldsymbol{y}_{\rm f}(t)$, $\boldsymbol{\zeta}_{\rm f}(t)$, $\boldsymbol{\varphi}_{\rm f}(t)$ and $\boldsymbol{\xi}(t)$ on the right-hand sides of (7)–(14) with their corresponding estimates $\hat{\boldsymbol{y}}_{\rm f}(t)$, $\hat{\boldsymbol{\zeta}}_{\rm f}(t)$, $\hat{\boldsymbol{\varphi}}_{\rm f}(t)$ and $\hat{\boldsymbol{\xi}}(t)$, replacing the unknown $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ with their estimates $\hat{\boldsymbol{\theta}}(t-1)$ and $\hat{\boldsymbol{\alpha}}(t-1)$ at the previous time $t-1$, and defining the innovation

$$\boldsymbol{e}(t) := \hat{\boldsymbol{y}}_{\rm f}(t) - \hat{\boldsymbol{\zeta}}_{\rm f}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\rm T}(t-1)\hat{\boldsymbol{\varphi}}_{\rm f}(t), \tag{22}$$

we can obtain the following least squares relations:

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}}(t-1) + \boldsymbol{P}_{\alpha}(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t), \tag{23}$$

$$\boldsymbol{P}_{\alpha}^{-1}(t) = \boldsymbol{P}_{\alpha}^{-1}(t-1) + \hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\hat{\boldsymbol{\zeta}}_{\rm f}(t), \tag{24}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{P}_{\theta}(t)\hat{\boldsymbol{\varphi}}_{\rm f}(t)\boldsymbol{e}^{\rm T}(t), \tag{25}$$

$$\boldsymbol{P}_{\theta}^{-1}(t) = \boldsymbol{P}_{\theta}^{-1}(t-1) + \hat{\boldsymbol{\varphi}}_{\rm f}(t)\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t), \tag{26}$$

$$\hat{\boldsymbol{\rho}}(t) = \hat{\boldsymbol{\rho}}(t-1) + \boldsymbol{P}_{\rho}(t)\hat{\boldsymbol{\xi}}^{\rm T}(t)[\boldsymbol{y}(t) - \hat{\boldsymbol{\zeta}}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\rm T}(t-1)\boldsymbol{\varphi}(t) - \hat{\boldsymbol{\xi}}(t)\hat{\boldsymbol{\rho}}(t-1)], \tag{27}$$

$$\boldsymbol{P}_{\rho}^{-1}(t) = \boldsymbol{P}_{\rho}^{-1}(t-1) + \hat{\boldsymbol{\xi}}^{\rm T}(t)\hat{\boldsymbol{\xi}}(t). \tag{28}$$

Eqs. (15)–(28) form the filtering based auxiliary model recursive least squares (F-AM-RLS) algorithm whose initial values $\hat{\boldsymbol{\alpha}}(0)$, $\hat{\boldsymbol{\theta}}(0)$ and $\hat{\boldsymbol{\rho}}(0)$ are taken as zero vectors or zero matrices of appropriate sizes, $\hat{\boldsymbol{y}}_{\rm f}(i)$, $\hat{\boldsymbol{\zeta}}_{\rm f}(i)$, $\hat{\boldsymbol{\varphi}}_{\rm f}(i)$, $\hat{\boldsymbol{\xi}}(i)$, $\boldsymbol{x}_{\rm a}(i)$, $\hat{\boldsymbol{w}}(i)$, $\hat{\boldsymbol{v}}(i)$, $\boldsymbol{u}(i)$ and $\boldsymbol{y}(i)$ as zero vectors or zero matrices of appropriate sizes for $i \leqslant 0$, and $\boldsymbol{P}_{\alpha}(0) = p_0\boldsymbol{I}$, $\boldsymbol{P}_{\theta}(0) = p_0\boldsymbol{I}$, $\boldsymbol{P}_{\rho}(0) = p_0\boldsymbol{I}$, $p_0 = 10^6$, $\boldsymbol{I}$ is an identity matrix of appropriate dimensions.

Remark 5. By using the linear filter $L(z)$, the MIMO model with ARMA noise is decomposed into a filtered model in (6) and a noise model in (4). The F-AM-RLS algorithm updates the estimates $\hat{\boldsymbol{\alpha}}(t)$, $\hat{\boldsymbol{\theta}}(t)$ and $\hat{\boldsymbol{\rho}}(t)$ by using the filtered output $\hat{\boldsymbol{y}}_{\rm f}(t)$ and the filtered information vector $\hat{\boldsymbol{\varphi}}_{\rm f}(t)$, the filtered information matrix $\hat{\boldsymbol{\zeta}}_{\rm f}(t)$ and the information matrix $\hat{\boldsymbol{\xi}}(t)$. Thus it is superior to the AM-RLS algorithm on the estimation accuracy.

