

Article

# Variational Bayesian Iterative Estimation Algorithm for Linear Difference Equation Systems

Junxia Ma <sup>1</sup> , Qiuling Fei <sup>1</sup>, Fan Guo <sup>2</sup> and Weili Xiong <sup>1,\*</sup>

<sup>1</sup> Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; junxia.20@163.com (J.M.); 6181905005@stu.jiangnan.edu.cn (Q.F.)

<sup>2</sup> Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada; fg3@ualberta.ca

\* Correspondence: weili\_xiong@jiangnan.edu.cn

Received: 30 October 2019; Accepted: 20 November 2019; Published: 22 November 2019



**Abstract:** Many basic laws of physics or chemistry can be written in the form of differential equations. With the development of digital signals and computer technology, the research on discrete models has received more and more attention. The estimates of the unknown coefficients in the discretized difference equation can be obtained by optimizing certain criterion functions. In modern control theory, the state-space model transforms high-order differential equations into first-order differential equations by introducing intermediate state variables. In this paper, the parameter estimation problem for linear difference equation systems with uncertain noise is developed. By transforming system equations into state-space models and on the basis of the considered priors of the noise and parameters, a variational Bayesian iterative estimation algorithm is derived from the observation data to obtain the parameter estimates. The unknown states involved in the variational Bayesian algorithm are updated by the Kalman filter. A numerical simulation example is given to validate the effectiveness of the proposed algorithm.

**Keywords:** iterative algorithm; variational Bayesian; parameter estimation; Kalman filter

## 1. Introduction

Differential equations arise in many instances when using the mathematical models to describe phenomena in economics, engineering, biology, etc. [1,2]. Considering the following  $n$ th-order linear differential equation with zero initial conditions

$$\begin{aligned} y_k^{(n)} + \alpha_1 y_k^{(n-1)} + \alpha_2 y_k^{(n-2)} + \cdots + \alpha_{n-1} y_k^{(1)} + \alpha_n y_k \\ = \beta_1 u_k^{(n-1)} + \beta_2 u_k^{(n-2)} + \cdots + \beta_{n-1} u_k^{(1)} + \beta_n u_k, \end{aligned} \quad (1)$$

where  $y_k$  is the system output,  $u_k$  is the system input,  $k$  is the time variable,  $\alpha_i$  and  $\beta_i$  are the unknown real parameters. Each derivative symbol is defined as

$$\begin{aligned} y_k^{(i)} &= \frac{d^i y_k}{d^i k}, & y_k^{(2)} &= \frac{d^2 y_k}{d^2 k} = \ddot{y}_k, & y_k^{(1)} &= \frac{d y_k}{d k} = \dot{y}_k, \\ u_k^{(i)} &= \frac{d^i u_k}{d^i k}, & u_k^{(2)} &= \frac{d^2 u_k}{d^2 k} = \ddot{u}_k, & u_k^{(1)} &= \frac{d u_k}{d k} = \dot{u}_k. \end{aligned}$$

Although the laws of the systems are continuously evolved over time, the sampled values of the systems are always discretized. In order to obtain the estimates of the unknown parameters in the system mathematical models, there are usually two approaches to further process the differential

equation models. One is called the frequency-domain analysis methods, and the other is called the time-domain analysis methods [3–6].

For linear continuous systems, Laplace transform of Equation (1) under zero initial conditions is given by

$$(s^{n-1} + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \dots + \alpha_{n-1} s + \alpha_n) Y(s) = (\beta_1 s^{n-1} + \beta_2 s^{n-2} + \dots + \beta_{n-1} s + \beta_n) U(s),$$

where the variable  $s$  is the laplace operator,  $Y(s)$  and  $U(s)$  are the Laplace transform of  $y_k$  and  $u_k$ :  $Y(s) := \mathcal{L}[y_k]$ ,  $U(s) := \mathcal{L}[u_k]$ . Then the transfer function of the system is written as

$$G(s) := \frac{Y(s)}{U(s)} = \frac{\beta_1 s^{n-1} + \beta_2 s^{n-2} + \dots + \beta_{n-1} s + \beta_n}{s^{n-1} + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \dots + \alpha_{n-1} s + \alpha_n}.$$

So the output equation can be further expressed as

$$Y(s) = G(s)U(s).$$

The system will be excited by designing different input signals to obtain different output responses. The parameters of the system are further obtained according to the analysis of the output responses [7–9]. By using the Laplace transform, the transfer function is one of the most important mathematical models in control theory. Under the zero initial conditions, the Laplace transform of the system output divided by the Laplace transform of the system input is called the transfer function. However, the transfer function is only suitable for describing linear constant coefficient differential equation systems.

In the time-domain analysis method, the system is represented in discrete time and the sampling interval is assumed to be one time unit. In this case, the discrete-time operator equations or difference equations are necessary [10–13]. A typical controlled autoregressive difference equation is written as

$$y_k + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_n y_{k-n} = b_1 u_{k-1} + b_2 u_{k-2} + \dots + b_n u_{k-n}.$$

Then the output  $y_k$  can be further written as a form of determining the next value for given previous observations:

$$\begin{aligned} y_k &= -a_1 y_{k-1} - a_2 y_{k-2} - \dots - a_n y_{k-n} + b_1 u_{k-1} + b_2 u_{k-2} + \dots + b_n u_{k-n} \\ &= \boldsymbol{\psi}_k^T \boldsymbol{\vartheta}, \end{aligned} \tag{2}$$

where

$$\begin{aligned} \boldsymbol{\psi}_k &:= [-y_{k-1}, -y_{k-2}, \dots, -y_{k-n}, u_{k-1}, u_{k-2}, \dots, u_{k-n}]^T \in \mathbb{R}^{2n}, \\ \boldsymbol{\vartheta} &:= [a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n]^T \in \mathbb{R}^{2n}. \end{aligned}$$

From Equation (2), it is easy to find that the calculation of  $y_k$  from past data depends on the parameters in  $\boldsymbol{\vartheta}$ . Call the calculated value  $\hat{y}_{k|\boldsymbol{\vartheta}}$  and write

$$\hat{y}_{k|\boldsymbol{\vartheta}} = \boldsymbol{\psi}_k^T \boldsymbol{\vartheta}.$$

The solution of obtaining the estimates of the parameters is to optimize a criterion function. The effective estimates make the prediction error  $\varepsilon(k, \boldsymbol{\vartheta}) = y_k - \hat{y}_{k|\boldsymbol{\vartheta}}$  as small as possible. Given a batch of inputs and outputs over a time interval  $1 \leq k \leq N$ :

$$Z^N = [u_1, y_1, u_2, y_2, \dots, u_N, y_N].$$

The quadratic criterion function is defined as [14]:

$$V_N(\boldsymbol{\theta}, Z^N) = \frac{1}{N} \sum_{k=1}^N [y_k - \hat{y}_{k|\boldsymbol{\theta}}]^2.$$

The process of computing the estimate  $\hat{\boldsymbol{\theta}}_N$  is the process of finding the solution of the following optimization problem:

$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} V_N(\boldsymbol{\theta}, Z^N),$$

where ‘ $\underset{\boldsymbol{\theta}}{\operatorname{argmin}}$ ’ represents the value of the considered parameter vector when the objective function takes the minimum value. The classical least squares-based and gradient descent-based numerical methods have been applied to search the estimates of the parameters by minimizing the criterion function [15–18]. For example, Xu et al. developed a least squares-based iterative parameter estimation algorithm for dynamic systems based on the parameter decomposition by using the hierarchical identification principle [19].

