Article

A Hierarchical Approach for Joint Parameter and State Estimation of a Bilinear System with Autoregressive Noise

Xiao Zhang 1,2,*, Feng Ding 1,2,*, Ling Xu 1, Ahmed Alsaedi 3 and Tasawar Hayat 3

1 Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; xzbang13@126.com (X.Z.); lingxu0848@163.com (L.X.)
2 College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China
3 Department of Mathematics, King Abdulaziz University, Jeddah 21589, Saudi Arabia; aalsaedi@hotmail.com (A.A.); fmgsak@gmail.com (T.H.)

*Correspondence: fding@jiangnan.edu.cn

Received: 7 March 2019; Accepted: 8 April 2019; Published: 17 April 2019

Abstract: This paper is concerned with the joint state and parameter estimation methods for a bilinear system in the state space form, which is disturbed by additive noise. In order to overcome the difficulty that the model contains the product term of the system input and states, we make use of the hierarchical identification principle to present new methods for estimating the system parameters and states interactively. The unknown states are first estimated via a bilinear state estimator on the basis of the Kalman filtering algorithm. Then, a state estimator-based recursive generalized least squares (RGLS) algorithm is formulated according to the least squares principle. To improve the parameter estimation accuracy, we introduce the data filtering technique to derive a data filtering-based two-stage RGLS algorithm. The simulation example indicates the efficiency of the proposed algorithms.

Keywords: bilinear system; hierarchical identification; parameter estimation; least squares; state estimator

1. Introduction

Nonlinear systems widely exist in practical industrial processes. Employing mathematical models of these processes becomes increasingly significant in system analysis [1–3], system control [4–6] and signal processing [7–10], so it is necessary to find efficient methods for system modeling [11–13]. Mathematical modeling methods are summarized into three categories, including mechanism modeling, experimental modeling, and their combination [14–17]. System identification is the field of getting appropriate models of dynamic systems directly from experimental data [18–21] and can be applied to many areas [22–26]. As a special class of nonlinear systems, bilinear systems are capable of approximately describing some complex dynamic processes such as a nuclear reactor, the heat transform process, and biology [27–29]. As a result, there are many active researches towards the parameter estimation problem of bilinear systems [30,31].

In the literature of bilinear system identification, Verdult and Verhaegen applied the subspace identification techniques for multi-input multi-output bilinear systems by using the separable least squares principle [32]. Larkowski et al. handled the parameter estimation problem of the diagonal bilinear errors-in-variables system and employed the bias-compensated least squares method to estimate the parameters [33]. Hizir et al. transformed a bilinear system into its equivalent linear
model and used the observer Kalman filter identification algorithm to identify the bilinear system [34]. Vicario et al. extended the interaction matrices to bilinear systems to present the relationship between the system states and the measurement data and developed the intersection subspace algorithm for the bilinear system identification [35].

The recursive least squares (RLS) method is regarded as the common parameter estimation approach among many different parameter estimation techniques [36–38]. Gan et al. proposed an efficient variable projection method to solve nonlinear least squares problems based on the matrix decomposition [39]. Compared with the transfer function representation [40,41], the state space model can reflect the motion of the inner mechanism and cover state estimation [42–44]. Many estimation methods such as sequential Monte Carlo methods [45] and subspace identification methods [46] have been extensively applied in many areas. Stroud et al. proposed a Bayesian method for the joint state and parameter estimation of state space models with additive Gaussian noise using the ensemble Kalman filter [47]. Urteaga et al. presented a sequential Monte Carlo method for filtering and prediction of time-varying signals by fusing the information from candidate models instead of using model selection [48]. Martino et al. presented the parallel particle filters based on the Bayesian model averaging principle for sequential tracking and online model selection [49]. Schön et al. derived an expectation maximization algorithm under the maximum likelihood framework and gave the parameter estimates of nonlinear state space models [50]. Li and Liu derived the input-output representation of the bilinear system through eliminating the state variables and proposed the filtering-based least squares iterative algorithm [51]. However, the state variables in the system were eliminated, so the state estimation of the bilinear system was not studied.

The bilinear system involves the products of control inputs and state variables in the state equation. Thus, linear system identification methods cannot be utilized to compute the parameter estimates due to the unmeasurable states. Different from the work in [51], this paper focuses on the joint parameter and state estimation problem for a bilinear state space system with colored noise. The bilinear model can be regarded as a time-varying state space model [52]. Thus, this paper presents a bilinear state estimator based on the Kalman filtering algorithm for computing the unknown states. Then, the state estimator-based recursive generalized least squares (RGLS) identification algorithm is developed for estimating the system parameters according to the least squares principle. For the purpose of improving the parameter estimation accuracy, the data filtering technique is introduced by transforming the bilinear model into several filtered submodels with smaller dimensions. The main contributions of this paper are listed as follows.

- We present a bilinear state estimator on the basis of the Kalman filtering algorithm.
- We derive a state estimator-based RGLS algorithm for joint state and parameter estimation of bilinear state space models on the basis of the hierarchical identification principle.
- We derive a filtering-based two-stage RGLS (F-TS-RGLS) algorithm using the state estimates for improving the parameter estimation accuracy by introducing the data filtering technique.

The remainder of this paper is organized as follows. Section 2 describes the bilinear system and gives its identification model. Section 3 derives a state estimator-based RGLS identification algorithm. Section 4 presents a state estimator-based F-TS-RGLS algorithm by using the data filtering technique. Section 5 provides an illustrative example to demonstrate the effectiveness of the proposed algorithms. Finally, some conclusions are given in Section 6.

2. System Description

Some symbols are introduced for convenience. “$A =: X$” or “$X: = A$” stands for “$A$ is defined as $X$”; $\hat{\theta}_t$ denotes the estimate of $\theta$ at time $t$; the symbol $I (I_n)$ represents an identity matrix of an appropriate size ($n \times n$); $z$ denotes a unit forward shift operator like $zx_t = x_{t+1}$ and $z^{-1}x_t = x_{t-1}$; the superscript T symbolizes the matrix/vector transpose.
Recently, a state observer-based multi-innovation stochastic gradient algorithm and a state observer-based recursive least squares identification algorithm were presented for a bilinear system with white noise [53]:

\[ x_{t+1} = Ax_t + Bx_t u_t + f u_t, \]
\[ y_t = h x_t + v_t. \]

Zhang et al. derived a state filtering-based hierarchical identification algorithm and a state filtering-based forgetting factor recursive least squares algorithm for a bilinear system with white noise [54]:

\[ x_{t+1} = Ax_t + Bx_t u_t + f u_t, \]
\[ y_t = h x_t + v_t. \]