Theorem 1. *For the system in (1) and (2), the identification models in (4) and (6) and the F-AM-RLS algorithm in (15)–(28), suppose that* $\{\boldsymbol{v}(t)\}$ *is a white noise sequence with zero mean and variance* σ^2*, i.e.,* $\mathrm{E}[\boldsymbol{v}(t)] = \boldsymbol{0}$, $\mathrm{E}[\|\boldsymbol{v}(t)\|^2] = \sigma^2$, $\mathrm{E}[\boldsymbol{v}(t)\boldsymbol{v}^{\rm T}(i)] = \boldsymbol{0}$ $(i \neq t)$*, and that there exist positive constants* η_i *such that for large* t*, the following persistent excitation conditions hold,*

$$\text{(C1)}\ \eta_1\boldsymbol{I} \leqslant \frac{1}{t}\sum_{j=1}^{t}\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(j)\hat{\boldsymbol{\zeta}}_{\rm f}(j) \leqslant \eta_2\boldsymbol{I}, \quad \text{a.s.},$$

$$\text{(C2)}\ \eta_3\boldsymbol{I} \leqslant \frac{1}{t}\sum_{j=1}^{t}\hat{\boldsymbol{\varphi}}_{\rm f}(j)\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(j) \leqslant \eta_4\boldsymbol{I}, \quad \text{a.s.},$$

$$\text{(C3)}\ \eta_5\boldsymbol{I} \leqslant \frac{1}{t}\sum_{j=1}^{t}\hat{\boldsymbol{\xi}}^{\rm T}(j)\hat{\boldsymbol{\xi}}(j) \leqslant \eta_6\boldsymbol{I}, \quad \text{a.s.}$$

Then the parameter estimation error given by the F-AM-RLS algorithm converges to zero.

The proof is given in the Appendix.

3. The AM-RLS algorithm

As a comparison, the following gives the auxiliary model based recursive least squares (AM-RLS) algorithm (Wang, Xu, & Ding, 2015) to demonstrate the superiority of the F-AM-RLS algorithm. For the MIMO system in (1) and (2), we define the parameter vector and information matrix as

$$\boldsymbol{\vartheta} := [\boldsymbol{\alpha}^{\rm T}, \boldsymbol{\rho}^{\rm T}]^{\rm T} \in \mathbb{R}^{n+n_c+n_d},$$

$$\boldsymbol{\psi}(t) := [\boldsymbol{\zeta}(t), \boldsymbol{\xi}(t)] \in \mathbb{R}^{m\times(n+n_c+n_d)}.$$

From (4)–(5), we have the following identification model

$$\boldsymbol{y}(t) = \boldsymbol{\psi}(t)\boldsymbol{\vartheta} + \boldsymbol{\theta}^{\rm T}\boldsymbol{\varphi}(t) + \boldsymbol{v}(t). \tag{29}$$

Referring to the derivation of the F-AM-RLS algorithm, we can obtain the following AM-RLS algorithm:

$$\hat{\boldsymbol{\vartheta}}(t) = \hat{\boldsymbol{\vartheta}}(t-1) + \boldsymbol{P}_1(t-1)\hat{\boldsymbol{\psi}}^{\rm T}(t)\boldsymbol{e}(t), \tag{30}$$

$$\boldsymbol{e}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\vartheta}}(t-1) - \hat{\boldsymbol{\theta}}^{\rm T}(t-1)\boldsymbol{\varphi}(t), \tag{31}$$

$$\boldsymbol{P}_1^{-1}(t) = \boldsymbol{P}_1^{-1}(t-1) + \hat{\boldsymbol{\psi}}^{\rm T}(t)\hat{\boldsymbol{\psi}}(t), \tag{32}$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{P}_2(t-1)\boldsymbol{\varphi}(t)\boldsymbol{e}^{\rm T}(t), \tag{33}$$

$$\boldsymbol{P}_2^{-1}(t) = \boldsymbol{P}_2^{-1}(t-1) + \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\rm T}(t), \tag{34}$$

$$\hat{\boldsymbol{\psi}}(t) = [\hat{\boldsymbol{\zeta}}(t), \hat{\boldsymbol{\xi}}(t)], \tag{35}$$

$$\hat{\boldsymbol{\zeta}}(t) = [-\boldsymbol{x}_{\rm a}(t-1), \boldsymbol{x}_{\rm a}(t-2), \ldots, -\boldsymbol{x}_{\rm a}(t-n)], \tag{36}$$

$$\hat{\boldsymbol{\xi}}(t) = [-\hat{\boldsymbol{w}}(t-1), -\hat{\boldsymbol{w}}(t-2), \ldots, -\hat{\boldsymbol{w}}(t-n_c), \hat{\boldsymbol{v}}(t-1), \hat{\boldsymbol{v}}(t-2), \ldots, \hat{\boldsymbol{v}}(t-n_d)], \tag{37}$$

$$\boldsymbol{x}_{\rm a}(t) = \hat{\boldsymbol{\zeta}}(t)\hat{\boldsymbol{\alpha}}(t) + \hat{\boldsymbol{\theta}}^{\rm T}(t)\boldsymbol{\varphi}(t), \tag{38}$$

$$\hat{\boldsymbol{w}}(t) = \boldsymbol{y}(t) - \boldsymbol{x}_{\rm a}(t), \tag{39}$$

$$\hat{\boldsymbol{v}}(t) = \boldsymbol{y}(t) - \hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\vartheta}}(t) - \hat{\boldsymbol{\theta}}^{\rm T}(t)\boldsymbol{\varphi}(t). \tag{40}$$

The initial values of the AM-RLS algorithm may be set by referring to the F-AM-RLS algorithm.