However, the gradient and least squares-based iterative algorithm only obtain the point estimates of parameters based on the known noise variance. They minimize a cost function on the parameter vector and then obtain the estimate of the parameter vector. Different from the stochastic gradient and least squares algorithm, the variational Bayesian (VB) algorithm needs the priors of parameters and then iteratively updates the posterior expression of the parameters by maximizing the lower bound marginal likelihood function [20,21]. The VB algorithm can obtain the distributions of the unknown parameters rather than the point estimates. For example, Yang and Yin applied the VB inference to estimate the parameters of finite impulse response (FIR) systems with randomly missing output data [22].

In this paper, the difference equation is further transformed to the state-space model. In modern control theory, the state-space model is widely used to describe the dynamics of systems, because it can not only reflect the internal states of systems but also reveals the relationship between the intermediate states and the external input and output variables [23–28]. A general linear state-space model

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \vdots \\ x_{n,k+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} u_k,$$

$$y_k = [c_1, c_2, \dots, c_n] \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n,k} \end{bmatrix} + v_k,$$

where  $y_k$  is the system output,  $u_k$  is the system input,  $v_k$  is the random noise, and  $x_{i,k}$  is the unmeasurable internal state, which establishes the link between system internal information and external measurements is presented. However, state estimation brings difficulty to the parameter estimation [29]. For linear Gaussian systems, the Kalman filter can achieve optimal estimation performance with a small amount of computational load, which is usually the preferred method of the state estimation for state-space models [30–32].

Based on the previous iterative identification schemes [33–36], this paper concerns a VB-based iterative estimation algorithm for difference equation models with unknown measurement noise. Different from the work studied in [37], which used an open-loop state observer to roughly calculate the system states without considering the effect of measurement noise in the procedure of states estimation, this paper considers the problem of state estimation by using the Kalman filter. The main contributions of this paper are as follows.

- By defining internal state variables, the difference equation model is rewritten in the form of the state-space model.
- A Kalman filter based VB (VB-KF) estimation algorithm is provided to estimate both the model states and parameters interactively. Based on the estimated states, the VB algorithm calculates the system parameters and the noise variance. Then the states are updated by means of the Kalman filter with new obtained parameters and noise variance.
- Different from the existing point parameter estimation methods, the proposed algorithm can give the distributions of the system parameters and noise variance to reflect the uncertainties of their estimates.

The remainder of this paper is organized as follows. Section 2 introduces the state-space model. Section 3 derives a VB iterative algorithm to estimate the unknown parameters of the systems by using the Kalman filter. The simulation is presented in Section 4 to verify the proposed method. Section 5 gives some conclusions.

## 2. Transformation of the Difference Equation

Reconsidering a general  $n$ th-order difference equation in Equation (2), in order to transform the difference equation to the state-space model,  $n$  state variables are needed. The choices of the state variables are not unique; they can be measurable or unmeasurable. Usually, the state variables which can help to convert the state equation into a standard form are selected as the intermediate states. Here, we introduce  $n$  state variables as follows:

$$\begin{aligned}
 x_{1,k} &= y_k, \\
 x_{2,k} &= y_{k+1} - b_1 u_k, \\
 x_{3,k} &= y_{k+2} - b_1 u_{k+1} - b_2 u_k, \\
 x_{4,k} &= y_{k+3} - b_1 u_{k+2} - b_2 u_{k+1} - b_3 u_k, \\
 &\vdots \\
 x_{n-1,k} &= y_{k+n-2} - b_1 u_{k+n-3} - b_2 u_{k+n-4} - \dots - b_{n-2} u_k, \\
 x_{n,k} &= -a_n y_{k-1} - a_{n-1} (y_k - b_1 u_{k-1}) - a_{n-2} (y_{k+1} - b_1 u_k - b_2 u_{k-1}) + \dots \\
 &\quad - a_1 (y_{k+n-2} - b_1 u_{k+n-3} - b_2 u_{k+n-4} - \dots - b_{n-2} u_k - b_{n-1} u_{k-1}) + b_n u_k.
 \end{aligned}
 \tag{3}$$

Further, replacing the time variable  $k$  with  $k + 1$  on both sides of Equation (3) gives the following equations:

$$\begin{aligned}
 x_{1,k+1} &= y_{k+1} = x_{2,k} + b_1 u_k, \\
 x_{2,k+1} &= y_{k+2} - b_1 u_{k+1} = x_{3,k} + b_2 u_k, \\
 &\vdots \\
 x_{j,k+1} &= y_{k+j} - b_1 u_{k+j-1} - \dots - b_{j-1} u_{k+1} = x_{j+1,k} + b_j u_k, \\
 &\vdots \\
 x_{n-1,k+1} &= x_{n,k} + b_{n-1} u_k, \\
 x_{n,k+1} &= -a_n x_{1,k} - a_{n-1} x_{2,k} - \dots - a_1 x_{n,k} + b_n u_k.
 \end{aligned}
 \tag{4}$$

Define the state vector  $\mathbf{x}_k := [x_{1,k}, x_{2,k}, \dots, x_{n,k}]^T \in \mathbb{R}^n$ . Rewrite the state equation for each state variable in Equation (4) by matrix representation; the observability canonical form state-space model of the difference Equation (2) can be described as

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ \vdots \\ x_{n-1,k+1} \\ x_{n,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n-1,k} \\ x_{n,k} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} u_k, \tag{5}$$

$$y_k = [1 \ 0 \ 0 \ \cdots \ 0] \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{n-1,k} \\ x_{n,k} \end{bmatrix}. \tag{6}$$

Further, we rewrite (5) and (6) as

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}u_k, \tag{7}$$

$$y_k = \mathbf{C}\mathbf{x}_k, \tag{8}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix},$$

$$\mathbf{C} = [1, 0, 0, \dots, 0].$$

The matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is called the system matrix, the vector  $\mathbf{B} \in \mathbb{R}^{n \times 1}$  is the system input vector, and the vector  $\mathbf{C} \in \mathbb{R}^{1 \times n}$  is the system output vector. Equation (7) is called the state equation, and Equation (8) is called the output equation. Specially, for the multi-input multi-output systems, the input and output vectors will be extended to matrices. It is easy to find that those unknown parameters are in the system matrix and input vector. In order to estimate the unknown parameters of the system, the output equation needs to be redescribed to include all unknown parameters. Transforming the state equation, the output equation in Equation (8) can be further written as

$$\begin{aligned} y_k &= x_{1,k} \\ &= x_{2,k-1} + b_1 u_{k-1} \\ &= x_{3,k-2} + b_2 u_{k-2} + b_1 u_{k-1} \\ &= x_{4,k-3} + b_3 u_{k-3} + b_2 u_{k-2} + b_1 u_{k-1} \\ &\vdots \\ &= -a_n x_{1,k-n} - a_{n-1} x_{2,k-n} - \cdots - a_1 x_{n,k-n} + b_n u_{k-n} + \cdots + b_2 u_{k-2} + b_1 u_{k-1}. \end{aligned} \tag{9}$$

Define the parameter vector  $\boldsymbol{\theta}$ , the information vector  $\boldsymbol{\psi}_k$  as

$$\boldsymbol{\theta} := [a_n, a_{n-1}, \dots, a_1, b_n, b_{n-1}, \dots, b_1]^T \in \mathbb{R}^{2n},$$

$$\boldsymbol{\psi}_k := [-x_{1,k-n}, -x_{2,k-n}, \dots, -x_{n,k-n}, u_{k-n}, u_{k-n+1}, \dots, u_{k-1}]^T \in \mathbb{R}^{2n}.$$

Considering that there will be some disturbance in the actual systems, introduce a noise term  $v_k$  to the output equation. Then the output in Equation (9) can be simplified into a vector form

$$y_k = \psi_k^T \theta + v_k, \tag{10}$$

where  $v_k$  is the random disturbance, which can be the white noise or colored noise. For practical systems, the effects of noise should be considered when dealing with the problem of parameter estimation.