On the basis of the data filtering technique, a bilinear state observer-based multi-innovation extended stochastic gradient algorithm and a data filtering-based multi-innovation extended stochastic gradient algorithm were developed for a bilinear system with moving average noise [55]:

\[ x_{t+1} = Ax_t + Bx_t u_t + f u_t, \]
\[ y_t = h x_t + D(z) v_t. \]

where \( D(z) = 1 + d_1 z^{-1} + \cdots + d_n z^{-n_d}. \)

Different from the work in [53–55], this paper considers a single-input single-output bilinear system with autoregressive noise:

\[ x_{t+1} = Ax_t + Bx_t u_t + f u_t, \]
\[ y_t = h x_t + w_t. \]

where \( x_t = [x_{1,t}, x_{2,t}, \cdots, x_{n,t}]^T \in \mathbb{R}^n \) is the state vector, \( u_t \in \mathbb{R} \) and \( y_t \in \mathbb{R} \) are the system input and output data, \( w_t \in \mathbb{R} \) is the measurement noise, and \( A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times n}, f \in \mathbb{R}^n \), and \( h \in \mathbb{R}^{1 \times n} \) are the system parameters:

\[
A: = \begin{bmatrix}
-a_1 & 1 & 0 & \cdots & 0 \\
-a_2 & 1 & \ddots & \ddots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
-a_{n-1} & 0 & \cdots & 0 & 1 \\
-a_n & 0 & \cdots & 0 & 0
\end{bmatrix} \in \mathbb{R}^{n \times n},
\]

\[
B: = \begin{bmatrix}
b_1 \\ b_2 \\ \vdots \\ b_n
\end{bmatrix} \in \mathbb{R}^{n \times n}, \quad b_i \in \mathbb{R}^{1 \times n}, \quad f: = \begin{bmatrix}
f_1 \\ f_2 \\ \vdots \\ f_n
\end{bmatrix} \in \mathbb{R}^n, \quad h: = [1, 0, \cdots, 0] \in \mathbb{R}^{1 \times n}.
\]

Without loss of generality, assume that \( u_t = 0, x_t = 0, y_t = 0 \), and \( w_t = 0 \) for \( t \leq 0 \).

**Assumption 1.** This paper assumes that the stochastic noise \( v_t \in \mathbb{R} \) is a Gaussian noise with zero mean and variance \( \sigma^2 \). The colored noise \( w_t \) can be a moving average process, an autoregressive process, or an autoregressive moving average process. In this paper, \( w_t \) is approximated by an autoregressive process:

\[ w_t: = -c_1 w_{t-1} - \cdots - c_{n_c} w_{t-n_c} + v_t, \]
Assumption 2. The system stability is the basis of system identification. This paper assumes that the bilinear system in (1) and (2) is stable, that is to say the bounded input leads to the bounded output, and the system is observable and controllable.

Assumption 3. System identification contains the determination of the system dimension and parameter estimation. This paper assumes that the orders $n$ and $n_i$ of the system are known. The parameters $a_i$, $b_{ij}$, $f_i$, and $c_i$ are to be identified from experimental data $u_i$ and $y_i$.

Referring to the method in [53] and from (1) and (2), we have:

$$x_{1,t} = - \sum_{i=1}^{n} a_i x_{1,t-i} + \sum_{i=1}^{n} b_i x_{t-i} u_{t-i} + \sum_{i=1}^{n} f_i u_{t-i}.$$  \hspace{1cm} (4)

Define the system parameter vectors $\theta$, $\theta_n$, and $\theta_n$ and the information vectors $\varphi_t$, $\varphi_{s,t}$, and $\varphi_{n,t}$ as:

$$\theta: \begin{bmatrix} \theta_h \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n^2 + 2n + n_c},$$

$$\theta_n: \begin{bmatrix} \alpha^T \\ \beta^T \\ \gamma^T \end{bmatrix} \in \mathbb{R}^{n^2 + 2n},$$

$$\alpha: [a_1, a_2, \cdots, a_{n-1}, a_n]^T \in \mathbb{R}^n,$$

$$\beta: [b_1, b_2, \cdots, b_{n-1}, b_n]^T \in \mathbb{R}^n,$$

$$\gamma: [f_1, f_2, \cdots, f_{n-1}, f_n]^T \in \mathbb{R}^n,$$

$$\theta_n: [c_1, c_2, \cdots, c_{n-1}, c_n]^T \in \mathbb{R}^n,$$

$$\varphi_t: \begin{bmatrix} \varphi_{s,t} \\ \varphi_{n,t} \end{bmatrix} \in \mathbb{R}^{n^2 + 2n + n_c},$$

$$\varphi_{s,t}: [\varphi_{s,t}^1, \varphi_{s,t}^2, \varphi_{s,t}^3, \varphi_{s,t}^4]^T \in \mathbb{R}^{n^2 + 2n},$$

$$\varphi_{s,t}: [-x_{1,t-1}, -x_{1,t-2}, \cdots, -x_{1,t-n}]^T \in \mathbb{R}^n,$$

$$\varphi_{s,t}: [x_{1,t-1} u_{t-1}, x_{1,t-2} u_{t-2}, \cdots, x_{1,t-n} u_{t-n}]^T \in \mathbb{R}^n,$$

$$\varphi_{s,t}: [u_{t-1}, u_{t-2}, \cdots, u_{t-n+1}, u_{t-n}]^T \in \mathbb{R}^n,$$

$$\varphi_{s,t}: [-w_{t-1}, -w_{t-2}, \cdots, -w_{t-n}]^T \in \mathbb{R}^n.$$  \hspace{1cm} (4)

Inserting (4) into (2) gives:

$$y_t = x_{1,t} + w_t$$

$$= \varphi_{t,1}^T a + \varphi_{s,t}^T b + \varphi_{g,t}^T f + w_t$$

$$= \varphi_{s,t}^T \theta_n + w_t.$$  \hspace{1cm} (5)

Then, Equation (3) can be written as:

$$w_t = -c_1 w_{t-1} - c_2 w_{t-2} - \cdots - c_n w_{t-n} + v_t$$

$$= \varphi_{n,t}^T \theta_n + v_t.$$  \hspace{1cm} (6)
Substituting (6) into (5), we obtain:

\[
\begin{align*}
  y_t &= x_{1,t} + w_t \\
  &= \phi_{x,t} x_{t} + \omega_t \\
  &= \phi'_{x,t} x_{t} + \phi'_{n,t} \theta_n + v_t \\
  &= \hat{\phi}' x_{t} + v_t. 
\end{align*}
\] (7)

Equation (7) is the identification model of the bilinear system in (1) and (2). The objective of this paper is to derive a bilinear state estimator to obtain the state estimates and explore an efficient parameter identification method to generate highly-accurate parameter estimates and to improve the computational efficiency by using available experimental data.