4. Example

Consider a two-input two-output system,

$$\boldsymbol{y}(t) = \frac{\boldsymbol{\beta}(z)}{\alpha(z)}\boldsymbol{u}(t) + \frac{d(z)}{c(z)}\boldsymbol{v}(t),$$

$$\alpha(z) = 1 + 0.22z^{-1},$$

Table 1
The AM-RLS estimates and errors ($\sigma^2 = 0.50^2$).

t	50	100	200	500	1000	True values
α_1	0.21585	0.23732	0.24574	0.23977	0.22938	0.22000
β_{11}	1.79075	1.79695	1.71612	1.65623	1.63078	1.62000
β_{12}	0.84566	0.94052	1.06014	1.07036	1.08288	1.10000
β_{21}	0.78608	0.86765	0.90644	0.99044	0.99633	1.00000
β_{22}	1.76976	1.63345	1.62782	1.59393	1.57765	1.58000
c_1	0.53055	0.62691	0.56064	0.49116	0.49486	0.49000
d_1	−0.01559	−0.00088	−0.12881	−0.28165	−0.36156	−0.44000
δ (%)	21.38811	19.23235	12.61638	5.98042	2.92668	

Table 2
The F-AM-RLS estimates and errors ($\sigma^2 = 0.50^2$).

t	50	100	200	500	1000	True values
α_1	0.21792	0.22431	0.22227	0.22097	0.21630	0.22000
β_{11}	1.59372	1.59027	1.62406	1.61615	1.61542	1.62000
β_{12}	1.15099	1.16451	1.14789	1.11842	1.10670	1.10000
β_{21}	1.01298	0.98926	0.98604	1.01047	1.01530	1.00000
β_{22}	1.54617	1.53349	1.55498	1.57592	1.57931	1.58000
c_1	0.33125	0.47949	0.46702	0.46091	0.49044	0.49000
d_1	−0.40306	−0.37678	−0.41106	−0.43496	−0.44603	−0.44000
δ (%)	6.31657	3.82809	2.40027	1.31592	0.66982	

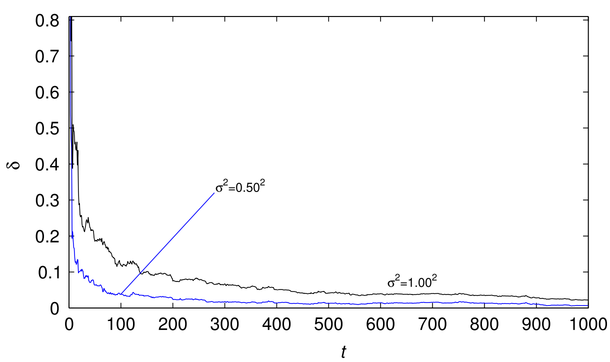

Fig. 2. The F-AM-RLS estimation errors δ versus t.

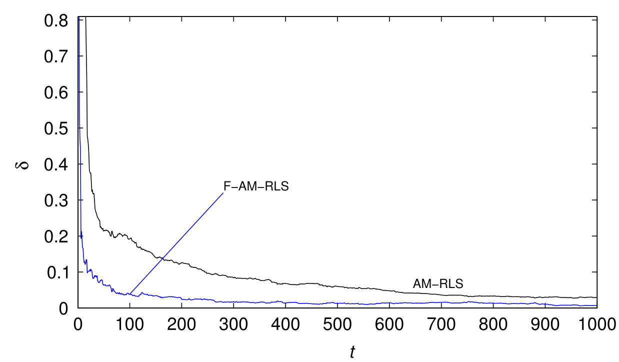

Fig. 3. The estimation errors δ versus t ($\sigma^2 = 0.50^2$).

$$\boldsymbol{\beta}(z) = \begin{bmatrix} 1.62z^{-1} & 1.10z^{-1} \\ 1.00z^{-1} & 1.58z^{-1} \end{bmatrix},$$

$$c(z) = 1 + 0.49z^{-1}, \qquad d(z) = 1 - 0.44z^{-1}.$$

In simulation, the inputs $\{u_1(t)\}$ and $\{u_2(t)\}$ are taken as two persistent excitation signal sequences with zero mean and unit variances, $\{v_1(t)\}$ and $\{v_2(t)\}$ are taken as two white noise sequences with zero mean and variances σ_1^2 for $v_1(t)$ and σ_2^2 for $v_2(t)$. Taking $\sigma_1^2 = \sigma_2^2 = \sigma^2$, applying the presented methods to estimate parameters of this system, the parameter estimates and the estimation errors are shown in Tables 1–3 and Figs. 2–3 with different σ^2, where δ is the relative parameter estimation error.

