Original Paper

Least square algorithm based on bias compensated principle for parameter estimation of canonical state space model

Longlong Liu[1], Zhen Long[1], Ahmad Taher Azar[2,3], Quanmin Zhu[4], Ibraheem Kasim Ibraheem[5] and Amjad J Humaidi[6]

Abstract

Due to the existence of system noise and unknown state variables, it is difficult to realize unbiased estimation with minimum variance for the parameter estimation of canonical state space model. This paper presents a new least squares estimator based on bias compensation principle to solve this problem, transforms canonical state space into the form suitable for the least square algorithm, introduces an augmented parameter vector and an auxiliary variable, derives parameter estimation formula based on noise compensation, realizes the unbiased estimation, and gives the specific algorithm. A simulation example is provided to verify the effectiveness of the estimator.

Introduction

State space model is an effective tool to describe systems. Due to its simplicity of equations and ease of understanding, state space model has been widely used in modeling and forecasting in economic, control, biological, and other fields. However, the state space model contains both unknown state vectors and unknown parameter vectors, and the nonlinear product relationship between parameter vectors and state vectors, which makes the identification of the model very complicated.[1] Therefore, the parameter estimation of the state space model has always been a research hotspot in control field. At present, the main methods to estimate the parameters of state space are subspace method,[2] Kalman filter method,[3,4] gradient search method,[5,6] least squares method,[7,8] etc.

Literature review

Least squares (LS) estimator is a mathematical optimization technique. The minimum variance of least squares estimator explained by Gauss Markov theorem shows that the effect of ordinary LS estimator is the best one compared with any other linear unbiased estimators. Therefore, the LS method has been widely used. However, the LS method can only use part of observed data, so it depends on the distribution of noise and the statistical properties of missing data, and it is difficult to realize unbiased estimation of parameters. An LS

algorithm based on bias compensated principle (BCP) was proposed in this paper[9] in view of the limitations of LS algorithm, which achieved the goal that online estimation can be realized when the order and variance of noise are unknown, and this method was widely used in parameter estimation of control systems.[10–12] Kenji Ikeda et al. analyzed the asymptotic bias of recursive least squares (RLS) estimation under closed-loop conditions, and presented compensation method under the assumption that the noise is white noise.[13] Wang et al. proposed a parameter and state estimation algorithm based on bias compensation for

[1]School of Mathematical Sciences, Ocean University of China, Qingdao, China
[2]College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia
[3]Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt
[4]Department of Engineering Design and Mathematics, University of the West of England, Bristol, UK
[5]Department of Computer Techniques Engineering, Dijlah University College, Baghdad, Iraq
[6]Control and Systems Engineering Department, University of Technology, Baghdad, Iraq

Corresponding author:
Amjad J. Humaidi, Control and Systems Engineering Department, University of Technology, Al-sina street, Baghdad 19006, Iraq.
Email: amjad.j.humaidi@uotechnology.edu.iq

observable canonical state space systems with colored noise, which can generate unbiased parameter estimation by the bias correction term.[14] For Hammerstein autoregressive moving average (ARMAX) systems with autoregressive variables, Zhang et al. proposed an RLS identification method based on the BCP.[15] Wu et al. combined bias compensation technology with LS estimation algorithm with forgetting factor to estimate parameters of output error model with moving average noise,[16] then, proposed an LS algorithm based on BCP for parameter estimation of MISO system under white noise.[17] Mu et al. proposed a globally convergent identification algorithm, modified BCP, and gave a consistent estimation of noise variance rather than introducing auxiliary vectors.[18] Jia et al. presented a unified framework based on the BCP, which introduces a nonsingular matrix and a noise-independent auxiliary vector to identify the system affected by correlated noise. Due to the rich possibility of selecting matrix and auxiliary vector, the framework has great flexibility.[19]

In the numerous literature studies on system parameter identification, there are not many cases of parameter identification for canonical state space model. By using hierarchical identification principle, Zhuang et al. proposed a new parameter and state estimation algorithm for canonical state space model by replacing noise term with estimation residual in an information vector and achieved good results.[20] However, even if the estimated residuals take the place of the noise item, the other items of algorithm in Ref. [20] still require system noise to be known. In actual systems, the noise is usually random noise, which is very difficult to be observed. When the noise is unknown, it is difficult to estimate parameters unbiased. In this paper, LS based on BCP is used to estimate parameters of canonical state space model, so that model transformation, bias analysis, and bias compensation are carried out to obtain unbiased estimation of system parameters.

Least squares algorithm based on bias compensated principle

Parameter estimation of canonical state space model based on least squares

By LS method, unknown data can be easily obtained, and square sum of errors between obtained data and actual data is the minimum (the least variance). In this section, LS method is applied to estimate parameters of canonical state space model, and state space model in the form of equation (1) is transformed into vector functions suitable for LS method.

Consider this state space equation (1)

$$\begin{cases} x(t+1) = Ax(t) + bu(t) \\ y(t) = cx(t) + v(t) \end{cases} \quad (1)$$

where, $A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \in \mathbb{R}^{n \times n}$ is

companion matrix $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n$, $\boldsymbol{c} = \begin{bmatrix} 1 & 0 & 0 & \dots 0 \end{bmatrix} \in$

$\mathbb{R}^{1 \times n}$ and $y(t) \in \mathbb{R}$ are input and output of the system, respectively, $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the unobservable state vector, and $v(t) \in \mathbb{R}$ is an uncorrelated random white noise with zero mean. Assume that the order n is known, $u(t) = 0, y(t) = 0$ and $v(t) = 0$ when $t \leq 0$. $\boldsymbol{A}, \boldsymbol{b}$ is the system parameter matrix and parameter vector to be identified, respectively.

Since $\boldsymbol{A}$ is a companion matrix and $\boldsymbol{c}$ has a standard structure, that is, the first element of $\boldsymbol{c}$ is 1 and the other elements are zeros, it can be seen that the model is observable canonical state space model, that is, parameters of observable canonical model are estimated.