3. The RGLS Algorithm Using the Bilinear State Estimates

In this section, a state estimator-based RGLS algorithm is employed to estimate the unknown states and parameters of the considered bilinear state space system.

Define a cost function:

\[
J(\theta) = \sum_{t=1}^{T} [y_t - \phi'_t \theta]^2. 
\]

Using the least squares principle and minimizing \(J(\theta)\) give the recursive least squares algorithm:

\[
\begin{align*}
  \hat{\theta}_t &= \hat{\theta}_{t-1} + L_t [y_t - \phi'_t \hat{\theta}_{t-1}], \\
  L_t &= P_{t-1} \phi'_t [1 + \phi'_t P_{t-1} \phi'_t]^{-1}, \\
  P_t &= [I - L_t \phi'_t] P_{t-1}. 
\end{align*}
\] (8–10)

The first identification difficulty occurs because only the observation data \(u_t\) and \(y_t\) are available, but \(\phi'_t, \phi'_{x,t}\) and \(\phi'_{xu,t}\) contain the unknown \(x_t\).

**Remark 1.** To solve this problem, one method is to eliminate the state vector \(x_t\) and to obtain a new identification model that involves only the system input and output variables [51]. However, this cannot work because the parameter matrix \(B\) here is not the special matrix with many zero entries like in [51]. Thus, it is necessary to study new state estimation methods to obtain the state estimate.

The second identification difficulty occurs because \(\phi'_t\) also contains the noise variable \(\{w_{t-i}, i = 1, 2, \ldots, n_c\}\), so the parameter estimates \(\hat{\theta}_t\) cannot be computed by (8)–(10).

**Remark 2.** The solution here is to introduce the hierarchical identification method for obtaining the parameter estimates and state estimates, respectively. Based on the auxiliary model identification idea, we replace the unknown disturbance \(w_{t-i}\) and the unknown state \(\{x_{t-i}, i = 1, 2, \ldots, n\}\) with their estimates \(\hat{w}_{t-i}\) and \(\hat{x}_{t-i}\) in the identification algorithms and define the estimated information vectors \(\hat{\phi}'_t, \hat{\phi}'_{x,t}, \hat{\phi}'_{xu,t}, \hat{\phi}'_{nu,t}\) of \(\phi'_t, \phi'_{x,t}, \phi'_{xu,t}, \phi'_{nu,t}\) as:

\[
\begin{align*}
  \hat{\phi}'_t &= \begin{bmatrix} \hat{\phi}'_{n,t} \\ \hat{\phi}'_{n,t} \end{bmatrix} \in \mathbb{R}^{n + 2n + n_c}, \\
  \hat{\phi}'_{n,t} &= \begin{bmatrix} \hat{\phi}'_{x,t}^T, \hat{\phi}'_{xu,t}^T, \hat{\phi}'_{nu,t}^T \end{bmatrix}^T \in \mathbb{R}^{n + 2n}, \\
  \hat{\phi}'_{x,t} &= [-\hat{x}_{t-1}, -\hat{x}_{t-2}, \ldots, -\hat{x}_{t-n}]^T \in \mathbb{R}^n, \\
  \hat{\phi}'_{xu,t} &= [-\hat{x}_{t-1}, -\hat{x}_{t-2}, \ldots, -\hat{x}_{t-n}]^T \in \mathbb{R}^{2n}, \\
  \hat{\phi}'_{nu,t} &= [-\hat{w}_{t-1}, -\hat{w}_{t-2}, \ldots, -\hat{w}_{t-n}] \in \mathbb{R}^{n_c}, 
\end{align*}
\] (11–15)
and the estimated parameter vectors \( \hat{\theta}_t \), \( \hat{\theta}_{n,t} \), \( \hat{\alpha}_t \), \( \hat{\beta}_t \), \( \hat{f}_t \), and \( \hat{\theta}_{n,t} \) of \( \theta \), \( \theta_n \), \( a \), \( b \), \( f \), and \( \theta_n \) as:

\[
\hat{\theta}_t = \begin{bmatrix} \hat{\theta}_{n,t} \\ \hat{\theta}_t \end{bmatrix} \in \mathbb{R}^{n+2n+u_t},
\]

\[
\hat{\theta}_{n,t} = [\hat{\alpha}_t; \hat{\beta}_t; \hat{f}_t]^T \in \mathbb{R}^{n^2},
\]

\[
\hat{\alpha}_t = [\hat{\alpha}_1; \hat{\alpha}_2; \cdots; \hat{\alpha}_{n,t}]^T \in \mathbb{R}^n,
\]

\[
\hat{\beta}_t = [\hat{\beta}_1; \hat{\beta}_2; \cdots; \hat{\beta}_{n,t}]^T \in \mathbb{R}^n,
\]

\[
\hat{f}_t = [\hat{f}_1; \hat{f}_2; \cdots; \hat{f}_{n,t}]^T \in \mathbb{R}^n,
\]

\[
\hat{\theta}_{n,t} = [\hat{\epsilon}_1; \hat{\epsilon}_2; \cdots; \hat{\epsilon}_{n,t}]^T \in \mathbb{R}^{n_t}.
\]

According to the structure of (5), the estimate \( \hat{\omega}_t \) can be calculated through:

\[
\hat{\omega}_t = y_t - \hat{\phi}_x^T \hat{\alpha}_t - \hat{\phi}_u^T \hat{\beta}_t - \hat{\phi}_f^T \hat{f}_t
\]

\[= y_t - \hat{\phi}_x^T \hat{\omega}_{n,t} \quad (16)
\]

According to the method in [55], we construct the following bilinear state estimator:

\[
\hat{x}_{t+1} = \hat{A}_t \hat{x}_t + \hat{B}_t \hat{u}_t + \hat{f}_t \hat{u}_t + L_{x,t} [y_t - \hat{h} \hat{x}_t - \hat{\phi}_n \hat{\theta}_n],
\]

\[= [A + B \hat{u}_t | P_{x,t} h^T | h P_{x,t} h^T + 1]^{-1},
\]

\[= [A + B \hat{u}_t] [A + B \hat{u}_t] - L_{x,t} h P_{x,t} [A + B \hat{u}_t]^T.
\]