From Tables 1–3 and Figs. 2–3, we can draw the following conclusions.

- The parameter estimation errors become smaller (in general) with the increasing of t and the F-AM-RLS algorithm has higher parameter estimation accuracy than the AM-RLS algorithm.
- The parameter estimates given by the F-AM-RLS algorithm are closer to their true values for large t, compared with the AM-RLS algorithm.
- For lower noise levels, the parameter estimation errors given by the F-AM-RLS algorithm become small.

We use the F-AM-RLS estimates with the noise variance $\sigma^2 = 0.50^2$ and $t = 1000$ to construct the estimated model (Ljung, 1999):

$$\hat{\boldsymbol{y}}(t) = \hat{H}^{-1}(z)\hat{\boldsymbol{\beta}}(z)/\hat{\alpha}(z)\boldsymbol{u}(t) + (1 - \hat{H}^{-1}(z))\boldsymbol{y}(t),$$

$$\frac{\hat{\boldsymbol{\beta}}(z)}{\hat{\alpha}(z)} = \begin{bmatrix} \dfrac{1.61542z^{-1}}{1+0.21630z^{-1}} & \dfrac{1.10670z^{-1}}{1+0.21630z^{-1}} \\ \dfrac{1.01530z^{-1}}{1+0.21630z^{-1}} & \dfrac{1.57931z^{-1}}{1+0.21630z^{-1}} \end{bmatrix},$$

$$\hat{H}(z) = \frac{1 - 0.44603z^{-1}}{1 + 0.49044z^{-1}}.$$

For the model validation, we use this estimated model and the rest 100 data from $t = 1000$ to $t = 1100$ to compute the predicted outputs $\hat{y}_1(t)$ and $\hat{y}_2(t)$ of the system. The predicted outputs and the actual outputs are plotted in Fig. 4.

From Fig. 4, we can see that the predicted outputs are very close to the true outputs. In other words, the estimated model can capture the dynamics of the system.

5. Conclusions

In this communique, we employ the auxiliary model and the data filtering technique to present a novel F-AM-RLS algorithm for MIMO systems with ARMA noise. Compared with the AM-RLS algorithm, the F-AM-RLS algorithm can generate higher estimation accuracy due to filtering the input–output data of the system.

Table 3
The F-AM-RLS estimates and errors ($\sigma^2 = 1.00^2$).

t	50	100	200	500	1000	True values
α_1	0.17917	0.21844	0.20730	0.20679	0.19988	0.22000
β_{11}	1.42201	1.45223	1.58120	1.58557	1.58974	1.62000
β_{12}	1.20338	1.21170	1.16122	1.11823	1.10034	1.10000
β_{21}	0.82573	0.86402	0.90612	0.98383	1.00712	1.00000
β_{22}	1.58616	1.49450	1.50725	1.55464	1.56310	1.58000
c_1	0.07834	0.30257	0.34950	0.39653	0.45545	0.49000
d_1	−0.56929	−0.53247	−0.50623	−0.48733	−0.47173	−0.44000
δ (%)	18.52593	11.87145	7.47420	4.16820	2.22211	

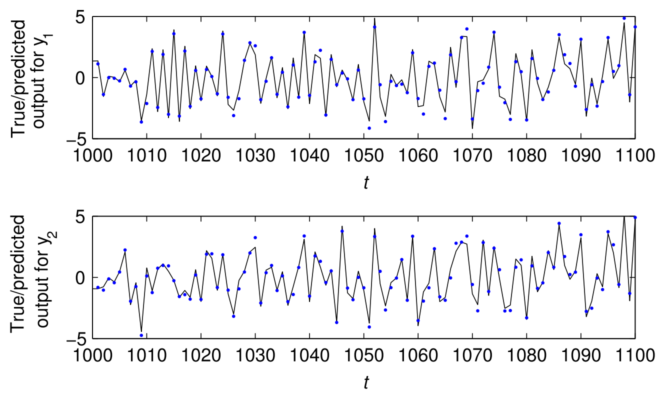

Fig. 4. The predicted outputs and the true outputs. Solid line: the true outputs, dots: the predicted outputs.

The proposed filtering based identification method can be applied to other multivariable systems with different structures and disturbance noise (e.g., ARMA noise).