Transform the form of equation (1). The equation of vector $\boldsymbol{Y}(t+n)$ can be obtained by deducing $y(t) = cx(t) + v(t)$ in equations (1) [20]:

$$\begin{aligned} y(t+1) &= cx(t+1) + v(t+1) \\ &= c(\boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}u(t)) + v(t+1) \\ &= cAx(t) + cbu(t) + v(t+1), \\ y(t+2) &= cAx(t+1) + cbu(t+1) + v(t+2) \\ &= cA[\boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}u(t)] + cbu(t+1) + v(t+2) \\ &= cA^2x(t) + cAbu(t) + cbu(t+1) + v(t+2), \\ &\qquad \cdots \\ y(t+n-1) &= cA^{n-1}x(t) + cA^{n-2}bu(t) + cA^{n-3}bu(t+1) \\ &\qquad + \cdots + cbu(t+n-2) + v(t+n-1), \end{aligned}$$

We have

$$\begin{cases} y(t) = cx(t) + v(t) \\ y(t+1) = cAx(t) + cbu(t) + v(t+1) \\ y(t+2) = cA^2x(t) + cAbu(t) + cbu(t+1) + v(t+2) \\ \qquad \cdots \\ y(t+n-1) = cA^{n-1}x(t) + cA^{n-2}bu(t) \\ \qquad\qquad + cA^{n-3}bu(t+1) + \dots \\ \qquad\qquad + cbu(t+n-2) + v(t+n-1) \\ y(t+n) = cA^n x(t) + cA^{n-1}bu(t) + cA^{n-2}bu(t+1) + \dots \\ \qquad\qquad + cbu(t+n-1) + v(t+n) \end{cases} \quad (2)$$

Define

$$\boldsymbol{Y}(t+n) = \begin{bmatrix} y(t) \\ y(t+1) \\ \vdots \\ y(t+n-1) \end{bmatrix} \in \mathbb{R}^n \quad (3)$$

$$\boldsymbol{U}(t+n) = \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+n-1) \end{bmatrix} \in \mathbb{R}^n \quad (4)$$

$$V(t+n) = \begin{bmatrix} v(t) \\ v(t+1) \\ \vdots \\ v(t+n-1) \end{bmatrix} \in \mathbb{R}^n \qquad (5)$$

$$M = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ cb & 0 & \cdots & 0 & 0 \\ cAb & cb & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ cA^{n-2}b & cA^{n-3}b & \cdots & cb & 0 \end{bmatrix} \in \mathbb{R}^{n \times n} \qquad (6)$$

$$Q_0 = \begin{bmatrix} c \\ cA \\ \vdots \\ cA^{n-1} \end{bmatrix} \in \mathbb{R}^{n \times n} \qquad (7)$$

Due to the special structures of c and A, we have $Q_0 = I_n$, which is an unit matrix. The first n equations of equation (2) can be rewritten as (8)

$$\begin{aligned} Y(t+n) &= Q_0 x(t) + MU(t+n) + V(t+n) \\ &= x(t) + MU(t+n) + V(t+n) \end{aligned} \qquad (8)$$

Transform (8) to get (9)

$$x(t) = Y(t+n) - MU(t+n) - V(t+n) \qquad (9)$$

Substituting (9) into equation (2) gives (10)

$$\begin{aligned} &y(t+n) \\ &= cA^n x(t) + cA^{n-1}bu(t) + cA^{n-2}bu(t+1) + \cdots \\ &\quad + cbu(t+n-1) + v(t+n) \\ &= cA^n[Y(t+n) - MU(t+n) - V(t+n)] + cA^{n-1}bu(t) \\ &\quad + cA^{n-2}bu(t+1) + \cdots + cbu(t+n-1) + v(t+n) \\ &= cA^n[Y(t+n) - V(t+n)] - cA^n MU(t+n) + cA^{n-1}b \\ &\quad u(t) + cA^{n-2}bu(t+1) + \cdots + cbu(t+n-1) + v(t+n) \\ &= cA^n[Y(t+n) - V(t+n)] - cA^n MU(t+n) \\ &\quad + [cA^{n-1}b, cA^{n-2}b, \cdots, cb] \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+n-1) \end{bmatrix} + v(t+n) \\ &= [Y^T(t+n) - V^T(t+n)](cA^n)^T - U^T(t+n)(cA^n M)^T \\ &\quad + [cA^{n-1}b, cA^{n-2}b, \cdots, cb] \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+n-1) \end{bmatrix} + v(t+n) \\ &= [Y^T(t+n) - V^T(t+n)](cA^n)^T - U^T(t+n)(cA^n M)^T \\ &\quad + U^T(t+n) \begin{bmatrix} cA^{n-1}b \\ \vdots \\ cb \end{bmatrix} + v(t+n) \\ &= [Y^T(t+n), U^T(t+n)] \begin{bmatrix} (cA^n)^T \\ (-cA^n M + [cA^{n-1}b, \cdots, cb])^T \end{bmatrix} \\ &\quad - V^T(t+n)(cA^n)^T + v(t+n). \end{aligned} \qquad (10)$$

Define

$$\phi(t+n) = \begin{bmatrix} Y(t+n) \\ U(t+n) \end{bmatrix} \in \mathbb{R}^{2n \times 1} \qquad (11)$$

$$\theta_1 = [cA^n]^T \qquad (12)$$

$$\theta_2 = \left[-cA^n M + [cA^{n-1}b, \cdots, cb]\right]^T \qquad (13)$$

$$\tilde{v}(t+n) = v(t+n) - V^T(t+n)(cA^n)^T \qquad (14)$$

Obtain

$$\phi^T(t+n) = \left[Y^T(t+n), U^T(t+n) \right] \qquad (15)$$

$$\tilde{V}(t+n) = \begin{bmatrix} \tilde{v}(t) \\ \tilde{v}(t+1) \\ \vdots \\ \tilde{v}(t+n-1) \end{bmatrix} \in \mathbb{R}^n \qquad (16)$$

So, equation (10) can be written as equation (17)

$$\begin{aligned} y(t+n) &= \phi^T(t+n) \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \tilde{v}(t+n) \\ &= \phi^T(t+n)\theta + \tilde{v}(t+n) \end{aligned} \qquad (17)$$

Replacing t in equation (17) with $t - n$ yields equation (18)

$$y(t) = \phi^T(t)\theta + \tilde{v}(t) \qquad (18)$$

Define

$$\Phi(t+n) = \begin{bmatrix} \phi^T(t) \\ \cdots \\ \phi^T(t+n-1) \end{bmatrix} \qquad (19)$$