**Remark 1.** *By introducing internal state variables, higher-order difference equations are transformed into first-order difference equations, which simplify the forms of the output equations of the models.*

This paper develops a combined parameters, intermediate states, and noise variance estimation algorithm for linear difference equation systems. On the basis of the model shown in Equation (10), the next section will drive a VB-KF iterative algorithm to get the solution for the problem of the parameter estimation.

### 3. The Kalman Filter-Based Variational Bayesian Iterative Estimation Algorithm

#### 3.1. The Variational Bayesian Parameter Estimation Algorithm

The VB algorithm is formed by introducing the variational approximation theory. It is difficult to deal with the integrations involved in the general posterior probability solution. The VB algorithm introduces a cluster of free factored distribution to approximate the complicated posterior probability, which avoids the direct multiple integrations on marginal distribution [38,39]. In this paper, the parameter vector together with the corresponding estimated error covariance matrix and noise variance are inferred by using the VB approach.

The unknown parameters that need to be estimated include the system model parameter vector  $\theta$ , the variance of parameter estimates  $\zeta^{-1}I$ , and the system noise variance  $\sigma^{-1}$ . Assuming their priors follow Gaussian, Gamma, and Gamma distributions, respectively,

$$\begin{aligned} p(\theta) &= N(\mathbf{0}, \zeta^{-1}I), \\ p(\zeta) &= \text{Gamma}(\zeta; d_0, e_0), \\ p(\sigma) &= \text{Gamma}(\sigma; f_0, h_0). \end{aligned}$$

The VB algorithm introduces a free distribution  $q(\cdot)$  as the approximate distribution of parameter to simplify the calculation of logarithmic likelihood function of the measurement data. Then the probability distribution of observation can be expressed as

$$\begin{aligned} \log p(y_{1:k}) &= \log p(\theta, \zeta, \sigma, y_{1:k}) - \log p(\theta, \zeta, \sigma | y_{1:k}) \\ &= \int q(\theta, \zeta, \sigma) \times \log \frac{p(\theta, \zeta, \sigma, y_{1:k})}{q(\theta, \zeta, \sigma)} d_\theta d_\zeta d_\sigma + \int q(\theta, \zeta, \sigma) \times \log \frac{q(\theta, \zeta, \sigma)}{p(\theta, \zeta, \sigma | y_{1:k})} d_\theta d_\zeta d_\sigma \\ &= F[q(\theta, \zeta, \sigma)] + \text{KL}[q(\theta, \zeta, \sigma) || p(\theta, \zeta, \sigma, y_{1:k})], \end{aligned}$$

where  $F[q(\theta, \zeta, \sigma)]$  is the lower bound of  $\log p(y_{1:k})$  (referring to the Jensen inequality [38]), that is

$$\log p(y_{1:k}) \geq \int q(\theta, \zeta, \sigma) \times \log \frac{p(\theta, \zeta, \sigma, y_{1:k})}{q(\theta, \zeta, \sigma)} d_\theta d_\zeta d_\sigma. \tag{11}$$

$\text{KL}[(q(\theta, \zeta, \sigma) || p(\theta, \zeta, \sigma | y_{1:k}))]$  is the Kullback–Leibler divergence between the factored approximation  $q(\theta, \zeta, \sigma)$  and the true posterior distribution  $p(\theta, \zeta, \sigma | y_{1:k})$ , which demonstrates the similarity between the two probability distributions.

We assume that the joint probability distribution is decomposable:

$$q(\boldsymbol{\theta}, \zeta, \sigma) = q(\boldsymbol{\theta})q(\zeta)q(\sigma).$$

We can find that when  $q(\boldsymbol{\theta})q(\zeta)q(\sigma)$  and  $p(\boldsymbol{\theta}, \zeta, \sigma|y_{1:k})$  are identically distributed, the KL divergence between them can reach minimum. Therefore, these independent distributions  $q(\boldsymbol{\theta}), q(\zeta), q(\sigma)$  can be solved by minimizing

$$\begin{aligned} & \text{KL} [q(\boldsymbol{\theta})q(\zeta)q(\sigma) || p(\boldsymbol{\theta}, \zeta, \sigma|y_{1:k})] \\ & = \int q(\boldsymbol{\theta}, \zeta, \sigma) \times \log \frac{p(\boldsymbol{\theta}, \zeta, \sigma|y_{1:k})}{q(\boldsymbol{\theta})q(\zeta)q(\sigma)} d\boldsymbol{\theta} d\zeta d\sigma. \end{aligned} \tag{12}$$

Using the theorems from the calculus of variation [31], the optimal solutions for the parameter distributions satisfy the following equations

$$\log q(\boldsymbol{\theta}) = \iint \log p(\boldsymbol{\theta}, \zeta, \sigma, y_{1:k}) q(\zeta) q(\sigma) d\zeta d\sigma + c_{\boldsymbol{\theta}}, \tag{13}$$

$$\log q(\zeta) = \iint \log p(\boldsymbol{\theta}, \zeta, \sigma, y_{1:k}) q(\boldsymbol{\theta}) q(\sigma) d\boldsymbol{\theta} d\sigma + c_{\zeta}, \tag{14}$$

$$\log q(\sigma) = \iint \log p(\boldsymbol{\theta}, \zeta, \sigma, y_{1:k}) q(\boldsymbol{\theta}) q(\zeta) d\boldsymbol{\theta} d\zeta + c_{\sigma}, \tag{15}$$

where  $c_{\boldsymbol{\theta}}, c_{\zeta}, c_{\sigma}$  denote the terms which are independent of  $\boldsymbol{\theta}, \zeta, \sigma$ , respectively.

The initial joint probability  $p(\boldsymbol{\theta}, \zeta, \sigma, y_{1:k})$  can be factored as

$$p(\boldsymbol{\theta}, \zeta, \sigma, y_{1:k}) = N(\boldsymbol{0}, \zeta^{-1}\mathbf{I}) \text{Gamma}(\zeta; c_0, d_0) \text{Gamma}(\sigma; e_0, f_0) \prod_{k=1}^N N(y_k; \boldsymbol{\psi}_k^T \boldsymbol{\theta}, \sigma^{-1}). \tag{16}$$

Using Equation (16) into Equation (13) and ignoring items which are unrelated to  $\boldsymbol{\theta}$  gives

$$\begin{aligned} \log q^{(s+1)}(\boldsymbol{\theta}) &= \iint q(\zeta) q(\sigma) \left[ -\frac{1}{2} \boldsymbol{\theta}^T \zeta \boldsymbol{\theta} - \frac{1}{2} \sum_{k=1}^N (y_k - \boldsymbol{\psi}_k^T \boldsymbol{\theta})^T \sigma (y_k - \boldsymbol{\psi}_k^T \boldsymbol{\theta}) \right] d\zeta d\sigma + c_{\boldsymbol{\theta}} \\ &= -\frac{1}{2} \left\{ \boldsymbol{\theta}^T [\zeta^{(s)} \mathbf{I} + \sum_{k=1}^N \boldsymbol{\psi}_k \sigma^{(s)} \boldsymbol{\psi}_k^T] \boldsymbol{\theta} + \sum_{k=1}^N y_k^2 - \sum_{k=1}^N \boldsymbol{\theta}^T \boldsymbol{\psi}_k \sigma^{(s)} y_k - \sum_{k=1}^N y_k \sigma^{(s)} \boldsymbol{\psi}_k^T \boldsymbol{\theta} \right\} \\ & \quad + c_{\boldsymbol{\theta}}, \end{aligned} \tag{17}$$

where  $q^{(s+1)}(\cdot)$  denotes the approximation of  $q(\cdot)$  at the  $(s + 1)$ th iteration. From Equation (17) we can find that  $q^{(s+1)}(\boldsymbol{\theta})$  is the probability density function of a Gaussian distribution, and the system parameter  $\boldsymbol{\theta}$  can be updated as

$$\bar{\boldsymbol{\theta}}^{(s+1)} = \text{var}[\boldsymbol{\theta}^{(s+1)}] \sum_{k=1}^N \boldsymbol{\psi}_k \sigma^{(s)} y_k, \tag{18}$$

$$\text{var}[\boldsymbol{\theta}^{(s+1)}] = [\zeta^{(s)} \mathbf{I} + \sum_{k=1}^N \boldsymbol{\psi}_k \sigma^{(s)} \boldsymbol{\psi}_k^T]^{-1}. \tag{19}$$