**Remark 3.** Another problem occurs because the parameters \( A, B, f \) are unknown. Therefore, the state estimation in (17)–(19) cannot be realized. The solution is to use \( \hat{a}_t \), \( \hat{b}_t \), and \( \hat{f}_t \) to construct the parameter estimates \( \hat{A}_t \), \( \hat{B}_t \), and \( \hat{f}_t \) of \( A, B, \) and \( f \) as:

\[
\hat{A}_t = \begin{bmatrix}
-\hat{a}_{1,t} & 1 & 0 & \cdots & 0 \\
-\hat{a}_{2,t} & 0 & 1 & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
-\hat{a}_{n-1,t} & 0 & \cdots & 0 & 1 \\
-\hat{a}_{n,t} & 0 & \cdots & 0 & 0
\end{bmatrix}, \quad \hat{B}_t = \begin{bmatrix}
\hat{b}_{1,t} \\
\hat{b}_{2,t} \\
\vdots \\
\hat{b}_{n,t} \\
\hat{b}_{n,t}
\end{bmatrix}, \quad \hat{f}_t = \begin{bmatrix}
\hat{f}_{1,t} \\
\hat{f}_{2,t} \\
\vdots \\
\hat{f}_{n,t}
\end{bmatrix} \quad (20)
\]

Then, replacing \( A, B, f, \) and \( \theta_n \) in (17)–(19) with their estimates \( \hat{A}_t, \hat{B}_t, \hat{f}_t, \) and \( \hat{\theta}_{n,t} \) obtains the bilinear state estimator:

\[
\hat{x}_{t+1} = \hat{A}_t \hat{x}_t + \hat{B}_t \hat{u}_t + \hat{f}_t \hat{u}_t + L_{x,t} [y_t - \hat{h} \hat{x}_t - \hat{\phi}_n \hat{\theta}_n],
\]

\[= [\hat{A}_t + \hat{B}_t | \hat{B}_t | h^T h + 1]^{-1},
\]

\[= [\hat{A}_t + \hat{B}_t] [\hat{A}_t + \hat{B}_u]^T - L_{x,t} h P_{x,t} [\hat{A}_t + \hat{B}_u]^T.
\]

Replacing the unknown \( \phi_i \) in (8)–(10) with its estimate \( \hat{\phi}_i \), and combining the bilinear state estimator in (20)–(23) give:

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + L_t [y_t - \hat{\phi}_i \hat{\theta}_{t-1}],
\]

\[= P_{t-1} \hat{\phi}_i [1 + \hat{\phi}_i P_{t-1} \hat{\phi}_i]^{-1},
\]

\[= [I - L_t \hat{\phi}_i] P_{t-1}^{-1}.
\]

Equations (11)–(26) form the recursive generalized least squares (RGLS) algorithm based on the state estimates.
Remark 4. In order to improve the RGLS parameter estimation accuracy, the next section will construct a linear filter to filter the experimental data for estimating the system parameters.

4. The F-TS-RGLS Algorithm Using the Bilinear State Estimator

For the purpose of improving the parameter estimation accuracy of the RGLS algorithm, this section presents an F-TS-RGLS algorithm for the bilinear system by constructing a linear filter \( C(z) \) to filter the observation data of the system and to decompose the identification model in (7) into two sub-identification models. Multiplying both sides of (1) and (2) by \( C \) to filter the observation data of the system and to decompose the identification model in (7) into two sub-identification models. Minimizing \( J \) gives the following recursive relations:

\[
\begin{align*}
C(z)x_{t+1} &= C(z)A x_t + C(z)B x_t u_t + C(z)f u_t, \quad (27) \\
C(z)y_t &= C(z)h x_t + v_t. \quad (28)
\end{align*}
\]

Define the filtered input \( u_{t,f} \), the filtered output \( y_{t,f} \), and the filtered state \( x_{t,f} \) as:

\[
\begin{align*}
u_{t,f} &= C(z)u_t = u_t + \sum_{i=1}^{n_f} c_i u_{t-i}, \\
y_{t,f} &= C(z)y_t = y_t + \sum_{i=1}^{n_f} c_i y_{t-i}, \\
x_{t,f} &= C(z)x_t = x_t + \sum_{i=1}^{n_f} c_i x_{t-i}.
\end{align*}
\]

Equations (27) and (28) can be transformed as:

\[
\begin{align*}
x_{t+1,f} &= A x_{t,f} + B x_{t,f} u_t + f u_t, \quad (32) \\
y_{t,f} &= h x_{t,f} + v_t, \quad (33)
\end{align*}
\]

where \( x_{t,f} = [x_{1,t,f}, x_{2,t,f}, \ldots, x_{(n-1),t,f}, x_{n,t,f}]^T \in \mathbb{R}^n \). From (32) and (33), we have:

\[
x_{t,f} = - \sum_{i=1}^{n_f} a_i x_{t-i,f} + \sum_{i=1}^{n_f} b_i x_{t-i,f} u_t + \sum_{i=1}^{n_f} f_i u_{t-i} = q_{t,f}^x \theta_s,
\]

where the filtered information vector:

\[
q_{t,f}^x = \begin{bmatrix} -x_{t,f-1} & -x_{t,f-2} & \cdots & -x_{t,f-n+1} & -x_{t,f-n} \\ x_{t,f-1} & x_{t,f-2} & \cdots & x_{t,f-n} & u_{t-1} & \cdots & u_{t-n} \\ u_{t-2} & \cdots & u_{t-n+1} & u_{t-n} \end{bmatrix}^T \in \mathbb{R}^{n^2 + 2n}.
\]

Equations (33) and (6) can be expressed as the following two filtered sub-identification models:

\[
\begin{align*}
y_{t,f} &= q_{t,f}^y \theta_s + v_t, \quad (36) \\
w_{t} &= q_{t,f}^w \theta_n + v_t. \quad (37)
\end{align*}
\]

On the basis of the two sub-identification models in (36) and (37), define two cost functions:

\[
\begin{align*}
J_1(\theta_s) &= \sum_{j=1}^{l_1} [y_{t,f} - q_{t,f}^y \theta_s]^2, \\
J_2(\theta_n) &= \sum_{j=1}^{l_2} [w_{t} - q_{t,f}^w \theta_n]^2.
\end{align*}
\]

Minimizing \( J_1(\theta_s) \) and \( J_2(\theta_n) \) gives the following recursive relations:
\[ \hat{\theta}_{n,t} = \hat{\theta}_{n,t-1} + L_n \{ y_{t,i} - \varphi_{n}^{i,t} \hat{\theta}_{n,t-1} \}, \quad (39) \]
\[ L_n = P_{n,t-1} \varphi_{n}^{i,t} \{ 1 + \varphi_{n}^{i,t} P_{n,t-1} \varphi_{n}^{i,t} \}^{-1}, \quad (40) \]
\[ P_n = [I - L_n \varphi_{n}^{i,t} P_{n,t-1}], \quad (41) \]
\[ \hat{\theta}_{n,t} = \hat{\theta}_{n,t-1} + L_n \{ w_t - \varphi_{n}^{i,t} \hat{\theta}_{n,t-1} \}, \quad (42) \]
\[ L_n = P_{n,t-1} \varphi_{n}^{i,t} \{ 1 + \varphi_{n}^{i,t} P_{n,t-1} \varphi_{n}^{i,t} \}^{-1}, \quad (43) \]
\[ P_n = [I - L_n \varphi_{n}^{i,t} P_{n,t-1}], \quad (44) \]

However, Equations (39)–(44) cannot generate the parameter estimates \( \hat{\theta}_{n,t} \) and \( \hat{\theta}_{n,t} \). Because the noise term \( w_t \) and \( \varphi_{n,t} \) are unmeasurable, the system state \( x_t \) in the system information vector \( \varphi_{n,t} \) is unmeasurable. In addition, \( C(z) \) is unknown, then \( y_{t,i}, u_{t,i}, x_{t,i} \), and \( \varphi_{t,i} \) cannot be obtained.