So, the vector function form suitable for LS method is obtained

$$Y(t+n) = \Phi(t+n)\theta + \tilde{V}(t+n) \qquad (20)$$

Replacing t in equation (20) with $t - n$ yields equation (21)

$$Y(t) = \Phi(t)\theta + \tilde{V}(t) \qquad (21)$$

According to the principle of LS, the loss function can be defined as (22)

$$J(\theta) = \sum_{i=1}^{t} \tilde{v}^2(i) = \sum_{i-1}^{t} \left[y(i) - \phi^T(i)\theta\right]^2 \qquad (22)$$

LS estimation can be obtained by solving the minimization of loss function

$$\widehat{\theta}_{LS} = \left(\Phi^T(t)\Phi(t)\right)^{-1} \Phi^T(t) Y(t) \qquad (23)$$

If the above equations are used to estimate $\widehat{\theta}_{LS}$ directly, and then A and b are deduced, parameter estimation of equation (1) can be achieved. However, it can be seen from equation (14) that noise $\tilde{v}(t+n)$ may no longer be random white noise with zero mean, so that estimation of parameters by LS method may not be the least variance, that is, unbiased estimation of parameters may not be possible. Therefore, bias

compensation method will be used to eliminate the bias; the above LS estimator will be modified in order to achieve unbiased estimation.

Least squares estimator based on bias compensated principle

In this section, equation (21) is continuously to be deduced by LS gives (24)

$$
\begin{aligned}
\widehat{\theta}_{LS}(n) &= \left(\frac{1}{n}\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\Phi}(t)\right)^{-1}\frac{1}{n}\boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{Y}(t) \\
&= \left(\frac{1}{n}(\boldsymbol{\phi}(t-n)\cdots\boldsymbol{\phi}(t-1))\begin{pmatrix}\boldsymbol{\phi}^{T}(t-n)\\ \vdots \\ \boldsymbol{\phi}^{T}(t-1)\end{pmatrix}\right)^{-1} \\
&\quad\left(\frac{1}{n}(\boldsymbol{\phi}(t-n)\cdots\boldsymbol{\phi}(t-1))\begin{pmatrix}y(t-n)\\ \vdots \\ y(t-1)\end{pmatrix}\right) \\
&= \left(\frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1}\left(\frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)y(i))\right) \\
&= \widehat{\boldsymbol{R}}_{\phi\phi}^{-1}(n)\widehat{\boldsymbol{R}}_{\phi y}(n).
\end{aligned}
\tag{24}
$$

where $\quad\widehat{\boldsymbol{R}}_{\phi\phi}^{-1}(n) = \left(\frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1}$, $\quad\widehat{\boldsymbol{R}}_{\phi y}(n) = \left(\frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)y(i))\right)$.

Substituting $y(t) = \boldsymbol{\phi}^{T}(i)\theta + \tilde{v}(t)$ into (24) gives

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_{LS}(n) &= \left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1} \\
&\quad \left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}\left(\boldsymbol{\phi}(i)\left(\boldsymbol{\phi}^{T}(i)\theta + \tilde{v}(t)\right)\right)\right) \\
&= \left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1} \\
&\quad \left(\theta\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i)) + \left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}\left(\boldsymbol{\phi}(i)\tilde{v}(t)\right)\right) \\
&= \theta\left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1}\left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right) \\
&\quad + \left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1}\left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}\left(\boldsymbol{\phi}(i)\tilde{v}(t)\right)\right) \\
&= \theta + \left(\left(\frac{1}{n}\right)\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i))\right)^{-1}\left(\frac{1}{n}\sum_{i=t-n}^{t-1}\left(\boldsymbol{\phi}(i)\tilde{v}(t)\right)\right) \\
&= \theta + \widehat{\boldsymbol{R}}_{\phi\phi}^{-1}(n)\widehat{\boldsymbol{r}}_{\phi v}(n).
\end{aligned}
\tag{25}
$$

Taking probability limit of (25) yields

$$
\begin{aligned}
plim_{n\to\infty}\widehat{\boldsymbol{\theta}}_{LS}(n) &= \boldsymbol{\theta} + plim_{n\to\infty}\widehat{\boldsymbol{R}}_{\phi\phi}^{-1}(n)\widehat{\boldsymbol{r}}_{\phi v}(n) \\
&= \boldsymbol{\theta} + \boldsymbol{R}_{\phi\phi}^{-1}\boldsymbol{r}_{\phi v}
\end{aligned}
\tag{26}
$$

where

$$
\boldsymbol{R}_{\phi\phi} = plim_{n\to\infty}\widehat{\boldsymbol{R}}_{\phi\phi}(n) = plim_{n\to\infty}\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\boldsymbol{\phi}^{T}(i)) \tag{27}
$$

$$
\boldsymbol{r}_{\phi v} = plim_{n\to\infty}\widehat{\boldsymbol{r}}_{\phi v}(n) = plim_{n\to\infty}\sum_{i=t-n}^{t-1}(\boldsymbol{\phi}(i)\tilde{v}(i)) \tag{28}
$$

It can be seen from equation (26) that LS estimator is biased, and bias term is $\boldsymbol{R}_{\phi\phi}^{-1}\boldsymbol{r}_{\phi v}$ caused by noise. In order to eliminate the influence of bias term, equation (29) is obtained

$$
\widehat{\boldsymbol{\theta}}(n) = \widehat{\boldsymbol{\theta}}_{LS}(n) - \widehat{\boldsymbol{R}}_{\phi\phi}^{-1}(n)\widehat{\boldsymbol{r}}_{\phi v}(n) \tag{29}
$$

