Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_107-1

Subspace Techniques in System Identification

Michel Verhaegen *
Delft Center for Systems and Control, Delft University, Delft, The Netherlands

Abstract

An overview is given of the class of subspace techniques (STs) for identifying linear, time-invariant state-space models from input-output data. STs do not require a parametrization of the system matrices and as a consequence do not suffer from problems related to local minima that often hamper successful application of parametric optimization- based identification methods.
The overview follows the historic line of development. It starts from Kronecker's result on the representation of an infinite power series by a rational function and then addresses, respectively, the deterministic realization problem, its stochastic variant, and finally the identification of a state-space model given in innovation form.
The overview summarizes the fundamental principles of the algorithms to solve the problems and summarizes the results about the statistical properties of the estimates as well as the practical issues like choice of weighting matrices and the selection of dimension parameters in using these STs in practice. The overview concludes with probing some future challenges and makes suggestions for further reading.

Keywords Hankel matrix • Extended observability matrix • SVD • Innovation model • State-space model

Introduction

Subspace techniques (STs) for system identification address the problem of identifying state-space models of MIMO dynamical systems. The roots of ST were laid by the German mathematician Leopold Kronecker (°1823–†1891). In Kronecker (1890) Kronecker established that a power series could be represented by a rational function when the *rank of the Hankel* operator with that power series as its symbol was *finite*. In the early 1990s of the twentieth century, new generalizations of the idea of Kronecker were presented for identifying linear, time-invariant (LTI) state-space models from input-output data or output data only. These new generalizations were formulated from different perspectives, namely, within the context of canonical variate analysis (Larimore 1990), within a linear algebra context (Van Overschee and De Moor 1994; Verhaegen 1994), and subspace splitting (Jansson and Wahlberg 1996). Despite their different origin, the close relationship between these methods was quickly established by a unifying theorem that interpreted these methods as a singular value decomposition (SVD) of a weighted matrix from which an estimate of the column space of the observability matrix or the row space of the state sequence of the given system or Kalman filter for observing the state of that system is derived (Van Overschee and De Moor 1995). This subspace calculation is the key feature that leads to the indication by ST for system identification or subspace identification methods (SIM).

*E-mail: m.verhaegen@tudelft.nl

DOI 10.1007/978-1-4471-5102-9_107-1

The STs are attractive *complementary* techniques to the maximum likelihood or prediction error framework. They do not require the user to specify a parametrization of the system matrices of the state-space model, and the user is not confronted with the problems due to possible local minima of a nonlinear parameter optimization method that is often necessary in estimating the parameters of a state-space model via, e.g., prediction error methods. Though the statistical properties such as consistency and efficiency have been investigated, such as in Bauer and Ljung (2002), the estimates obtained via ST are in general not optimal in the statistical minimum variance sense. However, practical evidence with the use of ST in a wide variety of problems has indicated that ST provides accurate estimates. As such they are often used as an initialization to the maximum likelihood or prediction error parametric identification methods.

In this chapter we make a distinction between *output only* or stochastic identification problems and *input–output* or combined deterministic-stochastic identification problems. The first occurs when identifying, e.g., the eigenmodes of a bridge from ambient acceleration responses of the bridge. The second occurs when, in addition to ambient excitations that cannot be directly measured, controlled excitations through actuators integrated in the system are used during the collection of the input–output data.

The outline of this chapter is as follows. In the next section, we formulate the LTI state-space model identification problems and outline the general strategy of ST. The presentation of ST is given according to the historical development of ST. It starts with a summary of the solution to the deterministic realization problem, which considers the noise-free "impulse" response of the system. Subsequently we present the stochastic realization problem which considers the output-only identification problem where the output is assumed to be a filtered zero-mean, white-noise sequence. The ST solution is discussed assuming samples of the covariance function of the output to be given. The deterministic-stochastic identification problem is considered in section "Combined Deterministic-Stochastic ST." In this section we first consider open-loop identification experiments. For this case, the basic linear regression problem is formulated that is at the heart of many ST. Second reference is made to a framework for analyzing and understanding the statistical properties of ST, the selection of the order, as well as to a number of open problems in the understanding of important choices the user has to made. Closed-loop identification experiments are considered in the third part of section "Combined Deterministic-Stochastic ST," while the fourth part makes a brief reference to ST papers that go beyond the LTI case.