Furthermore, we can calculate the mathematical expectation of  $\boldsymbol{\theta} \boldsymbol{\theta}^T$

$$\mathbb{E}^{(s+1)}(\boldsymbol{\theta} \boldsymbol{\theta}^T) = \text{var}(\boldsymbol{\theta}^{(s+1)}) + \boldsymbol{\theta}^{(s+1)} (\boldsymbol{\theta}^{(s+1)})^T. \tag{20}$$

According to Equation (14), we have

$$\begin{aligned} \log q^{(s+1)}(\zeta) &= \iint q(\boldsymbol{\theta}) q(\sigma) \left[ \frac{1}{2} \log \zeta - \frac{1}{2} (\boldsymbol{\theta}^T \zeta \boldsymbol{\theta} + (d_0 - 1) \log \zeta - e_0 \zeta) \right] d\boldsymbol{\theta} d\sigma + c_{\zeta} \\ &= -\left[ \frac{1}{2} \mathbb{E}^{(s+1)}(\boldsymbol{\theta}^T \boldsymbol{\theta}) + e_0 \right] \zeta + \left( \frac{1}{2} + d_0 - 1 \right) \log \zeta + c_{\zeta} \end{aligned}$$

$$= -\left\{ \frac{1}{2} \text{tr} \left[ \text{var}(\boldsymbol{\theta}^{(s+1)}) + \boldsymbol{\theta}^{(s+1)} (\boldsymbol{\theta}^{(s+1)})^\top \right] + e_0 \right\} \zeta + \left( \frac{1}{2} + d_0 - 1 \right) \log \zeta + c_\zeta,$$

which yields

$$q^{(s+1)}(\zeta) = \text{Gamma}(\zeta; d^{(s+1)}, e^{(s+1)}),$$

where  $d^{(s+1)}$  and  $e^{(s+1)}$  are the two parameters of the Gamma distribution and are given by

$$\begin{aligned} d^{(s+1)} &= \frac{1}{2} + d_0, \\ e^{(s+1)} &= e_0 + \frac{1}{2} \text{tr} \left[ \text{var}(\boldsymbol{\theta}^{(s+1)}) + \boldsymbol{\theta}^{(s+1)} (\boldsymbol{\theta}^{(s+1)})^\top \right]. \end{aligned}$$

Then, update the expectation of  $\zeta$  at iteration  $s + 1$  by

$$\bar{\zeta}^{(s+1)} = \frac{d^{(s+1)}}{e^{(s+1)}}. \tag{21}$$

Similarly, the variational form for  $q^{(s+1)}(\sigma)$  obeys

$$\begin{aligned} \log q^{(s+1)}(\sigma) &= \iint q(\boldsymbol{\theta}) q(\zeta) \left( (f_0 - 1) \log \sigma - h_0 \sigma + \frac{N}{2} \log \sigma - \frac{1}{2} \sum_{k=1}^N (y_k - \boldsymbol{\psi}_k^\top \boldsymbol{\theta})^\top \sigma (y_k - \boldsymbol{\psi}_k^\top \boldsymbol{\theta}) \right) d\boldsymbol{\theta} d\zeta + c_\sigma \\ &= (f_0 + \frac{N}{2} - 1) \log \sigma - \frac{1}{2} \left[ h_0 + \sum_{k=1}^N (y_k^2 - 2y_k \boldsymbol{\psi}_k^\top \boldsymbol{\theta}^{(s+1)} + \boldsymbol{\psi}_k^\top \boldsymbol{\Sigma}^{(s+1)} (\boldsymbol{\theta} \boldsymbol{\theta}^\top) \boldsymbol{\psi}_k) \right] \sigma + c_\sigma, \end{aligned}$$

which yields

$$q^{(s+1)}(\sigma) = \text{Gamma}(\sigma; f^{(s+1)}, h^{(s+1)}),$$

where  $f^{(s+1)}$  and  $h^{(s+1)}$  are given by

$$\begin{aligned} f^{(s+1)} &= \frac{N}{2} + f_0, \\ h^{(s+1)} &= h_0 + \sum_{k=1}^N \left\{ y_k^2 - 2y_k \boldsymbol{\psi}_k^\top \boldsymbol{\theta}^{(s+1)} + \boldsymbol{\psi}_k^\top \text{tr} \left[ \text{var}(\boldsymbol{\theta}^{(s+1)}) + \boldsymbol{\theta}^{(s+1)} (\boldsymbol{\theta}^{(s+1)})^\top \right] \boldsymbol{\psi}_k \right\}. \end{aligned}$$

Then  $\bar{\sigma}$  can be computed as

$$\bar{\sigma}^{(s+1)} = \frac{f^{(s+1)}}{h^{(s+1)}}. \tag{22}$$

### 3.2. The State Estimation Algorithm

Above all, we obtain the posterior probability distribution or optimal solution of system parameters and the noise variance. However, the state-space model contains both unknown system parameters and unavailable system states, which brings a difficulty to obtaining the parameter estimates. The Kalman filter is usually used to estimate the intermediate states for linear state-space systems with Gaussian noise. The state observer is a deterministic state estimation algorithm, which is simpler in structure but only obtains an approximate estimation of the system states. To improve the estimation accuracy, this paper adopts a modified Kalman filter to obtain state estimates under the Bayesian framework. According to the considered state-space model, the following Kalman filter algorithm [40] is designed to generate the state estimates

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{A}} \hat{\mathbf{x}}_k + \hat{\mathbf{b}} \bar{u}_k + \mathbf{L}_k (y_k - \mathbf{c} \hat{\mathbf{x}}_k), \quad \hat{\mathbf{x}}_1 = \mathbf{1}_n / p_0, \tag{23}$$

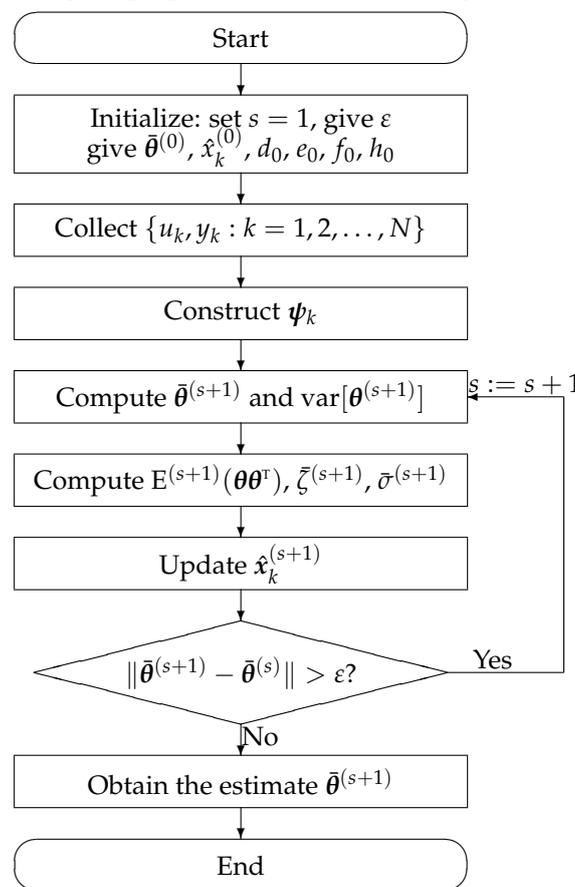
$$L_k = \hat{A}P_k c^T (\hat{\sigma}_k + cP_k c^T)^{-1}, \tag{24}$$

$$P_{k+1} = \hat{A}P_k \hat{A}^T - L_k c P_k \hat{A}^T, \quad P_1 = I_n. \tag{25}$$