In order to obtain an unbiased estimate of $\widehat{\boldsymbol{\theta}}$, a constant is needed to replace $\widehat{\boldsymbol{r}}_{\phi v}(n)$ to compensate for the influence of bias. Assuming that system operates in an open-loop mode, it means that inputs $u(i)$ and $v(i)$ are not statistically relevant, then equation (28) can be rewritten as equation (30)

$$
\begin{aligned}
\boldsymbol{r}_{\phi v} &= plim_{n\to\infty}\frac{1}{n}\sum_{i=t-n}^{t-1}\begin{pmatrix}\boldsymbol{Y}(i)\\ \boldsymbol{U}(i)\end{pmatrix}\tilde{v}(i) \\
&= plim_{n\to\infty}\frac{1}{n}\sum_{i=t-n}^{t-1}\begin{pmatrix}\boldsymbol{Y}(i)\tilde{v}(i)\\ \boldsymbol{U}(i)\tilde{v}(i)\end{pmatrix} \\
&= \begin{bmatrix}\boldsymbol{I}_{n}\\ \boldsymbol{0}_{n}\end{bmatrix}plim_{n\to\infty}\frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{Y}(i)\tilde{v}(i)) = \begin{bmatrix}\boldsymbol{I}_{n}\\ \boldsymbol{0}_{n}\end{bmatrix}\boldsymbol{r}_{yv}
\end{aligned}
\tag{30}
$$

where, $\widehat{\boldsymbol{r}}_{yv}(n) = \frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{Y}(i)\tilde{v}(i))$, $\boldsymbol{r}_{yv} = plim_{n\to\infty}\frac{1}{n}\sum_{i=t-n}^{t-1}(\boldsymbol{Y}(i)\tilde{v}(i))$.

Equation (29) can be written as equation (31)

$$
\widehat{\boldsymbol{\theta}}(n) = \widehat{\boldsymbol{\theta}}_{LS}(n) - \widehat{\boldsymbol{R}}_{\phi\phi}^{-1}(n)\begin{bmatrix}\boldsymbol{I}_{n}\\ \boldsymbol{0}_{n}\end{bmatrix}\widehat{\boldsymbol{r}}_{yv}(n) \tag{31}
$$

It can be seen from (31) that the key is to find $\widehat{\boldsymbol{r}}_{yv}(n)$, so augmented regression vector $\boldsymbol{\psi}(i)$ and augmented parameter vector $\overline{\boldsymbol{\theta}}$ are introduced to satisfy $\boldsymbol{\psi}^{T}(i)\overline{\boldsymbol{\theta}} = \widehat{\boldsymbol{\phi}}^{T}(i)\theta$, which is commonly used in BCP. It is worth noting that no matter how different the structures of $\overline{\boldsymbol{\theta}}$ and $\boldsymbol{\psi}(i)$ are, the calculation of $\widehat{\boldsymbol{r}}_{yv}(n)$ is similar.[19] Therefore, the method described below is universal.

Introduce an auxiliary vector $\boldsymbol{\xi}(i) \in \mathbb{R}^{m}(m \geq n)$ which satisfies equation (32) and a nonsingular matrix $\mathcal{F} \in \mathbb{R}^{(2n+m)\times(2n+m)}$

$$
plim_{n\to\infty}\frac{1}{n}\sum_{i=t-n}^{t-1}\left(\boldsymbol{\xi}(i)\tilde{v}(i)\right) = 0 \tag{32}
$$

Construct an augmented regression vector $\boldsymbol{\psi}(i)$ and an augmented parameter vector $\overline{\boldsymbol{\theta}}$ by introducing auxiliary vector $\boldsymbol{\xi}(i)$ and nonsingular matrix $\mathcal{F}$, this step is the most important step, the selection of $\boldsymbol{\psi}(i)$ and $\overline{\boldsymbol{\theta}}$ makes calculation of $\widehat{\boldsymbol{r}}_{yv}$ becomes possible

$$\psi(i) = \mathcal{F}\begin{bmatrix} \boldsymbol{\phi}(i) \\ \boldsymbol{\xi}(i) \end{bmatrix} \in \mathbb{R}^{(2n+m)} \qquad (33)$$

$$\overline{\boldsymbol{\theta}} = \left(\mathcal{F}^{-1}\right)^T \begin{bmatrix} \boldsymbol{\theta} \\ 0 \end{bmatrix} \in \mathbb{R}^{(2n+m)} \qquad (34)$$

The $\overline{\boldsymbol{\theta}}$ and $\psi(i)$ satisfy equation (35)

$$\psi^T(i)\overline{\boldsymbol{\theta}} = \boldsymbol{\phi}^T(i)\boldsymbol{\theta} \qquad (35)$$

Rewrite $y(t) = \boldsymbol{\phi}^T(i)\boldsymbol{\theta} + \tilde{v}(t)$ as $y(t) = \psi^T(t)\overline{\boldsymbol{\theta}} + \tilde{v}(t)$. It is easy to yield equation (36) from equation (34)

$$\mathcal{F}^T\overline{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta} \\ 0 \end{bmatrix} \qquad (36)$$

Let

$$\mathcal{F} = [\mathcal{F}_1 \ \mathcal{F}_2 \ \boldsymbol{Z}] = [\mathcal{F}_0 \ \boldsymbol{Z}] \qquad (37)$$

where $\mathcal{F}_1, \mathcal{F}_2 \in \mathbb{R}^{(2n+m)\times n}, \mathcal{F}_0 \in \mathbb{R}^{(2n+m)\times 2n}, \boldsymbol{Z} \in \mathbb{R}^{(2n+m)\times m}$.
Equation (38) can be obtained

$$\boldsymbol{Z}^T\overline{\boldsymbol{\theta}} = 0 \qquad (38)$$

Similar to equations (24) and (26), equations (39) and (40) can be obtained in the same way

$$\widehat{\overline{\boldsymbol{\theta}}}_{LS}(n) = \widehat{\boldsymbol{R}}_{\psi\psi}^{-1}(n)\widehat{\boldsymbol{r}}_{\psi y}(n) \qquad (39)$$

$$plim_{n\to\infty}\widehat{\overline{\boldsymbol{\theta}}}_{LS}(n) = \overline{\boldsymbol{\theta}} + \left(plim_{n\to\infty}\widehat{\boldsymbol{R}}_{\psi\psi}^{-1}\right)\mathcal{F}_1\boldsymbol{r}_{yv} \qquad (40)$$

where

$$\widehat{\boldsymbol{R}}_{\psi\psi}(n) = \frac{1}{n}\sum_{i=t-n}^{t-1}\psi(i)\psi^T(i) \qquad (41)$$

$$\widehat{\boldsymbol{r}}_{\psi y}(n) = \frac{1}{n}\sum_{i=t-n}^{t-1}\psi(i)y(i) \qquad (42)$$