**Remark 5.** The difficulty is overcome by utilizing the auxiliary model identification idea to take the place of the unknown variables with their estimates. Use the estimate \( \hat{\varphi}_{n,t} \) of \( \varphi_{n,t} \):

\[ \hat{\varphi}_{n,t} = \left[ -\hat{\varphi}_{t-1} - \hat{\varphi}_{t-2} \cdots - \hat{\varphi}_{t-n} \right]^T. \quad (45) \]

From (5), replacing \( \varphi_{n,t} \) and \( \theta_{n} \) with their corresponding estimates \( \hat{\varphi}_{n,t} \) and \( \hat{\theta}_{n,t} \) obtains the estimate of \( w_t \):

\[ \hat{\varphi}_{n} = y_t - \hat{\varphi}_{n,t} \hat{\theta}_{n,t-1}, \quad (46) \]

where:

\[ \hat{\varphi}_{n,t} = \begin{bmatrix} \hat{\varphi}_{t,i}^1, \hat{\varphi}_{n,i}, \varphi_{n}^i \end{bmatrix}^T, \quad (47) \]
\[ \hat{\varphi}_{t,i} = \begin{bmatrix} \hat{\varphi}_{t,i}^1, \hat{\varphi}_{t,i}^2, \cdots, \hat{\varphi}_{t,i}^n \end{bmatrix}^T, \quad (48) \]
\[ \hat{\varphi}_{n,i} = \begin{bmatrix} \hat{\varphi}_{n,i}^1, \hat{\varphi}_{n,i}^2, \cdots, \hat{\varphi}_{n,i}^n \end{bmatrix}^T, \quad (49) \]
\[ \varphi_{n,t} = \begin{bmatrix} u_{t-1}, u_{t-2}, \cdots, u_{t-n} \end{bmatrix}^T, \quad (50) \]
\[ \hat{\theta}_{n,t} = \begin{bmatrix} \hat{\theta}_{n,t}^1, \hat{\theta}_{n,t}^2, \cdots, \hat{\theta}_{n,t}^n \end{bmatrix}^T. \quad (51) \]

By virtue of the Kalman filter, a state estimator in the bilinear form can be established for generating the state estimates. Then, utilizing the parameter estimates:

\[ \hat{\theta}_{n,t} = [\hat{\varphi}_{1,i}^1, \hat{\varphi}_{2,i}^1, \cdots, \hat{\varphi}_{n,i}^1] \]

(52)

to construct the estimates of \( C(z) \) gives:

\[ \hat{C}(t,z) = 1 + \hat{\varphi}_{1,i} z^{-1} + \hat{\varphi}_{2,i} z^{-2} + \cdots + \hat{\varphi}_{n,i} z^{-n}. \]

Since the linear filter is unknown, we use its estimate to filter the system input \( u_t \), the system output \( y_t \), and the state system \( x_t \) to give the filtered estimates of \( u_{t,i}, y_{t,i}, \) and \( x_{t,i} \):

\[ \hat{u}_{t,i} = \hat{C}(t,z) u_t = u_t + \sum_{i=1}^{n_c} \hat{\varphi}_{i,i} u_{t-i}, \quad (53) \]
\[ \hat{y}_{t,i} = \hat{C}(t,z) y_t = y_t + \sum_{i=1}^{n_c} \hat{\varphi}_{i,i} y_{t-i}, \quad (54) \]
\[ \hat{x}_{t,i} = \hat{x}_t + \hat{C}(t,z) x_t = \hat{x}_t + \sum_{i=1}^{n_c} \hat{\varphi}_{i,i} \hat{x}_{t-i}. \quad (55) \]

Referring to the derivation of (21)–(23), we can obtain the following state estimator for obtaining the estimate of the filtered state \( x_{t,i} \):
\[ \dot{x}_{t+1} = \dot{A}_t \dot{x}_{t} + \dot{B}_t \dot{u}_t + \dot{f}_t + L_{t} [y_{t} - h \dot{x}_t - \dot{\phi}_{n,t}] , \]  
\[ L_{t} = [\dot{A}_t + \dot{B}_t u_t] P_{t} h^T [h P_{t} h^T + 1]^{-1} , \]  
\[ P_{t+1} = [\dot{A}_t + \dot{B}_t u_t] P_{t} [\dot{A}_t + \dot{B}_t u_t]^T - L_{t} h P_{t} [\dot{A}_t + \dot{B}_t u_t]^T , \]  
\[ \dot{A}_t = \begin{bmatrix} -\dot{a}_{1,t} & 0 & \cdots & 0 \\ -\dot{a}_{2,t} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -\dot{a}_{n,t} & 0 & \cdots & 0 \end{bmatrix}, \]  
\[ \dot{B}_t = \begin{bmatrix} \dot{b}_{1,t} \\ \dot{b}_{2,t} \\ \vdots \\ \dot{b}_{n,t} \end{bmatrix} , \]  
\[ \dot{f}_t = \begin{bmatrix} \dot{f}_{1,t} \\ \dot{f}_{2,t} \\ \vdots \\ \dot{f}_{n,t} \end{bmatrix} . \]  

By replacing \( x_{t-1}, x_{t-1} \) and \( u_{t-1} \) in (35) with their estimates \( \hat{x}_{1,t-1}, \hat{x}_{1,t-1} \) and \( \hat{u}_{t-1} \), define the estimated filtered information vector:

\[ \hat{\phi}_{1,t} = \begin{bmatrix} -\hat{x}_{1,t-1,1} - \hat{x}_{1,t-1,2}, \cdots, -\hat{x}_{1,t-1,n}, -\hat{x}_{1,t-1,1} \\ \hat{x}_{t-1} - \hat{x}_{t-1,1}, \hat{x}_{t-1,2}, -\hat{x}_{t-1,1}, \cdots, -\hat{x}_{t-1,1} - \hat{x}_{t-1,1}, \cdots, -\hat{x}_{t-1,1} - \hat{x}_{t-1,1} \end{bmatrix} \in \mathbb{R}^{2(n+1)} \]  