Since the system parameters are unknown, we have to utilize the estimated parameters to construct the system parameter vector and matrix

$$\hat{A} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\hat{a}_n & -\hat{a}_{n-1} & -\hat{a}_{n-2} & \cdots & -\hat{a}_1 \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \hat{b} := \begin{bmatrix} \hat{b}_{1,k} \\ \hat{b}_{2,k} \\ \vdots \\ \hat{b}_{n-1,k} \\ \hat{b}_{n,k} \end{bmatrix} \in \mathbb{R}^n.$$

The flowchart of executing the proposed VB-KF iterative algorithm is shown in Figure 1.



**Figure 1.** The flowchart of the proposed variational Bayesian Kalman filter (VB-KF) iterative algorithm for computing parameters and states.

The iterative steps for calculating parameter and state estimates are summarized as follows:

1. Initialization: set  $s = 1$ ,  $x_k^{(0)} = \mathbf{1}_n$ ,  $\bar{\theta}^{(0)} = \mathbf{1}_{n_2}$ , and give a group of random positive values to parameters  $d_0, e_0, f_0, h_0$  and a small positive value to  $\epsilon$ .
2. Collect input-output data  $(u_k, y_k : k = 1, 2, \dots, N)$ , and contract information vector  $\psi_k$  with Equation (10).
3. Update  $\bar{\theta}^{(s+1)}$ ,  $\text{var}[\theta^{(s+1)}]$ ,  $E^{(s+1)}(\theta\theta^T)$ , and  $\bar{\zeta}^{(s+1)}$  with Equations (18)–(21), respectively.
4. Update noise variance  $\bar{\sigma}^{(s+1)}$  with Equation (22).
5. Use the new obtained parameter estimates and noise variance to update states with Kalman filter algorithm in Equation (23)–(25).

- Evaluate the relative change between  $\bar{\theta}^{(s+1)}$  and  $\bar{\theta}^{(s)}$ , if  $\|\bar{\theta}^{(s+1)} - \bar{\theta}^{(s)}\| \leq \varepsilon$  ( $\|\cdot\|$  means L2-norm), for a very small positive number  $\varepsilon$ , then stop the iteration procedure and get the parameter estimates  $\bar{\theta}^{(s+1)}$ ; or else, let  $s = s + 1$  and go to Step 3.

The proposed methods can combine other estimation algorithms [41–45] and the mathematical tools [46–49] to study the parameter identification problems of other linear and nonlinear systems with colored noise [50–52] and can be applied to other literatures [53–56].

#### 4. Simulation Example

Consider the following 3rd-order difference equation:

$$y_k + 0.50y_{k-1} + 0.32y_{k-2} + 0.18y_{k-3} = 1.00u_{k-1} + 0.55u_{k-2} + 0.97u_{k-3} + v_k.$$

Taking into account the interference of noise, transform the difference equation into the state-space model:

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \\ x_{3,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.18 & -0.32 & -0.50 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \end{bmatrix} + \begin{bmatrix} 1.00 \\ 0.55 \\ 0.97 \end{bmatrix} u_k,$$

$$y_k = [1 \ 0 \ 0] \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \end{bmatrix} + v_k.$$

The parameter vector to be estimated is

$$\theta = [0.18, 0.32, 0.50, 1.00, 0.55, 0.97].$$

In this simulation example, the input is selected as a persistent excitation signal sequence with zero mean and unit variance,  $v_k$  is Gaussian noise sequence with zero mean and variance  $\sigma^2$ . Of the collected data length  $N = 2000$ , the first 1000 sets of data ( $N_1$ ) are used for the training set and the rest of the ( $N_2$ ) sets are used for the testing set. Design different noise variances to simulate the noise interferences on the output signal. Use the proposed VB-KF iterative estimation algorithm to estimate the parameter vector.

Taking the calculated mean values of the Gaussian distributions for the parameters as the estimates of the unknown parameters, the proposed VB-KF parameter estimates and estimation error  $\delta = \|\bar{\theta}^{(s)} - \theta\| / \|\theta\|$  with the increasing of the iteration  $s$  under different noise variances are presented in Table 1 and Figure 2. The curves of parameter estimates against  $s$  are plotted in Figure 3. The estimated noise variances are shown in Table 2.

**Table 1.** The parameter estimates and estimation errors versus  $s$  under the proposed VB-KF algorithm.

$\sigma^2$	$s$	$a_1$	$a_2$	$\kappa_1$	$\kappa_2$	$\kappa_3$	$b_2$	$\delta$ (%)
0.10 <sup>2</sup>	1	0.01016	-0.02289	-0.03700	1.00397	0.53194	0.97784	40.69214
	2	-0.07361	0.00728	0.89702	0.94687	0.52043	0.96392	35.08236
	5	0.17474	0.30156	0.49471	0.97522	0.55029	0.99253	2.40308
	10	0.17243	0.31521	0.49920	0.97571	0.54883	0.99432	2.19306
	20	0.17246	0.31520	0.49918	0.97571	0.54883	0.99432	2.19278
	40	0.17246	0.31520	0.49918	0.97571	0.54883	0.99432	2.19278
0.50 <sup>2</sup>	1	-0.05182	-0.05809	0.04132	1.00842	0.54880	0.99711	39.39494
	2	-0.11574	0.04963	0.90089	0.93849	0.52609	0.97003	35.19921
	5	0.17688	0.31997	0.49349	0.96759	0.56015	1.00064	2.85674
	10	0.17674	0.32778	0.49767	0.96774	0.55874	1.00189	2.90030
	20	0.17676	0.32777	0.49764	0.96774	0.55874	1.00189	2.90017
	40	0.17676	0.32777	0.49764	0.96774	0.55874	1.00189	2.90017
1.00 <sup>2</sup>	1	-0.01785	0.06661	-0.15694	0.98365	0.51835	1.01729	45.26212
	2	-0.01462	-0.04935	0.75665	0.89482	0.48419	0.96419	31.18481
	5	0.17291	0.28867	0.46187	0.91682	0.50836	0.98695	6.59330
	10	0.17082	0.29553	0.46557	0.91739	0.50776	0.98825	6.41325
	20	0.17083	0.29553	0.46556	0.91739	0.50776	0.98825	6.41343
	40	0.17083	0.29553	0.46556	0.91739	0.50776	0.98825	6.41343
true values		0.18000	0.32000	0.50000	1.00000	0.55000	0.97000	

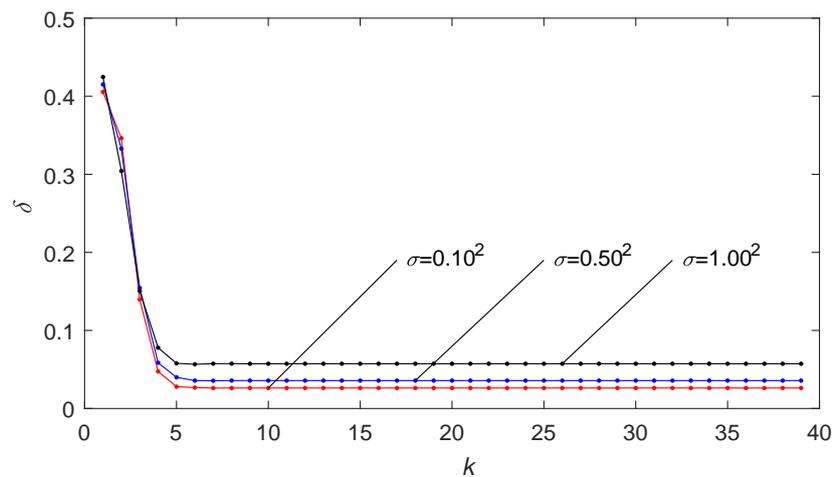


Figure 2. The VB parameter estimation error under different noise variance.