Pre-multiplying $\boldsymbol{Z}^T$ on both sides of equation (40) and using (38) yield

$$\boldsymbol{Z}^T plim_{n\to\infty}\widehat{\overline{\boldsymbol{\theta}}}_{LS}(n) = \boldsymbol{Z}^T\overline{\boldsymbol{\theta}} + \boldsymbol{Z}^T\left(plim_{n\to\infty}\widehat{\boldsymbol{R}}_{\psi\psi}^{-1}(n)\right)\mathcal{F}_1\boldsymbol{r}_{yv}$$
$$= \boldsymbol{Z}^T plim_{n\to\infty}\widehat{\boldsymbol{R}}_{\psi\psi}^{-1}(n)\mathcal{F}_1\boldsymbol{r}_{yv} \qquad (43)$$

Define

$$\widehat{\boldsymbol{D}} = \boldsymbol{Z}^T\widehat{\boldsymbol{R}}_{\psi\psi}^{-1}(n)\mathcal{F}_1 \in \mathbb{R}^{m\times n} \qquad (44)$$

Thus, the estimation of $\boldsymbol{r}_{yv}$ can be determined by the weighted LS as equation (45)

$$\widehat{\boldsymbol{r}}_{yv}(n) = \left(\widehat{\boldsymbol{D}}^T\boldsymbol{W}_{LS}\widehat{\boldsymbol{D}}\right)^{-1}\widehat{\boldsymbol{D}}^T\boldsymbol{W}_{LS}\boldsymbol{Z}^T\widehat{\overline{\boldsymbol{\theta}}}_{LS}(n) \qquad (45)$$

where $\boldsymbol{W}_{LS}$ is a positive definite weighting matrix. The choice of $\boldsymbol{W}_{LS}$ does not affect the consistency properties of $\widehat{\boldsymbol{r}}_{yv}(n)$, but may have a considerable impact on the statistical accuracy of $\widehat{\boldsymbol{r}}_{yv}(n)$.

One simple choice is $\boldsymbol{W}_{LS} = \boldsymbol{I}$.[19] In the case of $m = n$, if $\widehat{\boldsymbol{D}}$ is non-singular, $\widehat{\boldsymbol{r}}_{yv}(n)$ simply becomes as equation (46)

$$\widehat{\boldsymbol{r}}_{yv}(n) = \widehat{\boldsymbol{D}}^{-1}\boldsymbol{Z}^T\widehat{\overline{\boldsymbol{\theta}}}_{LS}(n) \qquad (46)$$

Similar to equations (24) and (26), equation (47) can be obtained from (40) in the same way

$$\widehat{\overline{\boldsymbol{\theta}}}(n) = \widehat{\overline{\boldsymbol{\theta}}}_{LS}(n) - \widehat{\boldsymbol{R}}_{\psi\psi}^{-1}(n)\mathcal{F}_1\widehat{\boldsymbol{r}}_{yv}(n) \qquad (47)$$

By substituting (45) into (47), consistent estimation of the augmented parameter vector $\overline{\boldsymbol{\theta}}$ can be obtained, finally, the global consistent estimation of original parameter $\boldsymbol{\theta}$ can be obtained by equation (36) and rewritten as equation (48)

$$\widehat{\boldsymbol{\theta}}(n) = \mathcal{F}_0^T\widehat{\overline{\boldsymbol{\theta}}}(n) \qquad (48)$$

Above is the unified BCP method. Through choosing appropriately $\boldsymbol{\xi}(i)$ and $\mathcal{F}$, various BCP-based methods can be obtained to achieve a global consistent estimation of parameters.

Parameter estimation of canonical state space model based on BCP-LS

For canonical state space model, this section uses the method introduced in 2.2 to derive a specific parameter estimation algorithm.
Choose

$$\mathcal{F} = \boldsymbol{I}_{3n}, \ \boldsymbol{\xi}(i) = \underline{\boldsymbol{U}}(i) \in \mathbb{R}^n \qquad (49)$$

where $\underline{\boldsymbol{U}}(i) = \begin{bmatrix} u(t-2n) \\ u(t-2n+1) \\ \vdots \\ u(t-n-1) \end{bmatrix}$, we obtain $\overline{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ 0 \end{bmatrix} \in \mathbb{R}^{3n}$,

$\psi(i) = \begin{bmatrix} \boldsymbol{Y}(i) \\ \boldsymbol{U}(i) \\ \underline{\boldsymbol{U}}(i) \end{bmatrix} \in \mathbb{R}^{3n}$.

Then, $\widehat{\boldsymbol{r}}_{yv}$ can be obtained according to equation (45), and the global unbiased estimate of the parameter $\boldsymbol{\theta}$ can be obtained by equations (47) and (48). Furthermore, parameters of the system can be estimated by input and output data. Since the relationship between parameter vector $\boldsymbol{\theta}$ and parameters $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}$ has been established in parameter estimation of canonical state space model based on least squares, the algorithm can be directly derived from algebraic operations.

Post-multiplying $\boldsymbol{Q}_0 = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{cA} \\ \vdots \\ \boldsymbol{cA}^{n-1} \end{bmatrix} = \boldsymbol{I}_n$ by $\boldsymbol{b}$ gives

$\begin{bmatrix} \boldsymbol{cb} \\ \boldsymbol{cAb} \\ \vdots \\ \boldsymbol{cA}^{n-1}\boldsymbol{b} \end{bmatrix} = \boldsymbol{b}$. Observing the structure of vector $\boldsymbol{b}$ to get equation (50)

$$\boldsymbol{cA}^{k-1}\boldsymbol{b} = b_k, \ k = 1, 2, \ldots, n \qquad (50)$$

The matrix $\boldsymbol{M}$ in parameter estimation of canonical state space model based on least squares can be simplified by (50) as

$$M = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ b_1 & 0 & \cdots & 0 & 0 \\ b_2 & b_1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ b_{n-1} & b_{n-2} & \cdots & b_1 & 0 \end{bmatrix} \quad (51)$$

Post-multiplying $\begin{bmatrix} c \\ cA \\ \vdots \\ cA^{n-1} \end{bmatrix} = I_n$ by A gives

$$\begin{bmatrix} cA \\ cA^2 \\ \vdots \\ cA^n \end{bmatrix} = A \quad (52)$$

Post-multiplying $cA^n = [-a_n, -a_{n-1}, \cdots, -a_1]$ by M gives

$$cA^n M = \begin{bmatrix} -(b_1 a_{n-1} + b_2 a_{n-2} + \cdots + b_{n-1} a_1) \\ -(b_1 a_{n-2} + b_2 a_{n-3} + \cdots + b_{n-2} a_1) \\ \vdots \\ -b_1 a_1 \\ 0 \end{bmatrix}^T \quad (53)$$