Finally we provide a brief overview on future research directions and conclude with some recommended literature for further exploration.

The ST Identification Problems and Strategy

The LTI system to be analyzed in this chapter is given by the following state-space model:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k) \\ y(k) &= Cx(k) + Du(k) + e(k) \end{aligned} \tag{1}$$

with $u(k) \in \mathbb{R}^m$ the (measurable) input, $e(k)$ a zero-mean, white-noise sequence with $E[e(k)e(k)^T] = R$, $y(k) \in \mathbb{R}^\ell$ the (measurable) output, and $x(k) \in \mathbb{R}^n$ the state vector. This model is in the so-called innovation form since the sequence $e(k)$ is the innovation signal in a Kalman filtering context.

DOI 10.1007/978-1-4471-5102-9_107-1

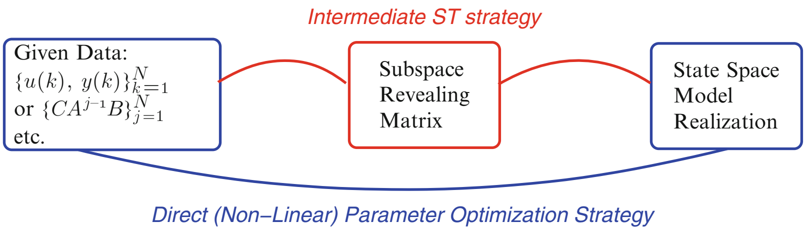

Fig. 1 Schematic representation of the intermediate step of ST to derive from the given data (input–output data $\{u(k), y(k)\}$, Markov parameters $\{CA^{j-1}B\}$, etc.) a subspace revealing matrix, from which the subspace of interest is computed via, e.g., singular value decomposition and that enables the computation of the state-space model realization by solving a (convex) linear least-squares problem. The commonly used approach to directly go from the given data to a state-space realization via in general nonlinear parameter optimization methods is indicated by the arrow directly connecting the given data box to the state-space realization box

The historical sequence of ST developments considers the following open-loop problem formulations. In the deterministic realization problem, the innovation sequence $e(k)$ is zero, and the input $u(k)$ is an impulse. The stochastic realization problem considers the case where the input $u(k)$ is zero and the given data is assumed to be samples of the covariance function of the output. The combined deterministic-stochastic identification problem considers the model (1) for generic input $u(k)$.

The general strategy of ST is to formulate an *intermediate step* in deriving the parameters of the system matrices of interest from the given data; see Fig. 1. This intermediate step makes the ST different from the parametric model identification framework that aims for a direct estimation of the parameters of the system matrices by (in general) nonlinear parameter optimization techniques. The intermediate step in ST aims to determine a matrix from the given data that *reveals* an (approximation of an) essential subspace of the unknown system. This essential subspace can be the extended observability matrix of (1) as given by the matrix $\mathcal{O}_s$:

This sais that you can use this method to reveal different kinds of subspaces

$$\mathcal{O}_s = \begin{bmatrix} C \\ CA \\ \cdots \\ CA^{s-1} \end{bmatrix} \quad \text{for} \quad s > n,$$

or the state sequence of a Kalman filter designed for (1). Essential for ST is that both the intermediate step to reveal the subspace of interest and the subsequent derivation of the system matrices from that subspace and the given data are done via convex optimization methods and/or linear algebra methods.