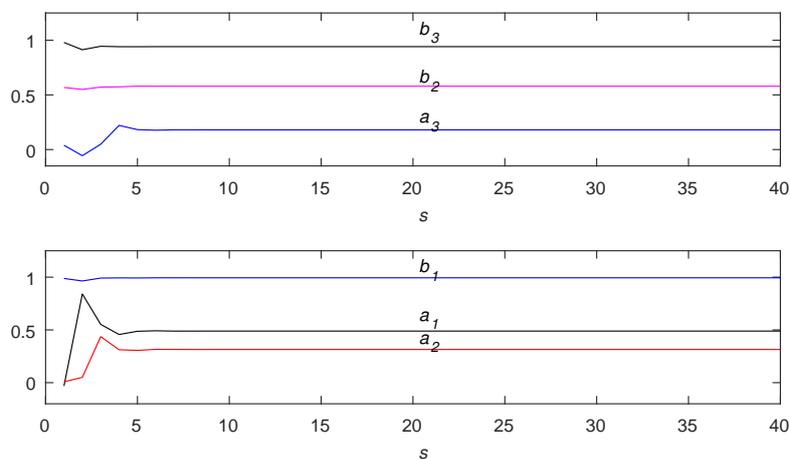


Figure 3. The VB parameter estimates against  $s$  with noise variance  $\sigma^2 = 0.50^2$ .

When the noise variance is  $\sigma^2 = 0.10^2$ , use the estimated parameters to construct the approximate difference equation:

$$y_k + 0.49918y_{k-1} + 0.31520y_{k-2} + 0.17246y_{k-3} = 0.97571u_{k-1} + 0.54883u_{k-2} + 0.99432u_{k-3} + v_k.$$

The computed prediction  $\hat{y}_k$  and the true output  $y_k$  are compared in Figure 4. From Figure 4, we can see that the predictions are very close to the true outputs.

For comparison, we also use the open-loop state observer to generate the estimates of unknown states [37]. Here, we combine the proposed VB parameter estimation method with state observer to form the state observer-based variational Bayesian (VB-SO) iterative estimation algorithm. In [57], the KF-based least squares iterative algorithm is used to identify the state space system, where the unknown process noise variance is set to one ( $R = 1.00$ ) in the Kalman filter. Here, we use this KF based-variational Bayesian algorithm (VB-KF<sub>2</sub>) and the proposed method (VB-KF<sub>1</sub>), where the noise variance is accurately computed, to estimate system parameters and states under different noises. Furthermore, the parameter estimation errors of the three different algorithms are compared in Figure 5 with  $\sigma^2 = 0.50^2$ .

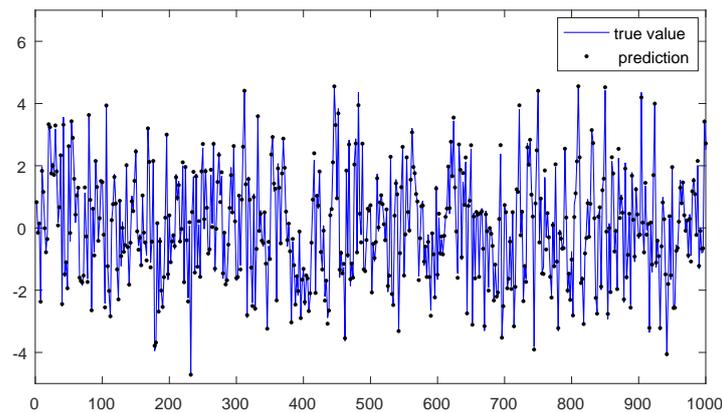


Figure 4. The true outputs and predictions with noise variance  $\sigma^2 = 0.10^2$ .

Table 2. The estimates of noise variance versus  $s$  under the proposed VB algorithm.

$\sigma^2/s$	1	2	3	5	10	15	40
$0.10^2$	0.2000	0.7509	0.0356	0.0280	0.0108	0.0108	0.0108
$0.50^2$	0.2000	0.9935	0.3010	0.2748	0.2664	0.2664	0.2664
$1.00^2$	0.2000	1.6252	1.0241	1.0097	0.9950	0.9950	0.9950

In addition, we adopt the other batch of data set for model validation. The parameter estimates with the least estimation error are chosen for the verification test. The root mean square errors (RMSE) of the three estimate models are computed as

$$\begin{aligned} \text{RMSE}_1 &= \sqrt{\frac{1}{N_2} \sum_{i=1001}^{2000} [y_i - \hat{y}_{\text{vbkf}_1}]^2} = 0.2833, \\ \text{RMSE}_2 &= \sqrt{\frac{1}{N_2} \sum_{i=1001}^{2000} [y_i - \hat{y}_{\text{vbso}}]^2} = 0.3007, \\ \text{RMSE}_3 &= \sqrt{\frac{1}{N_2} \sum_{i=1001}^{2000} [y_i - \hat{y}_{\text{vbkf}_2}]^2} = 0.2930. \end{aligned}$$

From the calculated results, it can be seen that the proposed VB-KF algorithm has the smallest root mean square error.

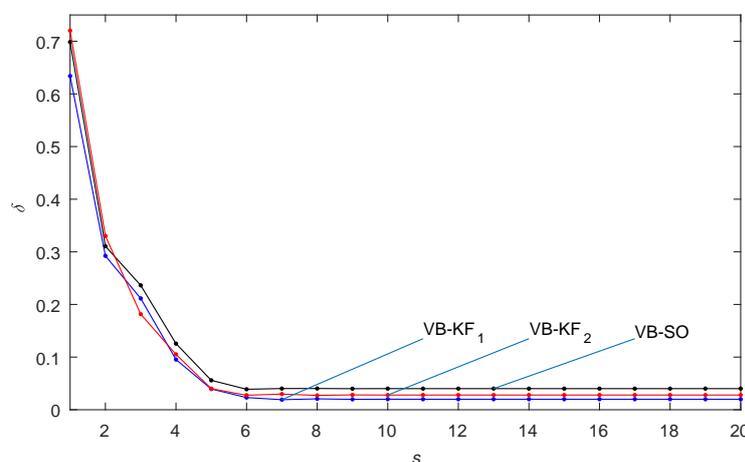


Figure 5. Comparison of estimation errors  $\delta$  versus  $s$  under different algorithms ( $\sigma^2 = 0.50^2$ ).

## 5. Conclusions

In this paper, a Kalman filter-based variational Bayesian iterative algorithm is proposed to estimate the unknown parameters for linear difference equation systems. The main advantage of the presented method is that it can give the distribution of the parameter rather than a point estimate. Besides, the proposed algorithm obtains the noise statistics and state estimation. From the simulation results shown in Tables 1 and 2 and Figures 2–5, it is easy to find the following conclusions.

- The parameter estimation errors are generally becoming smaller with iteration  $s$  increasing and with the noise variance decreasing. The estimated parameters are very close to the real values.
- The estimated noise variances in Table 2 show that variance estimates are relatively large in the previous period because of the inaccurate parameters and states, but with the increasing of the iteration  $s$ , the noise variance estimates can converge to their true value.
- Under the same batch of data and noise variance, the estimation accuracy of the proposed VB algorithm is higher than that of VB-SO and VB-KF<sub>2</sub> algorithms.
- The predicted outputs fit well with true outputs, which indicates the effectiveness of the proposed algorithm.