Equation (54) can be obtained according to equations (52) and (12)

$$\theta_1 = [cA^n]^T = \begin{bmatrix} -a_n \\ -a_{n-1} \\ \vdots \\ -a_1 \end{bmatrix} \quad (54)$$

Equation (55) can be obtained according to equations (53) and (13)

$$\theta_2 = \left[-cA^n M + \left[cA^{n-1} b, \cdots, cb \right] \right]^T$$

$$= \begin{bmatrix} b_1 a_{n-1} + b_2 a_{n-2} + \cdots + b_{n-1} a_1 + b_n \\ b_1 a_{n-2} + b_2 a_{n-3} + \cdots + b_{n-2} a_1 + b_{n-1} \\ \vdots \\ b_1 a_1 + b_2 \\ b_1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{n-1} & a_{n-2} & \cdots & a_1 & 1 \\ a_{n-2} & a_{n-3} & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \quad (55)$$

Since the estimated value $\widehat{\theta}(n)$ of $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$ has been obtained previously, the estimated value of a_i has been obtained,

that is, the estimated value $\widehat{A}$ of A can be obtained. Then equation (56) can be obtained due to the expression of θ_2

$$\widehat{b} = \begin{bmatrix} \widehat{b}_1 \\ \widehat{b}_2 \\ \vdots \\ \widehat{b}_{n-1} \\ \widehat{b}_n \end{bmatrix} = \begin{bmatrix} \widehat{a}_{n-1} & \widehat{a}_{n-2} & \cdots & \widehat{a}_1 & 1 \\ \widehat{a}_{n-2} & \widehat{a}_{n-3} & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{a}_1 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}^{-1} \widehat{\theta}_2 \quad (56)$$

A consistent estimated value $\widehat{b}$ of b can be obtained from (56). Now we have realized the global consistent unbiased estimation of parameters of canonical state space model (1).

The specific algorithm is summarized as follows:

1) Transform equation (1) into the form of $Y(t) = \Phi(t)\theta + \tilde{V}(t)$ according to equations (2)–(21);
2) collect the input/output data of $\{u(t), y(t)\}$;
3) introduce auxiliary vector ξ and matrix $\mathcal{F}$, then construct ψ and $\overline{\theta}$ according to equations (33) and (34);
4) calculate $\widehat{R}_{\psi\psi}$ by equation (41);
5) calculate $\widehat{r}_{\psi y}$ by equation (42);
6) calculate $\widehat{\overline{\theta}}_{LS}$ by equation (39);
7) calculate $\widehat{D}$ by equation (44);
8) define $W_{LS} = I$, calculate $\widehat{r}_{yv}$ by equation (45);
9) calculate $\widehat{\overline{\theta}}(n)$ by equation (47);
10) calculate $\widehat{\theta}(n)$ by equation (48), then the global consistent least squares estimation of θ is obtained;
11) calculate $\widehat{A}, \widehat{b}$ by equations (54) and (56), algorithm is end.

Simulation research

In this section, several examples of a canonical state space model are used to verify performance of presented estimator.

Example 1

For equation (57), $\{u(t)\}$ is taken as an independent random signal sequence with zero mean and unit variance, and $\{v(t)\}$ as a white noise sequence with zero mean and variances $\sigma_1^2 = 0.1^2$ and $\sigma_2^2 = 1.0^2$, respectively. Time is represented as t and parameter estimation error is calculated by $\varepsilon = \widehat{\theta} - \theta_2$, which is similar to the example in Ref. 20

$$\begin{cases} x(t+1) = \begin{bmatrix} 0 & 1 \\ -0.8 & 1.6 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u(t) \\ y(t) = [1 \quad 0] x(t) + v(t). \end{cases} \quad (57)$$

First, true value of θ is calculated:

Table 1. Estimated value and error of parameter θ.

	θ_{11}	θ_{12}	θ_{21}	θ_{22}	ε
True value	-0.80000	1.60000	-2.60000	1.00000	
Estimated value (σ_1)	-0.80039	1.59986	-2.60101	1.00049	0.0011966
Estimated value (σ_2)	-0.80079	1.60029	-2.60211	1.00149	0.0027167

Table 2. Estimated values and true values of parameters A and b.

	A	b
True value	$\begin{bmatrix} 0 & 1 \\ -0.80000 & 1.60000 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$
Estimated value (σ_1)	$\begin{bmatrix} 0 & 1 \\ -0.80039 & 1.59986 \end{bmatrix}$	$\begin{bmatrix} 1.00049 \\ -1.00037 \end{bmatrix}$
Estimated value (σ_2)	$\begin{bmatrix} 0 & 1 \\ -0.80079 & 1.60029 \end{bmatrix}$	$\begin{bmatrix} 1.00149 \\ -0.99944 \end{bmatrix}$
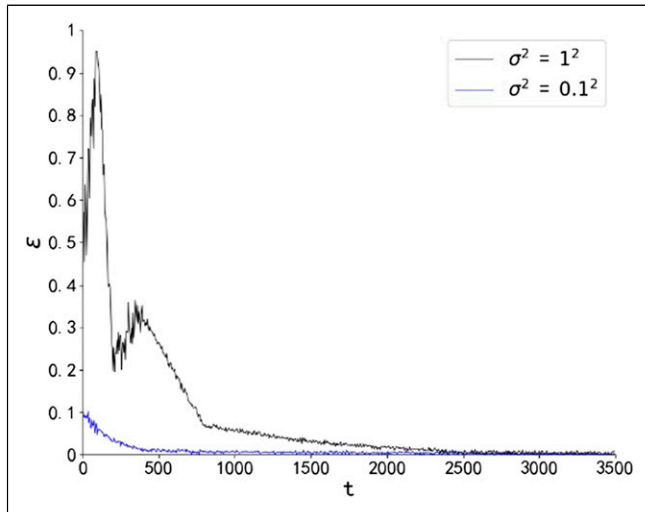
Figure 1. Parameter estimation error ε with time t.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -0.8 \\ 1.6 \\ -2.6 \\ 1 \end{bmatrix},$$

Let

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \theta_{21} \\ \theta_{22} \end{bmatrix} = \begin{bmatrix} -0.8 \\ 1.6 \\ -2.6 \\ 1 \end{bmatrix}.$$

Then algorithm in 2.3 is used to estimate parameters. When $t = 3500$, the estimated value and error of θ are shown in Table 1:

After obtaining estimated value of θ, the estimated values of parameters A and b are obtained, as shown in Table 2.