Similarly, replacing the unmeasurable information vector \( \phi_{n,t} \) and \( \hat{\phi}_{n,t} \) in (39)–(44) with \( \hat{\phi}_{1,t} \) and \( \hat{\phi}_{n,t} \) obtains:

\[ \hat{\theta}_{t} = \hat{\theta}_{t-1} + L_{n,t} [\hat{y}_{t} - \hat{\phi}_{1,t}^T \hat{\theta}_{t-1}] , \]  
\[ L_{t} = P_{s,t} [1 + \hat{\phi}_{1,t}^T P_{s,t-1} \hat{\phi}_{1,t}]^{-1} , \]  
\[ P_{s,t} = [I - L_{s,t} \hat{\phi}_{1,t}^T] P_{s,t-1} , \]  
\[ \hat{n}_{t} = \hat{n}_{t-1} + L_{n,t} [\hat{w}_{t} - \hat{\phi}_{n,t}^T \hat{\theta}_{t-1}] , \]  
\[ L_{n,t} = P_{n,t} [1 + \hat{\phi}_{n,t}^T P_{n,t-1} \hat{\phi}_{n,t}]^{-1} , \]  
\[ P_{n,t} = [I - L_{n,t} \hat{\phi}_{n,t}^T] P_{n,t-1} . \]  

The procedures of calculating \( \hat{\theta}_{s,t} \) and \( \hat{n}_{t} \) by the filtering-based recursive generalized least squares algorithm in (45)–(66) are listed as follows.

1. To initialize: let \( t = 1 \), \( \hat{\theta}_{s,0} = \frac{1}{p_0} \), \( \hat{n}_{s,0} = \frac{1}{p_0^2 + 2n} / p_0 \), \( \hat{y}_{t} = \frac{1}{p_0} \), \( \hat{u}_{t} = \frac{1}{p_0} \), \( \hat{x}_{t+1} = \frac{1}{p_0} \), \( \hat{x}_{t-1} = \frac{1}{p_0} \), \( \hat{w}_{t} = \frac{1}{p_0} \), \( \hat{f}_t = \frac{1}{p_0} \), \( \hat{P}_t = \frac{1}{p_0} \), \( \hat{L}_t = \frac{1}{p_0} \), \( p_0 = 10^6 \), \( i = 1, 2, \ldots, n \), \( j = 1, 2, \ldots, n \).
2. Collect the measurement data \( u_t \) and \( y_t \). Form \( \hat{\phi}_{n,t} \) using (45) and \( \hat{\phi}_{s,t} \) using (47)–(49).
3. Compute \( \hat{w}_t \) using (46), the gain vector \( L_{n,t} \) using (65), and the covariance matrix \( P_{s,t} \) using (66).
4. Update the parameter estimates \( \hat{\theta}_{s,t} \) using (64). Read \( \hat{c}_{ij} \) from \( \hat{n}_{s,t} \) in (52).
5. Compute \( \hat{y}_{t} \) and \( \hat{u}_{t} \) from (53) and (54). Construct \( \hat{\phi}_{1,t} \) using (60).
6. Compute the gain vector \( L_{s,t} \) using (62) and the covariance matrix \( P_{s,t} \) using (63). Update the parameter estimates \( \hat{\theta}_{s,t} \) using (61).
7. Read \( \hat{u}_{t}, \hat{b}_{j,t} \) and \( \hat{f}_{j,t} \) from \( \hat{\theta}_{s,t} \) in (51). Construct \( \hat{A}_t, \hat{B}_t, \hat{f}_t \) using (59). Compute \( L_{t} \) and \( P_{t+1} \) using (57) and (58).
8. Compute \( \hat{x}_{t+1} \) and \( \hat{x}_t \) using (56) and (55).
9. Increase \( t \) by one, and turn to Step 2.

Remark 6. By using the hierarchical identification principle, the unknown parameters and states can be estimated interactively.

Remark 7. By introducing a linear filter, the original identification model is decomposed into two filtered sub-identification models. Then, the system parameters and the noise parameters are estimated interactively.
Remark 8. For identification methods, the large dimensions of the parameter vector $\theta$ cause a heavy computational cost. Thus, this paper uses the hierarchical identification principle for improving the computational efficiency of the proposed methods based on the decomposition.

Remark 9. The number of the multiplication and addition operations is used to measure the computational cost of an algorithm. The computational cost of the RGLS and F-TS-RGLS algorithms is shown in Table 1.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Number of Multiplications</th>
<th>Number of Additions</th>
<th>Total Flop</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGLS</td>
<td>$2(n^2 + 2n + n_c)^2 +$</td>
<td>$2(n^2 + 2n + n_c)^2 +$</td>
<td>$N_1: = 4(n^2 + 2n + n_c)^2 +$</td>
</tr>
<tr>
<td></td>
<td>$4n^2 + 2n + n_c + n^2 + 2n$</td>
<td>$2(n^2 + 2n + n_c) + n^2 + 2n$</td>
<td>$6(n^2 + 2n + n_c) + 2n^2 + 4n$</td>
</tr>
<tr>
<td>F-TS-RGLS</td>
<td>$(n^2 + 2n)^2 + 5n^2 +$</td>
<td>$2(n^2 + 2n)^2 + 3n^2 +$</td>
<td>$N_2: = 4(n^2 + 2n)^2 + 8n^2 +$</td>
</tr>
<tr>
<td></td>
<td>$10n + 2n^2 + 6n_c$</td>
<td>$6n + 2n^2 + 4n_c$</td>
<td>$16n + 4n^2 + 10n_c$</td>
</tr>
</tbody>
</table>

The difference of the computational cost of the RGLS and the F-TS-RGLS algorithms is:

$$N_1 - N_2 = (4n^2 + 2n + n_c)^2 + 6(n^2 + 2n + n_c) + 2n^2 + 4n - [4(n^2 + 2n)^2 + 8n^2 + 16n + 4n^2 + 10n_c]$$

$$= (8n^2 + 2n) - 4n_c.$$

Remark 10. Since $n \geq 1$ and $n_c \geq 1$, the difference $N_1 - N_2$ is positive. That is to say, the computational cost of the F-TS-RGLS algorithm is less than that of the RGLS algorithm.