The proposed variational Bayesian iterative estimation algorithm in this paper for state space models from the observation data can combine some mathematical strategies [58–61] to study the parameter estimation algorithms of other stochastic systems with colored noise and can be applied to other fields [62–66].

**Author Contributions:** J.M., Q.F., F.G., and W.X. conceived and worked together to achieve this work; J.M. and F.G. discussed the method and the derivation process of the algorithm; Q.F. and W.X. compiled the simulation program by Matlab and analyzed the data; J.M. and Q.F. wrote the original draft. Finally, all the authors have read and approved the final manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (nos. 61803183, 21706096 and 61773182), the Fundamental Research Funds for the Central Universities (no. JUSRP11923), the 111 Project (no. B12018) and the Natural Science Foundation of Jiangsu Province (nos. BK20180591 and BK20160162).

**Acknowledgments:** The authors thank the anonymous reviewers for their careful reading and constructive comments.

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

## References

1. Dallon, J.C.; Evans, Emily J.; Grant, C.P.; Smith, W.V. Results from a differential equation model for cell motion with random switching show that the model cell velocity is asymptotically independent of force. *J. Differ. Equ.* **2019**, *268*, 301–317. [[CrossRef](#)]
2. Senthil, R.; Janarthanan, K.; Prakash, J. Nonlinear state estimation using fuzzy Kalman filter. *Ind. Eng. Chem. Res.* **2006**, *45*, 8678–8688. [[CrossRef](#)]
3. Sekizawa, S.; Inagaki, S.; Suzuki, T.; Hayakawa, S.; Tsuchida, N.; Tsuda, T.; Fujinami, H. Modeling and Recognition of Driving Behavior Based on Stochastic Switched ARX Model. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 593–606. [[CrossRef](#)]
4. Garus, D.; Gogolla, T.; Krebber, K.; Schliep, F. Brillouin optical-fiber frequency-domain analysis for distributed temperature and strain measurements. *J. Light. Technol.* **1997**, *15*, 654–662. [[CrossRef](#)]
5. John, V.; Ye, Z.; Kolwalkar, A. Investigation of anti-islanding protection of power converter based distributed generators using frequency domain analysis. *IEEE Trans. Power Electron.* **2004**, *19*, 1177–1183. [[CrossRef](#)]
6. Worden, K.; Hickey, D.; Haroon, M.; Adams, D.E. Nonlinear system identification of automotive dampers: A time and frequency-domain analysis. *Mech. Syst. Signal Process.* **2009**, *23*, 104–126. [[CrossRef](#)]
7. Xu, L.; Ding, F. Parameter estimation algorithms for dynamical response signals based on the multi-innovation theory and the hierarchical principle. *IET Signal Process.* **2017**, *11*, 228–237. [[CrossRef](#)]
8. Xu, L.; Chen, L.; Xiong, W.L. Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. *Nonlinear Dyn.* **2015**, *79*, 2155–2163. [[CrossRef](#)]

9. Xu, L. The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. *Signal Process.* **2016**, *120*, 660–667. [[CrossRef](#)]
10. Ding, F.; Liu, X.P.; Liu, G. Gradient based and least-squares based iterative identification methods for OE and OEMA systems. *Digit. Signal Process.* **2010**, *20*, 664–677. [[CrossRef](#)]
11. Pan, J.; Jiang, X.; Wan, X.K.; Ding, W. A filtering based multi-innovation extended stochastic gradient algorithm for multivariable control systems. *Int. J. Control Autom. Syst.* **2017**, *15*, 1189–1197. [[CrossRef](#)]
12. Ding, F. Decomposition based fast least squares algorithm for output error systems. *Signal Process.* **2013**, *93*, 1235–1242. [[CrossRef](#)]
13. Ding, F. Two-stage least squares based iterative estimation algorithm for CARARMA system modeling. *Appl. Math. Model.* **2013**, *37*, 4798–4808. [[CrossRef](#)]
14. Ljung, L. *System Identification: Theory for the User*, 2nd ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 1999.
15. Wan, L.J.; Ding, F. Decomposition- and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory. *Circuits Syst. Signal Process.* **2019**, *38*, 2971–2991. [[CrossRef](#)]
16. Ding, F.; Pan, J.; Alsaedi, A.; Hayat, T. Gradient-based iterative parameter estimation algorithms for dynamical systems from observation data. *Mathematics* **2019**, *7*, 428. [[CrossRef](#)]
17. Liu, S.Y.; Ding, F.; Xu, L.; Hayat, T. Hierarchical principle-based iterative parameter estimation algorithm for dual-frequency signals. *Circuits Syst. Signal Process.* **2019**, *38*, 3251–3268. [[CrossRef](#)]
18. Ding, F.; Liu, Y.J.; Bao, B. Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2012**, *226*, 43–55. [[CrossRef](#)]
19. Xu, L.; Ding, F.; Zhu, Q.M. Hierarchical Newton and least squares iterative estimation algorithm for dynamic systems by transfer functions based on the impulse responses. *Int. J. Syst. Sci.* **2019**, *50*, 141–151. [[CrossRef](#)]
20. Daunizeau, J.; Friston, K.J.; Kiebel, S.J. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Phys. D* **2009**, *238*, 2089–2118. [[CrossRef](#)]
21. Khatibisepehr, S.; Huang, B. Dealing with irregular data in soft sensors: Bayesian method and comparative study. *Ind. Eng. Chem. Res.* **2008**, *47*, 8713–8723. [[CrossRef](#)]
22. Yang, X.; Yin, S. Variational Bayesian inference for FIR models with randomly missing measurements. *IEEE Trans. Ind. Electron.* **2017**, *64*, 4217–4225. [[CrossRef](#)]
23. Ding, F. State filtering and parameter estimation for state space systems with scarce measurements. *Signal Process.* **2014**, *104*, 369–380. [[CrossRef](#)]
24. Liu, Y.J.; Ding, F.; Shi, Y. An efficient hierarchical identification method for general dual-rate sampled-data systems. *Automatica* **2014**, *50*, 962–970. [[CrossRef](#)]
25. Ding, F. Combined state and least squares parameter estimation algorithms for dynamic systems. *Appl. Math. Modell.* **2014**, *38*, 403–412. [[CrossRef](#)]
26. Viberg, M. Subspace-based methods for the identification of linear time-invariant systems. *Automatica* **1995**, *21*, 1835–1851. [[CrossRef](#)]
27. Mao, Y.; Guo, X.; Zhao, Y. A state space force identification method based on Markov parameters precise computation and regularization technique. *J. Sound Vib.* **2010**, *329*, 300–3019. [[CrossRef](#)]
28. Ding, F.; Liu, G.; Liu, X.P. Partially coupled stochastic gradient identification methods for non-uniformly sampled systems. *IEEE Trans. Autom. Control* **2010**, *55*, 1976–1981. [[CrossRef](#)]
29. Zhang, X.; Ding, F.; Yang, E.F. State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 1157–1173. [[CrossRef](#)]
30. Zhou, L.; Li, X.; Shan, L.; Xia, J. Hierarchical recursive least squares parameter estimation of non-uniformly sampled Hammerstein nonlinear systems based on Kalman filter. *J. Frankl. Inst.* **2017**, *354*, 4231–4246. [[CrossRef](#)]
31. Sarkka, S.; Nymmenmaa, A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Trans. Automat. Control* **2009**, *54*, 596–600. [[CrossRef](#)]
32. Huang, Y.; Zhang, Y.; Li, N.; Chambers, J. A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises. *IEEE Signal Process. Lett.* **2016**, *23*, 468–472. [[CrossRef](#)]
33. Li, M.H.; Liu, X.M.; Ding, F. The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle. *Int. J. Adapt. Control Signal Process.* **2019**, *33*, 1189–1211. [[CrossRef](#)]