The curve of parameter θ estimation error ε with time t is shown in Figure 1.

Example 2

For equations (58), in the same way, $\{u(t)\}$ is taken as an independent random signal sequence with zero mean and unit variance, and $\{v(t)\}$ as a white noise sequence with zero mean and variances $\sigma_1^2 = 0.1^2$ and $\sigma_2^2 = 1.0^2$, respectively. Time is represented as t, and parameter estimation error is calculated by $\varepsilon = \widehat{\theta} - \theta_2$.

$$\begin{cases} x(t+1) = \begin{bmatrix} 0 & 1 \\ -1.2 & 0.6 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) + v(t) \end{cases} \quad (58)$$

First, true value of θ is calculated.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -1.2 \\ 0.6 \\ -1.6 \\ 1 \end{bmatrix},$$

Let

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \theta_{21} \\ \theta_{22} \end{bmatrix} = \begin{bmatrix} -1.2 \\ 0.6 \\ -1.6 \\ 1 \end{bmatrix}.$$

Then algorithm in 2.3 is used to estimate parameters. When $t = 3500$, the estimated value and error of θ are shown in Table 3:

After obtaining estimated value of θ, the estimated values of parameters A and b are obtained, as shown in Table 4.

The curve of parameter θ estimation error ε with time t is shown in Figure 2.

Example 3

For equation (59), in the same way, $\{u(t)\}$ is taken as an independent random signal sequence with zero mean and unit

Table 3. Estimated value and error of parameter θ.

	θ_{11}	θ_{12}	θ_{21}	θ_{22}	ε
True value	-1.20000	0.60000	-1.60000	1.00000	
Estimated value (σ_1)	-1.20041	0.59787	-1.60103	1.00037	0.0024296
Estimated value (σ_2)	-1.20088	0.60129	-1.60215	1.00109	0.0028721

Table 4. Estimated values and true values of parameters A and b.

	A	b
True value	$\begin{bmatrix} 0 & 1 \\ -1.20000 & 0.60000 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$
Estimated value (σ_1)	$\begin{bmatrix} 0 & 1 \\ -1.20041 & 0.59787 \end{bmatrix}$	$\begin{bmatrix} 1.00042 \\ -1.00031 \end{bmatrix}$
Estimated value (σ_2)	$\begin{bmatrix} 0 & 1 \\ -1.20088 & 0.60129 \end{bmatrix}$	$\begin{bmatrix} 1.00138 \\ -0.99896 \end{bmatrix}$
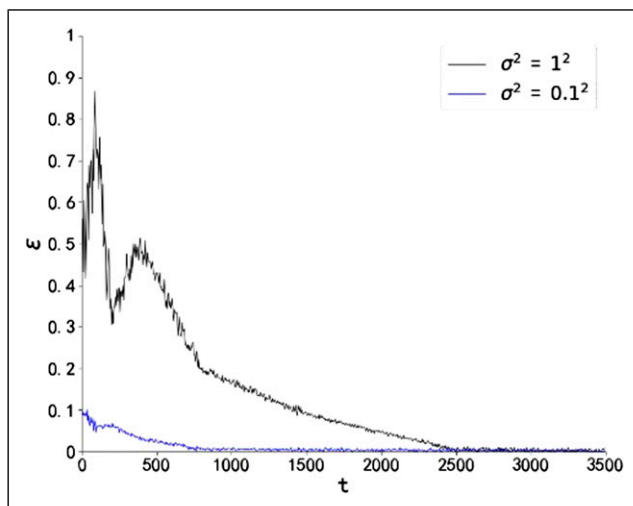
Figure 2. Parameter estimation error ε with time t.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0.2 \\ 0.8 \\ -2 \\ 0.2 \\ 1 \end{bmatrix},$$

Let

$$\begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \theta_{13} \\ \theta_{21} \\ \theta_{22} \\ \theta_{23} \end{bmatrix} = \begin{bmatrix} -1 \\ 0.2 \\ 0.8 \\ -2 \\ 0.2 \\ 1 \end{bmatrix}.$$

variance, and $\{v(t)\}$ as a white noise sequence with zero mean and variances $\sigma_1^2 = 0.1^2$ and $\sigma_2^2 = 1.0^2$, respectively. Time is represented as t, and parameter estimation error is calculated by $\varepsilon = \hat{\theta} - \theta_2$..

$$\begin{cases} x(t+1) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0.2 & 0.8 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x(t) + v(t) \end{cases} \quad (59)$$

First, true value of θ is calculated.

Then algorithm in 2.3 is used to estimate parameters. When $t = 3500$, the estimated value and error of θ are shown in Table 5.

After obtaining estimated value of θ, the estimated values of parameters A and b are obtained, as shown in Table 6.

The curve of parameter θ estimation error ε with time t is shown in Figure 3.

It can be seen from the three examples and Tables 1, 2, 3, 4, 5, and 6 that estimation error of parameters increases with the increase of noise, but it is still acceptable, so that the accuracy of estimator is relatively high. It is easy to see from Figures 1, 2, and 3 that the estimation error of parameter θ decreases with the increase of time, and the smaller the noise variance σ is, the faster the convergence speed of error curve is. The error ε will fluctuate greatly when noise

Table 5. Estimated value and error of parameter θ.

	θ_{11}	θ_{12}	θ_{13}	θ_{21}	θ_{22}	θ_{23}	ε
True value	−1.00000	0.20000	0.80000	−2.00000	0.20000	1.00000	
Estimated value (σ_1)	−1.00139	0.20076	0.80109	−2.00101	0.20049	1.00320	0.0038397
Estimated value (σ_2)	−1.00288	0.19930	0.80300	−2.00411	0.19149	1.00527	0.0116133

Table 6. Estimated values and true values of parameters A and b.