5. Numerical Example

This section provides an example to illustrate the effectiveness of the proposed algorithms. Here, we consider a bilinear state space model:

$$x_{t+1} = \begin{bmatrix} 0.46 & 1 \\ -0.30 & 0 \end{bmatrix} x_t + \begin{bmatrix} 0.08 & 0.17 \\ -0.07 & -0.22 \end{bmatrix} x_t u_t + \begin{bmatrix} 0.40 \\ 0.15 \end{bmatrix} u_t,$$

$$y_t = [1, 0] x_t + w_t,$$

$$w_t = -c w_{t-1} + v_t.$$

The parameter vector to be identified is:

$$\theta = [a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, f_1, f_2, c]^T = [-0.46, 0.30, 0.08, 0.17, -0.07, -0.22, 0.40, 0.15, 0.08]^T.$$

In the simulation, the input signal $[u_t]$ was chosen as a pseudo-random binary sequence with zero mean and unit variance and $[v_t]$ as a white noise sequence with zero mean and variances $\sigma^2$. Set the data length $L = 3000$, and employ the proposed algorithms to estimate the parameters and states of this example system. The simulation input and output data are shown in Figure 1, which indicate that the bounded input signal leads to the bounded output. The parameter estimates and their estimation errors of the RGLS algorithm and the F-TS-RGLS algorithm are shown in Tables 2 and 3. The RGLS estimation error $\delta$ versus $t$ is shown in Figure 2 with $\sigma^2 = 0.10^2$ and $\sigma^2 = 0.15^2$. The F-TS-RGLS estimation error versus $t$ is shown in Figure 3 with different variances. The predicted outputs and the true outputs are shown in Figures 4 and 5. The states $x_{i,t}$ and their estimates $\hat{x}_{i,t}$ versus $t$ are shown in Figure 6.
Table 2. The parameter estimates and their errors with $\sigma^2 = 0.15^2$.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>$t$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_{11}$</th>
<th>$b_{12}$</th>
<th>$b_{21}$</th>
<th>$b_{22}$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$c$</th>
<th>$\delta(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGLS</td>
<td>100</td>
<td>-0.00316</td>
<td>0.05268</td>
<td>0.07544</td>
<td>0.05230</td>
<td>-0.06414</td>
<td>-0.05137</td>
<td>0.42414</td>
<td>0.34337</td>
<td>0.09662</td>
<td>77.79370</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.02327</td>
<td>0.12465</td>
<td>0.06667</td>
<td>0.02426</td>
<td>-0.08934</td>
<td>-0.03437</td>
<td>0.41314</td>
<td>0.31430</td>
<td>-0.00151</td>
<td>73.30282</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.16392</td>
<td>0.16651</td>
<td>0.03428</td>
<td>0.08944</td>
<td>-0.05452</td>
<td>-0.03641</td>
<td>0.42364</td>
<td>0.26468</td>
<td>0.02093</td>
<td>53.53272</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.26324</td>
<td>0.20629</td>
<td>0.03025</td>
<td>0.11604</td>
<td>-0.04730</td>
<td>-0.07382</td>
<td>0.41937</td>
<td>0.22886</td>
<td>0.01985</td>
<td>38.31102</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>-0.36424</td>
<td>0.25406</td>
<td>0.03468</td>
<td>0.13072</td>
<td>-0.04674</td>
<td>-0.10435</td>
<td>0.41210</td>
<td>0.19167</td>
<td>0.03129</td>
<td>24.68460</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>-0.40788</td>
<td>0.25214</td>
<td>0.04352</td>
<td>0.12255</td>
<td>-0.04747</td>
<td>-0.12509</td>
<td>0.40931</td>
<td>0.18626</td>
<td>0.05481</td>
<td>18.65007</td>
</tr>
<tr>
<td>F-TS-RGLS</td>
<td>100</td>
<td>-0.17207</td>
<td>0.13225</td>
<td>0.08149</td>
<td>0.04762</td>
<td>-0.08050</td>
<td>-0.06387</td>
<td>0.38554</td>
<td>0.24812</td>
<td>-0.24016</td>
<td>67.37094</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.26379</td>
<td>0.21163</td>
<td>0.05110</td>
<td>0.10094</td>
<td>-0.07089</td>
<td>-0.08849</td>
<td>0.39815</td>
<td>0.22078</td>
<td>-0.20425</td>
<td>51.91655</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.41609</td>
<td>0.27122</td>
<td>0.04448</td>
<td>0.17605</td>
<td>-0.04003</td>
<td>-0.19822</td>
<td>0.40894</td>
<td>0.15349</td>
<td>-0.06039</td>
<td>20.84531</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.44506</td>
<td>0.28907</td>
<td>0.05723</td>
<td>0.16636</td>
<td>-0.05408</td>
<td>-0.19663</td>
<td>0.40519</td>
<td>0.15223</td>
<td>-0.06005</td>
<td>12.54315</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>-0.45785</td>
<td>0.28984</td>
<td>0.06293</td>
<td>0.16491</td>
<td>-0.06176</td>
<td>-0.19095</td>
<td>0.40193</td>
<td>0.14834</td>
<td>0.02730</td>
<td>8.43565</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>-0.45744</td>
<td>0.29468</td>
<td>0.06717</td>
<td>0.15764</td>
<td>-0.06434</td>
<td>-0.19611</td>
<td>0.40149</td>
<td>0.14983</td>
<td>0.05888</td>
<td>4.92240</td>
</tr>
<tr>
<td>True values</td>
<td></td>
<td>-0.46000</td>
<td>0.30000</td>
<td>0.08000</td>
<td>0.17000</td>
<td>-0.07000</td>
<td>-0.22000</td>
<td>0.40000</td>
<td>0.15000</td>
<td>0.08000</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. The parameter estimates and their errors with $\sigma^2 = 0.10^2$.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>$t$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$b_{11}$</th>
<th>$b_{12}$</th>
<th>$b_{21}$</th>
<th>$b_{22}$</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$c$</th>
<th>$\delta(%)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGLS</td>
<td>100</td>
<td>-0.01344</td>
<td>0.08368</td>
<td>0.04929</td>
<td>0.12366</td>
<td>-0.04534</td>
<td>-0.08160</td>
<td>0.42729</td>
<td>0.30373</td>
<td>0.01047</td>
<td>71.76858</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.04819</td>
<td>0.15699</td>
<td>0.03634</td>
<td>0.12767</td>
<td>-0.03469</td>
<td>-0.10328</td>
<td>0.41799</td>
<td>0.25865</td>
<td>-0.09361</td>
<td>65.81981</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.21525</td>
<td>0.21335</td>
<td>0.02114</td>
<td>0.18346</td>
<td>-0.02157</td>
<td>-0.15342</td>
<td>0.42234</td>
<td>0.19440</td>
<td>-0.08012</td>
<td>42.77206</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.33044</td>
<td>0.24492</td>
<td>0.03458</td>
<td>0.18170</td>
<td>-0.03375</td>
<td>-0.16150</td>
<td>0.41530</td>
<td>0.17384</td>
<td>-0.06981</td>
<td>30.24311</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>-0.39697</td>
<td>0.26064</td>
<td>0.04731</td>
<td>0.17039</td>
<td>-0.04645</td>
<td>-0.16265</td>
<td>0.40883</td>
<td>0.15616</td>
<td>-0.02604</td>
<td>19.40043</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>-0.42738</td>
<td>0.27001</td>
<td>0.05540</td>
<td>0.15703</td>
<td>-0.05159</td>
<td>-0.16863</td>
<td>0.40674</td>
<td>0.15902</td>
<td>0.00337</td>
<td>14.22233</td>
</tr>
<tr>
<td>F-TS-RGLS</td>
<td>100</td>
<td>-0.17072</td>
<td>0.12873</td>
<td>0.08786</td>
<td>0.04669</td>
<td>-0.08179</td>
<td>-0.08320</td>
<td>0.39734</td>
<td>0.23700</td>
<td>-0.08772</td>
<td>56.19096</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.26673</td>
<td>0.21478</td>
<td>0.05515</td>
<td>0.09642</td>
<td>-0.07483</td>
<td>-0.09063</td>
<td>0.40715</td>
<td>0.20974</td>
<td>-0.07335</td>
<td>40.40498</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.45453</td>
<td>0.28197</td>
<td>0.04590</td>
<td>0.18088</td>
<td>-0.04140</td>
<td>-0.21202</td>
<td>0.41195</td>
<td>0.13243</td>
<td>0.04761</td>
<td>8.46768</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.46578</td>
<td>0.29562</td>
<td>0.05933</td>
<td>0.17264</td>
<td>-0.05829</td>
<td>-0.20062</td>
<td>0.40718</td>
<td>0.14093</td>
<td>0.06403</td>
<td>4.89664</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>-0.47371</td>
<td>0.29198</td>
<td>0.06315</td>
<td>0.17165</td>
<td>-0.06534</td>
<td>-0.19084</td>
<td>0.40294</td>
<td>0.13951</td>
<td>0.06910</td>
<td>5.33512</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>-0.46830</td>
<td>0.29693</td>
<td>0.06704</td>
<td>0.16129</td>
<td>-0.06756</td>
<td>-0.19803</td>
<td>0.40240</td>
<td>0.14385</td>
<td>0.08923</td>
<td>4.02737</td>
</tr>
<tr>
<td>True values</td>
<td></td>
<td>-0.46000</td>
<td>0.30000</td>
<td>0.08000</td>
<td>0.17000</td>
<td>-0.07000</td>
<td>-0.22000</td>
<td>0.40000</td>
<td>0.15000</td>
<td>0.08000</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. The simulated input-output data versus $t$. 
Figure 2. The RGLS parameter estimation errors $\delta$ versus $t$ with different variances.

