Automatica 45 (2009) 1086–1091

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Subspace like identification incorporating prior information[☆]

Pavel Trnka*, Vladimír Havlena

Honeywell Prague Laboratory, Honeywell Intl., V Parku 2326/18, 148 00 Prague, Czech Republic
Department of Control Engineering, Czech Technical University in Prague, Karlovo náměstí 13, 121 35 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 2 March 2007
Received in revised form
23 October 2008
Accepted 11 December 2008
Available online 24 February 2009

Keywords:
System identification
Subspace methods
A priori information
Grey-box models
State space realization

ABSTRACT

The subspace identification methods have proved to be a powerful tool, which can further benefit from the prior information incorporation algorithm proposed in this note. In the industrial environment, there is often some knowledge about the identified system (known static gains, dominant time constants, low frequency character, etc.), which can be used to improve model quality and its compliance with first principles. The proposed algorithm has two stages. The first one is similar to the subspace methods as it uses their interpretation as an optimization problem of finding parameters of an optimal multi-step linear predictor for the experimental data. This problem is reformulated in the Bayesian framework allowing prior information incorporation in the form of the mean value and the covariance of the impulse response, which is shown to be useful for the incorporation of several prior information types. The second stage with state space model realization from the posterior impulse response estimate is different from the standard subspace methods as it is based on the structured weighted lower rank approximation, which is necessary to preserve the prior information incorporated in the first stage.

1. Introduction

The Subspace State Space System IDentification (4SID) has shown its suitability for industrial applications (Favoreel, De Moor, & Van Overschee, 2000), mainly due to its numerical robustness and the ability to identify MIMO (Multiple Inputs Multiple Outputs) systems with the same complexity as that for SISO (Single Input Single Output) systems without a need for extensive structural parameterization. However, it is quite usual that input/output data obtained from identification experiments in the industrial environment do not always have sufficient quality to give a good model by themselves. This may be caused by the fact that process excitation during identification experiments is limited by economical and safety reasons, which often results in data without proper excitation and with strong noise contamination. The black-box approach as in 4SID, relying only on experimental data, may provide biased models in such cases. In practical applications there is often strong prior information (PI) about the system, which can be exploited by the identification algorithm to improve the quality of the identified model. Such information can be approximate knowledge of time constants, known static gains, an input/output stability, step response smoothness, etc. It can be obtained from process operator experience, first principles model, by the analysis of process history data, etc. Previous efforts in 4SID incorporating PI were directed to specific cases (Chen, Van Huffel, Van den Boom, & Van den Bosch, 1997; Laudadio et al., 2004). A more general and systematic solution is still unavailable. In order to obtain a solution, the projections in 4SID have to be treated in a more general way. The proposed algorithm uses the interpretation of subspace identification as an optimization problem of finding a model as an optimal multi-step predictor (Van Overschee & De Moor, 1996) for the experimental data. Further, the problem is reformulated in the Bayesian framework allowing a combination of available PI with information from the experimental data. The PI is included by a convenient choice of the mean value and the covariance of a prior impulse response estimate. A straightforward realization of a state space model from a posterior impulse response estimate by Kung's algorithm (Kung, 1978) would lead to the loss of the included PI. To retain it, the realization is found using Hankel structured weighted lower rank approximation (SWLRA).

2. Notation and overview

In this paper a state space model of a stochastic system in the innovation form (Ljung, 1987) is considered

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Ke_k, \\ y_k &= Cx_k + Du_k + e_k, \end{aligned} \tag{1}$$

[☆] This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Brett Ninness, under the direction of Editor Torsten Söderström.

* Corresponding author at: Honeywell Prague Laboratory, Honeywell Intl., V Parku 2326/18, 148 00 Prague, Czech Republic. Tel.: +420 732 544725; fax: +420 234 625900.

E-mail addresses: pavel.trnka@honeywell.com (P. Trnka), vladimir.havlena@honeywell.com (V. Havlena).

0005-1098/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2008.12.005

where $u_k \in \mathbb{R}^m$ is the input, $x_k \in \mathbb{R}^n$ is the state, $y_k \in \mathbb{R}^l$ is the output, K is the steady state Kalman gain and $e_k \in \mathbb{R}^l$ is an unknown innovation sequence with $E[e_k] = 0$ and the covariance matrix $E\left[e_k e_k^{\mathrm{T}}\right] = R_e$, $E\left[e_p e_q^{\mathrm{T}}\right] = 0$ for $p \neq q$.

2.1. Signal related matrices

In 4SID algorithms, all signals are arranged in Hankel matrices. Assume known input/output data set $\left\{u_0^{N-1}, y_0^{N-1}\right\}$[1]arranged into block Hankel matrices with i and h block rows and j columns

$$\begin{pmatrix} U_p \\ U_f \end{pmatrix} \triangleq \left(\begin{array}{cccc} u_0 & u_1 & \dots & u_{j-1} \\ u_1 & u_2 & \dots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{i-1} & u_i & \dots & u_{i+j-2} \\ \hline u_i & u_{i+1} & \dots & u_{i+j-1} \\ u_{i+1} & u_{i+2} & \dots & u_{i+j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{i+h-1} & u_{i+h} & \dots & u_{N-1} \end{array}\right),$$

where $U_p \in \mathbb{R}^{im \times j}$ is the matrix denoted as the past inputs and $U_f \in \mathbb{R}^{hm \times j}$ is the matrix denoted as the future inputs. The coefficients i and h are selected larger than the upper bound of the expected system order n and $j = N - i - h + 1$. Identifiability results of Willems, Rapisarda, Markovsky, and De Moor (2008) require persistently exciting input of order $i + h + n$. For the outputs y_k and noises e_k similar Hankel matrices Y_p, Y_f and E_p, E_f can be constructed. A combination of U_p and Y_p is denoted as $W_p \triangleq (Y_p^{\mathrm{T}} \quad U_p^{\mathrm{T}})^{\mathrm{T}}$. The system state sequence is also arranged in a matrix form

$$X_p \triangleq \left(x_0\, x_1\, \dots\, x_{j-1}\right), \quad X_f \triangleq \left(x_i\, x_{i+1}\, \dots\, x_{i+j-1}\right).$$