	A	b
True value	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0.2 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$
Estimated value (σ_1)	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1.00139 & 0.20076 & 0.80109 \end{bmatrix}$	$\begin{bmatrix} 1.00027 \\ 1.00019 \\ -1.00033 \end{bmatrix}$
Estimated value (σ_2)	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1.00288 & 0.19930 & 0.80300 \end{bmatrix}$	$\begin{bmatrix} 1.00052 \\ 1.00043 \\ -1.00069 \end{bmatrix}$
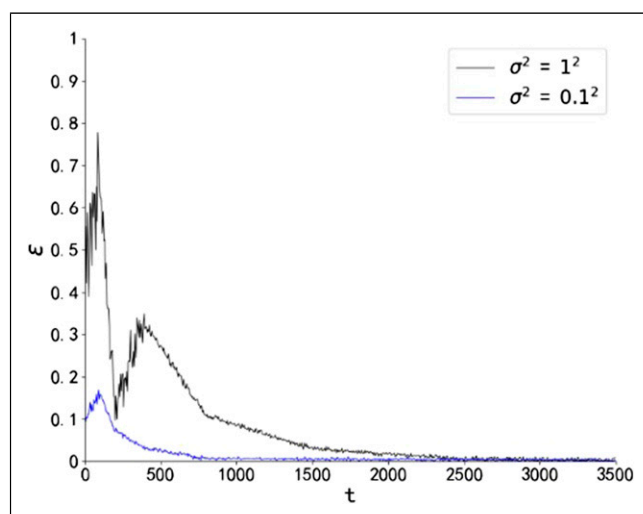
Figure 3. Parameter estimation error ε with time t.

variance $\sigma^2 = 1.0^2$ and $t < 500$, but with the increase of time t, it will show a downward trend, and finally achieve satisfactory results.

Conclusion

In this paper, an LS algorithm based on BCP is presented for parameter estimation of canonical state space model and a satisfactory result is achieved. The canonical state space model is transformed into vector function form suitable for LS algorithm, and then, augmented vector and auxiliary variable are introduced to realize unbiased estimation based on noise compensation; finally, the effectiveness of the estimator is verified by an example. This method of parameter estimation is universally useful, which can be extended to parameter estimation of other systems with noise, for example,

parameter estimation of closed-loop system can be realized by selecting different augmented vectors, this is also our future research work. Therefore, the method presented in our paper has widespread application and practical significance.

ORCID iDs

Ibraheem K. Ibraheem ⓘ https://orcid.org/0000-0001-7009-3634
Amjad J. Humaidi ⓘ https://orcid.org/0000-0002-9071-1329
Ahmad Taher Azar ⓘ https://orcid.org/0000-0002-7869-6373

References

1. Ma X and Ding F. Standard state space system identification method. *J Nanjing Univ Information Technology: Nat Sci Edition* 2014; 6: 481–504.
2. Juillet F, Schmid PJ and Huerre P. Control of amplifier flows using subspace identification techniques. *J Fluid Mech* 2013; 725: 522–565.
3. Ma X and Ding F. Gradient-based parameter identification algorithms for observer canonical state space systems using state estimates. *Circuits, Systems, Signal Process* 2015; 34(5): 1697–1709.

4. Varshney D, Bhushan M and Patwardhan SC. State and parameter estimation using extended kitanidis kalman filter. *J Process Control* 2019; 76: 98–111.

5. Zhang X, Ding F, Alsaadi FE, et al. Recursive parameter identification of the dynamical models for bilinear state space systems. *Nonlinear Dyn* 2017.

6. Xuehai W and Feng D. Recursive parameter and state estimation for an input nonlinear state space system using the hierarchical identification principle. *Signal Process* 2015; 117: 208–218.

7. Krzysztof O and Klaudia D. New parameter identification method for the fractional order, state space model of heat transfer process. In: *Conference on Automation*. Cham, Switzerland: Springer; 2018.

8. Wei Xing Zheng WX. On a least-squares-based algorithm for identification of stochastic linear systems. *IEEE Trans Signal Process* 1998; 46(6): 1631–1638.

9. Sagara S and Wada K. On-line modified least-squares parameter estimation of linear discrete dynamic systems. *Int J Control* 1977; 25(3): 329–343.

10. Stoica P and Söderström T. Bias correction in least-squares identification. *Int J Control* 1982; 35: 449–457.

11. Zhang Y, Lie TT and Soh CB. Consistent parameter estimation of systems disturbed by correlated noise. *IEE Proc - Control Theor Appl* 1997; 144: 40–44.

12. Stoica P, Simonyte V and Soderstrom T. Study of a bias-free least squares parameter estimator. *IEE Proc - Control Theor Appl* 1995; 142: 1–6.

13. Kenji I, Yoshio M and Takao S. Bias compensation of recursive least squares estimate in closed loop environment. *SICE J ControlMeasurement. and System Integration* 2008.

14. Wang X, Ding F, Liu Q, et al. The bias compensation based parameter and state estimation for observability canonical state-space models with colored noise. *Algorithms* 2018; 11(11).

15. Zhang B and Mao Z. Bias compensation principle based recursive least squares identification method for hammerstein nonlinear systems. *J Franklin Inst* 2016: 1340–1355.

16. Wu AG, Qian YY and Wu WJ. Bias compensation-based recursive least-squares estimation with forgetting factors for output error moving average systems. *Iet Signal Process* 2013; 8(5): 483–494.

17. Wu AG, Qi WN, Dong RQ Bias compensation based recursive least-squares identification algorithm for MISO system with input and output noises. In Proceedings of the 2017 IEEE 11th Asian Control Conference (ASCC), Gold Coast, QLD, 17-20 December 2017; pp. 812–816.

18. Mu B, Bai E-W, Zheng WX, et al. A globally consistent nonlinear least squares estimator for identification of nonlinear rational systems. *Automatica* 2017; 77: 322–335.

19. Jia L-J, Tao R, Kanae S, et al. A unified framework for bias compensation based methods in correlated noise case. *IEEE Trans Automatic Control* 2011; 56(3): 625–629.

20. Zhuang L, Pan F and Ding F. Parameter and state estimation algorithm for single-input single-output linear systems using the canonical state space models. *Appl Math Model* 2012; 36(8): 3454–3463.