Appendix. Proof

Proof of Theorem 1. Define the parameter estimation error vectors/matrix $\tilde{\boldsymbol{\alpha}}(t) := \hat{\boldsymbol{\alpha}}(t) - \boldsymbol{\alpha}$, $\tilde{\boldsymbol{\theta}}(t) := \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}$, $\tilde{\boldsymbol{\rho}}(t) := \hat{\boldsymbol{\rho}}(t) - \boldsymbol{\rho}$, and the vectors $\tilde{\boldsymbol{y}}_{\rm f}(t) := \hat{\boldsymbol{\zeta}}_{\rm f}(t)\tilde{\boldsymbol{\alpha}}(t-1) + \tilde{\boldsymbol{\theta}}^{\rm T}(t-1)\hat{\boldsymbol{\varphi}}_{\rm f}(t)$, $\tilde{\boldsymbol{w}}(t) := \hat{\boldsymbol{\xi}}(t)\tilde{\boldsymbol{\rho}}(t-1)$ and $\Delta_1(t) := \hat{\boldsymbol{y}}_{\rm f}(t) - \boldsymbol{y}_{\rm f}(t) + [\boldsymbol{\zeta}_{\rm f}(t) - \hat{\boldsymbol{\zeta}}_{\rm f}(t)]\boldsymbol{\alpha} + \boldsymbol{\theta}^{\rm T}[\boldsymbol{\varphi}_{\rm f}(t) - \hat{\boldsymbol{\varphi}}_{\rm f}(t)]$, $\Delta_2(t) := \hat{\boldsymbol{w}}(t) - \boldsymbol{w}(t) + [\boldsymbol{\xi}(t) - \hat{\boldsymbol{\xi}}(t)]\boldsymbol{\rho}$. Let

$$\begin{aligned}\boldsymbol{e}_v(t) &:= \boldsymbol{y}(t) - \hat{\boldsymbol{\zeta}}(t)\hat{\boldsymbol{\alpha}}(t-1) - \hat{\boldsymbol{\theta}}^{\rm T}(t-1)\boldsymbol{\varphi}(t) - \hat{\boldsymbol{\xi}}(t)\hat{\boldsymbol{\rho}}(t-1)\\ &= \hat{\boldsymbol{w}}(t) - \hat{\boldsymbol{\xi}}(t)\hat{\boldsymbol{\rho}}(t-1).\end{aligned}$$

Using (6) and (4), we have

$$\begin{aligned}\boldsymbol{e}(t) &:= -\tilde{\boldsymbol{y}}_{\rm f}(t) + \Delta_1(t) + \boldsymbol{v}(t),\\ \boldsymbol{e}_v(t) &:= -\tilde{\boldsymbol{w}}(t) + \Delta_2(t) + \boldsymbol{v}(t).\end{aligned}$$

Subtracting $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ from (23), (25) and (27) gives

$$\begin{aligned}\tilde{\boldsymbol{\alpha}}(t) &= \tilde{\boldsymbol{\alpha}}(t-1) + \boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t),\\ \tilde{\boldsymbol{\theta}}(t) &= \tilde{\boldsymbol{\theta}}(t-1) + \boldsymbol{P}_\theta(t)\hat{\boldsymbol{\varphi}}_{\rm f}(t)\boldsymbol{e}^{\rm T}(t),\\ \tilde{\boldsymbol{\rho}}(t) &= \tilde{\boldsymbol{\rho}}(t-1) + \boldsymbol{P}_\rho(t)\hat{\boldsymbol{\xi}}^{\rm T}(t)\boldsymbol{e}_v(t).\end{aligned}$$

Define the non-negative functions:

$$\begin{aligned}&T_1(t) := \tilde{\boldsymbol{\alpha}}^{\rm T}(t)\boldsymbol{P}_\alpha^{-1}(t)\tilde{\boldsymbol{\alpha}}(t),\\ &T_2(t) := {\rm tr}[\tilde{\boldsymbol{\theta}}^{\rm T}(t)\boldsymbol{P}_\theta^{-1}(t)\tilde{\boldsymbol{\theta}}(t)], \qquad T_{12}(t) := T_1(t) + T_2(t),\\ &T_3(t) := \tilde{\boldsymbol{\rho}}^{\rm T}(t)\boldsymbol{P}_\rho^{-1}(t)\tilde{\boldsymbol{\rho}}(t), \qquad T(t) := T_{12}(t) + T_3(t),\\ &\gamma_1(t) := {\rm tr}[\hat{\boldsymbol{\zeta}}_{\rm f}(t)\boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)] + \hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\boldsymbol{P}_\theta(t)\hat{\boldsymbol{\varphi}}_{\rm f}(t),\\ &\gamma_2(t) := {\rm tr}[\hat{\boldsymbol{\xi}}(t)\boldsymbol{P}_\rho(t)\hat{\boldsymbol{\xi}}^{\rm T}(t)].\end{aligned}$$

Then, we have

$$\begin{aligned}T_1(t) &= \{\tilde{\boldsymbol{\alpha}}(t-1) + \boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t)\}^{\rm T}\boldsymbol{P}_\alpha^{-1}(t)\\ &\quad\times\{\tilde{\boldsymbol{\alpha}}(t-1) + \boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t)\},\\ &= \tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)\boldsymbol{P}_\alpha^{-1}(t)\tilde{\boldsymbol{\alpha}}(t-1) + 2\tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t)\\ &\quad + \boldsymbol{e}^{\rm T}(t)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t)\\ &= \tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)[\boldsymbol{P}_\alpha^{-1}(t-1) + \hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\hat{\boldsymbol{\zeta}}_{\rm f}(t)]\tilde{\boldsymbol{\alpha}}(t-1)\\ &\quad + 2\tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t) + \boldsymbol{e}^{\rm T}(t)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t)\\ &= T_1(t-1) + \tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\tilde{\boldsymbol{\alpha}}(t-1) + 2\tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)\\ &\quad\times\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t) + \boldsymbol{e}^{\rm T}(t)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t)\\ &\le T_1(t-1) + \|\hat{\boldsymbol{\zeta}}_{\rm f}(t)\tilde{\boldsymbol{\alpha}}(t-1)\|^2 + 2\tilde{\boldsymbol{\alpha}}^{\rm T}(t-1)\\ &\quad\times\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t) + {\rm tr}[\hat{\boldsymbol{\zeta}}_{\rm f}(t)\boldsymbol{P}_\alpha(t)\hat{\boldsymbol{\zeta}}_{\rm f}^{\rm T}(t)]\|\boldsymbol{e}(t)\|^2.\end{aligned}$$