2.2. Parameter related matrices

The extended observability matrix Γ_k is an extension of the observability matrix for a number of block rows higher than or equal to the system order n

$$\Gamma_k \triangleq \left(C^{\mathrm{T}} \quad (CA)^{\mathrm{T}} \quad \dots \quad (CA^{k-1})^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^{kl \times n}. \tag{2}$$

Similarly, *the extended controllability matrix* Δ_k

$$\Delta_k \triangleq \left(B \quad AB \quad \dots \quad A^{k-1}B\right) \in \mathbb{R}^{n \times km}. \tag{3}$$

The block Toeplitz matrix H_k^d composed from the deterministic impulse response sequence $\left\{g_0^{k-1}\right\}$

$$H_k^d \triangleq \begin{pmatrix} g_0 & 0 & \dots & 0 \\ g_1 & g_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{k-1} & g_{k-2} & \dots & g_0 \end{pmatrix} \in \mathbb{R}^{kl \times km}. \tag{4}$$

3. 4SID as multi-step optimal predictor

This section recalls that an oblique projection in 4SID can be viewed as a minimization result of the multi-step prediction error criterion (Van Overschee & De Moor, 1996). Assume a bank of h linear predictors for 1 to h step predictions based on history input/output window of length i and future input window of length h. The kth predictor with parameters $\theta^{(k)}$ is (Ljung & McKelvey, 1996)

$$\hat{y}(t+k|t) = \left(y_{t-i+1}^t \quad u_{t-i+1}^t \quad | \quad u_{t+1}^{t+h}\right)\theta^{(k)}, \quad k = 1, \dots, h.$$

The predictions on the whole data set can be written using Hankel notation as

$$\widehat{Y}_f = \left(L_w \quad H_h^d\right)\begin{pmatrix} W_p \\ U_f \end{pmatrix}, \tag{5}$$

where $\left(L_w \quad H_h^d \right) = \left(\theta^{(1)} \quad \dots \quad \theta^{(h)} \right)^{\mathrm{T}}$. Optimizing parameters for the minimal overall quadratic error

$$\min_{L_w, H_h^d} \left\| Y_f - \widehat{Y}_f \right\|_F^2 = \min_{L_w, H_h^d} \left\| Y_f - \left(L_w \quad H_h^d\right)\begin{pmatrix} W_p \\ U_f \end{pmatrix}\right\|_F^2, \tag{6}$$

where $\|\bullet\|_F$ is Frobenius norm and denoting $\mathcal{D} = \left(W_p^{\mathrm{T}} \quad U_f^{\mathrm{T}} \right)^{\mathrm{T}}$, the optimal L_w and H_h^d are

$$\left(L_w \quad H_h^d\right) = Y_f \mathcal{D}^{\mathrm{T}} \left(\mathcal{D}\mathcal{D}^{\mathrm{T}}\right)^{\dagger}. \tag{7}$$

Using the previous result, the estimated zero-input initial state response $L_w W_p$ can be computed as

$$L_w W_p = Y_f \mathcal{D}^{\mathrm{T}} \left[\left(\mathcal{D}\mathcal{D}^{\mathrm{T}}\right)^{\dagger}\right] \begin{pmatrix} \mathbf{I}_{r \times r} \\ \mathbf{0}_{hm \times r} \end{pmatrix} W_p, \quad r = \mathrm{i}(l + m),$$

which is an expression for the oblique projection

$$L_w W_p = Y_f \underset{U_f}{/} W_p, \tag{8}$$

showing the equivalence between the oblique projection in 4SID and estimation of parameters of an optimal multi-step predictor.

3.1. Enforcing causality and uniqueness of parameters

The oblique projection (7) does not ensure proper parameter structure, i.e. H_h^d should have zeros above the main diagonal and Toeplitz structure. This leads to predictor non-causality and over-parameterization. A solution was proposed in Peternell, Cherrer, and Deistler (1996). It uses a formula for the vectorization of a matrix product on (5)

$$\mathrm{vec}\left(\widehat{Y}_f\right) = \left((W_p^{\mathrm{T}} \quad U_f^{\mathrm{T}}) \otimes \mathbf{I}_{hl \times hl}\right) \mathrm{vec}\left(\left(L_w \quad H_h^d \right)\right)$$

and the fact that it is possible to find a certain matrix N such that

$$\mathrm{vec}\left(\left(L_w \quad H_h^d\right)\right) = N \begin{pmatrix} l_w \\ g \end{pmatrix}, \qquad \begin{array}{l} l_w = \mathrm{vec}\left(L_w\right), \\ g = \mathrm{vec}\left(\left(g_0 \quad \dots \quad g_{h-1}\right)\right). \end{array}$$

A set of equations equivalent to (5) with the enforced H_h^d structure and reduced number of parameters is then

$$\underbrace{\mathrm{vec}\left(\widehat{Y}_f\right)}_{y} = \underbrace{\left((W_p^{\mathrm{T}} \quad U_f^{\mathrm{T}}) \otimes \mathbf{I}_{hl \times hl}\right) N}_{Z} \underbrace{\begin{pmatrix} l_w \\ g \end{pmatrix}}_{\theta}. \tag{9}$$

[1] Notation $x_{t_1}^{t_2}$ stands for $(x_{t_1} \quad \dots \quad x_{t_2})$.