Figure 3. The F-TS-RGLS parameter estimation errors versus $t$ with different variances.

Figure 4. The true outputs and the RGLS predicted outputs.
From the simulation results in Tables 2 and 3 and Figures 1–6, we can draw the following conclusions.

- The estimation errors of the RGLS algorithm and the F-TS-RGLS algorithm became smaller with the data length increasing. This means that the proposed algorithms are effective.
- Under the same noise levels, the state estimator-based F-TS-RGLS algorithm could generate more accurate parameter estimates than the state estimator-based RGLS algorithm.
- The F-TS-RGLS algorithm could decrease the dimensions of the covariance matrices and improve the computational efficiency.
- The state estimates obtained from the bilinear state estimator could track their true values as $t$ increased.
6. Conclusions

This paper presents a recursive parameter and state estimation algorithm on the basis of the hierarchical identification principle for bilinear systems. In this approach, the system parameters and states could be estimated interactively. The bilinear state estimator was derived based on the Kalman filter. Then, the state estimator-based RGLS algorithm and the state estimator-based F-TS-RGLS algorithm were proposed based on the data filtering technique and the decomposition-coordination principle. The simulation results showed that the state estimator-based F-TS-RGLS algorithm had higher parameter estimation accuracy compared with the state estimator-based RGLS algorithm. Moreover, the state estimator-based F-TS-RGLS algorithm could greatly improve the computational efficiency. The methods proposed in this paper could be combined the particle Monte Carlo methods [56,57] and some statistical methods to study the model selection and parameter estimation [58–60] for different systems [61–66] and could be applied to other fields [67–72] such as communication networks [73–75].

Author Contributions: Conceptualization and methodology, X.Z. and F.D.; software, X.Z. and L.X.; validation and analysis, A.A. and T.H. Finally, all the authors have read and approved the final manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61873111), the 111 Project (B12018), and the Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX18_1854).

Conflicts of Interest: The authors declare no conflict of interest.

References
32. Verdult, V.; Verhaegen, M. Identification of multivariable bilinear state space systems based on subspace techniques and separable least squares optimization. *Int. J. Control* 2001, 74, 1824–1836. [CrossRef]
33. Larkowski, T.; Linden, J.G.; Vinsonneau, B.; Burnham, K.J. Frisch scheme identification for dynamic diagonal bilinear models. *Int. J. Control* 2009, 82, 1591–1604. [CrossRef]
40. Ding, F.; Liu, G.; Liu, X.P. Parameter estimation with scarce measurements. *Automatica* 2011, 47, 1646–1655. [CrossRef]
46. Van Wingerden, J.W.; Verhaegen, M. Subspace identification of bilinear and LPV systems for open-and closed-loop data. *Automatica* 2009, 45, 372–381. [CrossRef]
50. Schön, T.B.; Wills, A.; Ninness, B. System identification of nonlinear state-space models. *Automatica* 2011, 47, 39–49. [CrossRef]
59. Pan, J.; Ma, H.; Jiang, X.; Ding, W.; Ding, F. Adaptive gradient-based iterative algorithm for multivariate controlled autoregressive moving average systems using the data filtering technique. *Complexity* 2018, 2018, 9598307. [CrossRef]
60. Pan, J.; Li, W.; Zhang, H.P. Control algorithms of magnetic suspension systems based on the improved double exponential reaching law of sliding mode control. *Int. J. Control Autom. Syst.* 2018, 16, 2878–2887. [CrossRef]


68. Cao, Y.; Zhang, Y.; Wen, T.; Li, P. Research on dynamic nonlinear input prediction of fault diagnosis based on fractional differential operator equation in high-speed train control system. *Chaos* 2019, 29, 013130. [CrossRef]


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).