Similarly, we have

$$\begin{aligned}T_2(t) &= T_2(t-1) + \|\tilde{\boldsymbol{\theta}}^{\rm T}(t-1)\hat{\boldsymbol{\varphi}}_{\rm f}(t)\|^2 + 2\boldsymbol{e}^{\rm T}(t)\\ &\quad\times\tilde{\boldsymbol{\theta}}^{\rm T}(t-1)\hat{\boldsymbol{\varphi}}_{\rm f}(t) + \hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\boldsymbol{P}_\theta(t)\hat{\boldsymbol{\varphi}}_{\rm f}(t)\|\boldsymbol{e}(t)\|^2,\\ T_3(t) &\le T_3(t-1) + \|\hat{\boldsymbol{\xi}}(t)\tilde{\boldsymbol{\rho}}(t-1)\|^2 + 2\tilde{\boldsymbol{\rho}}^{\rm T}(t-1)\hat{\boldsymbol{\xi}}^{\rm T}(t)\boldsymbol{e}_v(t)\\ &\quad + {\rm tr}[\hat{\boldsymbol{\xi}}(t)\boldsymbol{P}_\rho(t)\hat{\boldsymbol{\xi}}^{\rm T}(t)]\|\boldsymbol{e}_v(t)\|^2\\ &= T_3(t-1) + \|\tilde{\boldsymbol{w}}(t)\|^2 + 2\tilde{\boldsymbol{w}}^{\rm T}(t)\boldsymbol{e}_v(t) + \gamma_2(t)\|\boldsymbol{e}_v(t)\|^2\\ &= T_3(t-1) + \|\tilde{\boldsymbol{w}}(t)\|^2 + 2\tilde{\boldsymbol{w}}^{\rm T}(t)[-\tilde{\boldsymbol{w}}(t) + \Delta_2(t) + \boldsymbol{v}(t)]\\ &\quad + \gamma_2(t)\|-\tilde{\boldsymbol{w}}(t) + \Delta_2(t) + \boldsymbol{v}(t)\|^2\\ &= T_3(t-1) - [1-\gamma_2(t)]\|\tilde{\boldsymbol{w}}(t)\|^2\\ &\quad + 2[1-\gamma_2(t)]\tilde{\boldsymbol{w}}^{\rm T}(t)[\Delta_2(t) + \boldsymbol{v}(t)]\\ &\quad + \gamma_2(t)[\|\boldsymbol{v}(t)\|^2 + \|\Delta_2(t)\|^2 + 2\Delta_2^{\rm T}(t)\boldsymbol{v}(t)],\\ T_{12}(t) &\le T_{12}(t-1) + \|\tilde{\boldsymbol{y}}_{\rm f}(t)\|^2 - 2\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\tilde{\boldsymbol{\alpha}}(t-1)\\ &\quad + 2\tilde{\boldsymbol{y}}_{\rm f}^{\rm T}(t)\boldsymbol{e}(t) + \gamma_1(t)\|\boldsymbol{e}(t)\|^2\\ &\le T_{12}(t-1) + \|\tilde{\boldsymbol{y}}_{\rm f}(t)\|^2 - 2\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\tilde{\boldsymbol{\alpha}}(t-1)\\ &\quad + 2\tilde{\boldsymbol{y}}_{\rm f}^{\rm T}(t)[-\tilde{\boldsymbol{y}}_{\rm f}(t) + \Delta_1(t) + \boldsymbol{v}(t)]\\ &\quad + \gamma_1(t)\|-\tilde{\boldsymbol{y}}_{\rm f}(t) + \Delta_1(t) + \boldsymbol{v}(t)\|^2\\ &= T_{12}(t-1) - [1-\gamma_1(t)]\|\tilde{\boldsymbol{y}}_{\rm f}(t)\|^2\\ &\quad - 2\hat{\boldsymbol{\varphi}}_{\rm f}^{\rm T}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\rm f}(t)\tilde{\boldsymbol{\alpha}}(t-1)\\ &\quad + 2[1-\gamma_1(t)]\tilde{\boldsymbol{y}}_{\rm f}^{\rm T}(t)[\Delta_1(t) + \boldsymbol{v}(t)]\\ &\quad + \gamma_1(t)[\|\boldsymbol{v}(t)\|^2 + \|\Delta_1(t)\|^2 + 2\Delta_1^{\rm T}(t)\boldsymbol{v}(t)],\end{aligned}$$