34. Xu, L. The parameter estimation algorithms based on the dynamical response measurement data. *Adv. Mech. Eng.* **2017**, *9*, 1–12. [[CrossRef](#)]
35. Ma, H.; Pan, J.; Ding, F.; Xu, L.; Ding, W. Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like autoregressive moving average systems. *IET Control Theory Appl.* **2019**, *13*, 3040–3051. [[CrossRef](#)]
36. Xu, L. Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. *J. Comput. Appl. Math.* **2015**, *288*, 33–43. [[CrossRef](#)]
37. Xu, L.; Ding, F.; Gu, Y.; Alsaedi, A.; Hayat, T. A multi-innovation state and parameter estimation algorithm for a state space system with d-step state-delay. *Signal Process.* **2017**, *140*, 97–103. [[CrossRef](#)]
38. Tzikas, D.G.; Likas, A.C.; Galatsanos, N.P. The variational approximation for Bayesian inference. *IEEE Signal Process. Mag.* **2008**, *25*, 131–146. [[CrossRef](#)]
39. Latouche, P.; Birmele, E.; Ambroise, C. Variational Bayesian inference and complexity control for stochastic block models. *Stat. Model.* **2012**, *12*, 93–115. [[CrossRef](#)]
40. Ma, J.X.; Xiong, W.L.; Chen, J.; Feng, D. Hierarchical identification for multivariate Hammerstein systems by using the modified Kalman filter. *IET Control Theory Appl.* **2017**, *11*, 857–869. [[CrossRef](#)]
41. Ding, F.; Liu, G.; Liu, X.P. Parameter estimation with scarce measurements. *Automatica* **2011**, *47*, 1646–1655. [[CrossRef](#)]
42. Ding, J.; Chen, J.Z.; Lin, J.X.; Wan, L.J. Particle filtering based parameter estimation for systems with output-error type model structures. *J. Frankl. Inst.* **2019**, *356*, 5521–5540. [[CrossRef](#)]
43. Ding, J.; Chen, J.Z.; Lin, J.X.; Jiang, G.P. Particle filtering-based recursive identification for controlled auto-regressive systems with quantised output. *IET Control Theory Appl.* **2019**, *13*, 2181–2187 [[CrossRef](#)]
44. Wang, Y.J.; Ding, F.; Wu, M.H. Recursive parameter estimation algorithm for multivariate output-error systems. *J. Frankl. Inst.* **2018**, *355*, 5163–5181. [[CrossRef](#)]
45. Ding, F.; Wang, F.F.; Xu, L.; Wu, M.H. Decomposition based least squares iterative identification algorithm for multivariate pseudo-linear ARMA systems using the data filtering. *J. Frankl. Inst.* **2017**, *354*, 1321–1339. [[CrossRef](#)]
46. Yin, C.C.; Wang, C.W. The perturbed compound Poisson risk process with investment and debit interest. *Methodol. Comput. Appl. Probab.* **2010**, *12*, 391–413. [[CrossRef](#)]
47. Yin C.C.; Yuen, K.C. Optimality of the threshold dividend strategy for the compound Poisson model. *Stat. Probab. Lett.* **2011**, *81*, 1841–1846. [[CrossRef](#)]
48. Yin, C.C.; Wen, Y.Z. Optimal dividend problem with a terminal value for spectrally positive Levy processes. *Insur. Math. Econom.* **2013**, *53*, 769–773. [[CrossRef](#)]
49. Yin, C.C.; Wen, Y.Z. Exit problems for jump processes with applications to dividend problems. *J. Comput. Appl. Math.* **2013**, *245*, 30–52. [[CrossRef](#)]
50. 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](#)]
51. Ma, J.X.; Ding, F. Filtering-based multistage recursive identification algorithm for an input nonlinear output-error autoregressive system by using the key term separation technique. *Circuits Syst. Signal Process.* **2017**, *36*, 577–599. [[CrossRef](#)]
52. Ma, P.; Ding, F. New gradient based identification methods for multivariate pseudo-linear systems using the multi-innovation and the data filtering. *J. Frankl. Inst.* **2017**, *354*, 1568–1583 [[CrossRef](#)]
53. Wan, X.K.; Li, Y.; Xia, C.; Wu, M.H.; Liang, J.; Wang, N. A T-wave alternans assessment method based on least squares curve fitting technique. *Measurement* **2016**, *86*, 93–100. [[CrossRef](#)]
54. Zhao, N. Joint Optimization of cooperative spectrum sensing and resource allocation in multi-channel cognitive radio sensor networks. *Circuits Syst. Signal Process.* **2016**, *35*, 2563–2583. [[CrossRef](#)]
55. Zhao, X.L.; Liu, F.; Fu, B.; Na, F. Reliability analysis of hybrid multi-carrier energy systems based on entropy-based Markov model. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2016**, *230*, 561–569. [[CrossRef](#)]
56. Zhao, N.; Liang, Y.; Pei, Y. Dynamic contract incentive mechanism for cooperative wireless networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10970–10982. [[CrossRef](#)]
57. Ding F., Liu X., Ma X. Kalman state filtering based least squares iterative parameter estimation for observer canonical state space systems using decomposition. *J. Comput. Appl. Math.* **2016**, *301*, 135–143. [[CrossRef](#)]

58. Yin, C.C.; Wen, Y.Z. An extension of Paulsen-Gjessing's risk model with stochastic return on investments. *Insur. Math. Econom.* **2013**, *52*, 469–476
59. Yin, C.C.; Wen, Y.Z.; Zhao, Y.X. On the optimal dividend problem for a spectrally positive levy process. *Astin Bull.* **2014**, *44*, 635–651. [[CrossRef](#)]
60. Yin, C.C.; Yuen, K.C. Exact joint laws associated with spectrally negative Levy processes and applications to insurance risk theory. *Front. Math. China* **2014**, *9*, 1453–1471.
61. Yin, C.C.; Yuen, K.C. Optimal dividend problems for a jump-diffusion model with capital injections and proportional transaction costs. *J. Ind. Manag. Optim.* **2015**, *11*, 1247–1262.
62. Gong, P.C.; Wang, W.Q.; Li, F.C.; Cheung, H. Sparsity-aware transmit beamspace design for FDA-MIMO radar. *Signal Process.* **2018**, *144*, 99–103. [[CrossRef](#)]
63. Zhao, X.L.; Lin, Z.Y.; Fu, B.; He, L.; Na, F. Research on automatic generation control with wind power participation based on predictive optimal 2-degree-of-freedom PID strategy for multi-area interconnected power system. *Energies* **2018**, *11*, 3325. [[CrossRef](#)]
64. Zhao, X.L.; Lin, Z.Y.; Fu, B.; He, L.; Li, C.S. Research on the predictive optimal PID plus second order derivative method for AGC of power system with high penetration of photovoltaic and wind power. *J. Electr. Eng. Technol.* **2019**, *14*, 1075–1086. [[CrossRef](#)]
65. Shi, W.X.; Liu, N.; Zhou, Y.M.; Cao, X.A. Effects of postannealing on the characteristics and reliability of polyfluorene organic light-emitting diodes. *IEEE Trans. Electron Devices* **2019**, *66*, 1057–1062. [[CrossRef](#)]
66. Wu, T.Z.; Shi, X.; Liao, L.; Zhou, C.J.; Zhou, H.; Su, Y.H. A capacity configuration control strategy to alleviate power fluctuation of hybrid energy storage system based on improved particle swarm optimization. *Energies* **2019**, *12*, 642. [[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/>).