$$
\begin{aligned}
T(t) \le\ & T(t-1) - [1-\gamma_1(t)]\|\tilde{\boldsymbol{y}}_{\mathrm{f}}(t)\|^2 \\
& - 2\hat{\boldsymbol{\varphi}}_{\mathrm{f}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\mathrm{f}}(t)\tilde{\boldsymbol{\alpha}}(t-1) \\
& + 2[1-\gamma_1(t)]\tilde{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t)[\Delta_1(t)+\boldsymbol{v}(t)] \\
& + \gamma_1(t)[\|\boldsymbol{v}(t)\|^2 + \|\Delta_1(t)\|^2 + 2\Delta_1^{\mathrm{T}}(t)\boldsymbol{v}(t)] \\
& - [1-\gamma_2(t)]\|\tilde{\boldsymbol{w}}(t)\|^2 \\
& + 2[1-\gamma_2(t)]\tilde{\boldsymbol{w}}^{\mathrm{T}}(t)[\Delta_2(t)+\boldsymbol{v}(t)] \\
& + \gamma_2(t)[\|\boldsymbol{v}(t)\|^2 + \|\Delta_2(t)\|^2 + 2\Delta_2^{\mathrm{T}}(t)\boldsymbol{v}(t)] \\
\le\ & T(t-1) - [1-\gamma_1(t)]\|\tilde{\boldsymbol{y}}_{\mathrm{f}}(t)\|^2 \\
& - 2\hat{\boldsymbol{\varphi}}_{\mathrm{f}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\mathrm{f}}(t)\tilde{\boldsymbol{\alpha}}(t-1) \\
& + 2[1-\gamma_1(t)]\tilde{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t)[\Delta_1(t)+\boldsymbol{v}(t)] - [1-\gamma_2(t)]\|\tilde{\boldsymbol{w}}(t)\|^2 \\
& + 2[1-\gamma_2(t)]\tilde{\boldsymbol{w}}^{\mathrm{T}}(t)[\Delta_2(t)+\boldsymbol{v}(t)] \\
& + [\gamma_1(t)+\gamma_2(t)][2\|\boldsymbol{v}(t)\|^2 + \|\Delta_1(t)\|^2 \\
& + 2\Delta_1^{\mathrm{T}}(t)\boldsymbol{v}(t) + \|\Delta_2(t)\|^2 + 2\Delta_2^{\mathrm{T}}(t)\boldsymbol{v}(t)] \\
=\ & T(t-1) - [1-\gamma_1(t)]\|\tilde{\boldsymbol{y}}_{\mathrm{f}}(t)\|^2 \\
& - 2\hat{\boldsymbol{\varphi}}_{\mathrm{f}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\mathrm{f}}(t)\tilde{\boldsymbol{\alpha}}(t-1) \\
& + 2[1-\gamma_1(t)]\tilde{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t)\Delta_1(t) - [1-\gamma_2(t)]\|\tilde{\boldsymbol{w}}(t)\|^2 \\
& + 2[1-\gamma_2(t)]\tilde{\boldsymbol{w}}^{\mathrm{T}}(t)\Delta_2(t) \\
& + [\gamma_1(t)+\gamma_2(t)][2\|\boldsymbol{v}(t)\|^2 + \|\Delta_1(t)\|^2 + \|\Delta_2(t)\|^2] \\
& + 2[1-\gamma_1(t)]\tilde{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t)\boldsymbol{v}(t) + 2[1-\gamma_2(t)]\tilde{\boldsymbol{w}}^{\mathrm{T}}(t)\boldsymbol{v}(t) \\
& + [\gamma_1(t)+\gamma_2(t)][2\Delta_1^{\mathrm{T}}(t)\boldsymbol{v}(t) + 2\Delta_2^{\mathrm{T}}(t)\boldsymbol{v}(t)].
\end{aligned}
$$

Here, we have used the relation $ab + cd \le (a+c)(b+d)$ for $a, b, c, d \ge 0$. Let

$$
\begin{aligned}
\gamma_3(t) :=\ & [1-\gamma_1(t)]\|\tilde{\boldsymbol{y}}_{\mathrm{f}}(t)\|^2 + 2\hat{\boldsymbol{\varphi}}_{\mathrm{f}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\mathrm{f}}(t)\tilde{\boldsymbol{\alpha}}(t-1) \\
& - 2[1-\gamma_1(t)]\tilde{\boldsymbol{y}}_{\mathrm{f}}^{\mathrm{T}}(t)\Delta_1(t) + [1-\gamma_2(t)]\|\tilde{\boldsymbol{w}}(t)\|^2 \\
& - 2[1-\gamma_2(t)]\tilde{\boldsymbol{w}}^{\mathrm{T}}(t)\Delta_2(t).
\end{aligned}
$$

Since $\gamma_1(t)$, $\gamma_2(t)$, $\tilde{\boldsymbol{y}}_{\mathrm{f}}(t)$, $\hat{\boldsymbol{\varphi}}_{\mathrm{f}}^{\mathrm{T}}(t)\tilde{\boldsymbol{\theta}}(t-1)\hat{\boldsymbol{\zeta}}_{\mathrm{f}}(t)\tilde{\boldsymbol{\alpha}}(t-1)$, $\Delta_1(t)$, $\Delta_2(t)$ and $\tilde{\boldsymbol{w}}(t)$ are uncorrelated with $\boldsymbol{v}(t)$, taking the mathematical expectation yields

$$
\begin{aligned}
\mathrm{E}[T(t)] \le\ & \mathrm{E}[T(t-1)] - \mathrm{E}[\gamma_3(t)] \\
& + \mathrm{E}\{[\gamma_1(t)+\gamma_2(t)][2\|\boldsymbol{v}(t)\|^2 \\
& + \|\Delta_1(t)\|^2 + \|\Delta_2(t)\|^2]\} + 0.
\end{aligned}
$$

Case I: If $\gamma_3(t) \ge 0$ and $\|\Delta_i(t)\|^2 \le \varepsilon$ for given $\varepsilon < \infty$, we have

$$
\begin{aligned}
\mathrm{E}[T(t)] \le\ & \mathrm{E}[T(t-1)] \\
& + \mathrm{E}\{[\gamma_1(t)+\gamma_2(t)][2\|\boldsymbol{v}(t)\|^2 + \|\Delta_1(t)\|^2 + \|\Delta_2(t)\|^2]\} \\
\le\ & \mathrm{E}[T(t-1)] + \mathrm{E}\{[\gamma_1(t)+\gamma_2(t)](2\sigma^2 + 2\varepsilon)\}.
\end{aligned}
$$

Case II: If $\gamma_3(t) < 0$ or $\|\Delta_i(t)\|^2 > \varepsilon$, let $\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}}(t-1)$, $\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1)$ and $\hat{\boldsymbol{\rho}}(t) = \hat{\boldsymbol{\rho}}(t-1)$, then we have $\mathrm{E}[T(t)] = \mathrm{E}[T(t-1)]$.

Thus, we always have

$$
\begin{aligned}
\mathrm{E}[T(t)] \le\ & \mathrm{E}[T(t-1)] + \mathrm{E}\{[\gamma_1(t)+\gamma_2(t)](2\sigma^2 + 2\varepsilon)\} \\
\le\ & \mathrm{E}[T(0)] + \mathrm{E}\left\{\sum_{j=1}^{t}[\gamma_1(j)+\gamma_2(j)]\right\}(2\sigma^2 + 2\varepsilon).
\end{aligned}
$$

Thus, no matter whether $\gamma_3(t)$ is more than zero or less than zero, the above inequality always holds. The rest of the proof can be done in a similar way in Ding and Gu (2012). □

References

Ding, F., & Chen, T. (2005). Hierarchical gradient-based identification of multivariable discrete-time systems. *Automatica*, *41*(2), 315–325.

Ding, F., & Gu, Y. (2012). Performance analysis of the auxiliary model based least squares identification algorithm for one-step state delay systems. *International Journal of Computer Mathematics*, *89*(15), 2019–2028.

Goodwin, G. C., & Sin, K. S. (1984). *Adaptive filtering prediction and control*. Englewood Cliffs, New Jersey: Prentice Hall.

Li, S., Zhu, Y., & Dickinson, B. W. (1989). A comparison of two linear methods of estimating the parameters of ARMA models. *IEEE Transactions on Automatic Control*, *34*(8), 915–917.

Ljung, L. (1999). *System identification: theory for the user* (2nd ed.). Englewood Cliffs, New Jersey: Prentice Hall.

Mercèrea, G., & Bako, L. (2011). Parameterization and identification of multivariable state-space systems: A canonical approach. *Automatica*, *47*(8), 1547–1555.

Qaisar, S. M., Fesquet, L., & Renaudin, M. (2014). Adaptive rate filtering a computationally efficient signal processing approach. *Signal Processing*, *94*, 620–630.

Wang, Y. J., & Ding, F. (2016a). The auxiliary model based hierarchical gradient algorithms and convergence analysis using the filtering technique. *Signal Processing*, *128*, 212–221.

Wang, Y. J., & Ding, F. (2016b). The filtering based iterative identification for multivariable systems. *IET Control Theory & Applications*, *10*(8), 894–902.

Wang, D. Q., Ding, F., & Zhu, D. Q. (2013). Data filtering based least squares algorithms for multivariable CARAR-like systems. *International Journal of Control, Automation and Systems*, *11*(4), 711–717.

Wang, Y. J., Xu, L., & Ding, F. (2015). Data filtering based parameter estimation algorithms for multivariable Box–Jenkins-like systems. *IFAC-PapersOnLine*, *48*(8), 849–852.

Zhang, H., Shi, Y., & Saadat, M. A. (2011). Robust static output feedback control and remote PID design for networked motor systems. *IEEE Transactions on Industrial Electronics*, *58*(12), 5396–5405.