Realization Theory: The Progenitor of ST

The Deterministic Realization Problem

In the 1960s, the cited result of Kronecker inspired independently Ho and Kalman, Silverman and Youla, and Tissi to present an algorithm to construct a state-space model from a Hankel matrix of impulse response coefficients (Schutter 2000). This breakthrough gave rise to the field of *realization theory*. One key problem in realization theory that paved the way for subspace identification is the determination of a minimal realization from a finite number of samples of the impulse response of a deterministic system, assumed to have a minimal representation as in (1) for $e(k) \equiv 0$. The samples of the impulse response are called the *Markov parameters*. The minimal realization sought for is the LTI model with quadruple of system matrices $[A_T, B_T, C_T, D]$, with $A_T \in \mathbb{R}^{n\times n}$ and n minimal such that the pair (A_T, C_T) is observable, the pair (A_T, B_T) is controllable, and the transfer function $D + C_T(zI - A_T)^{-1}B_T$ equals $D + C(zI - A)^{-1}B$ with z the complex variable of the z-transform. When A is stable, the latter transfer function can be written into the matrix power series:

$$D + C(zI - A)^{-1}B = D + \sum_{j=1}^{\infty} CA^{j-1}Bz^{-j} \tag{2}$$

Following the cited result of Kronecker, the solution to the minimum realization problem is based on the construction of the (block-)Hankel matrix $H_{s,N}$ constructed from the Markov parameters $\{CA^{j-1}B\}_{j=1}^{N}$ as

$$H_{s,N} = \begin{bmatrix} CB & CAB & \cdots & CA^{N-s}B \\ \vdots & & \ddots & \vdots \\ CA^{s-1}B & CA^{s}B & \cdots & CA^{N-1}B \end{bmatrix} \tag{3}$$

For the deterministic realization problem, the *intermediate ST step* simply is the storage of the impulse response data into a Hankel matrix. The subsequent step is to derive from this matrix a subspace from which the system matrices can be either read-off or computed via linear least squares. How this is done is outlined next.

When the order n of the minimal realization is known and the Hankel matrix dimension parameters s, N are chosen such that

$$s > n \quad N \geq 2n - 1 \tag{4}$$

the Hankel matrix $H_{s,N}$ has **rank** n. A numerically reliable way to compute that rank is via the SVD of $H_{s,N}$. Under the assumption that the rank of $H_{s,N}$ is n, we can denote that SVD as $U_n \Sigma_n V_n^T$, with $\Sigma_n \in \mathbb{R}^{n\times n}$ positive definite and with the columns of the matrices U_n and V_n orthonormal. By the minimality of (1) (for $e(k) \equiv 0$), $H_{s,N}$ can be factored as $\mathcal{O}_s \left[B \ AB \ \cdots \ A^{N-s}B\right] = \mathcal{O}_s \mathcal{C}_{N-s+1}$ or as $\left(U_n \Sigma_n^{\frac{1}{2}}\right)\left(\Sigma_n^{\frac{1}{2}} V_n^T\right)$, and these factors are related as

$$U_n \Sigma_n^{\frac{1}{2}} = \mathcal{O}_s T^{-1} = \mathcal{O}_{s,T} \quad \Sigma_n^{\frac{1}{2}} V_n^T = T\mathcal{C}_{N-s+1} = \mathcal{C}_{N-s+1,T}$$

for $T \in \mathbb{R}^{n\times n}$ a nonsingular transformation. Therefore $\mathcal{O}_{s,T}$ resp. $\mathcal{C}_{N-s+1,T}$ act as the extended observability resp. controllability matrix of a similarly equivalent triplet of system matrices (A_T, B_T, C_T). This correspondence allows to read-off the system matrices C_T and B_T as the first ℓ rows of the matrix $\mathcal{O}_{s,T}$ and the first m columns of $\mathcal{C}_{N-s+1,T}$ resp. Further the *shift-invariance* property of the extended observability resp. controllability matrices allows to find the system matrix A_T of the minimal realization. For example, consider the extended observability matrix $\mathcal{O}_s$, then the shift-invariance property states that:

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_107-1

$$\mathcal{O}_{s,T}(1:(s-1)\ell,:)A_T = \mathcal{O}_{s,T}(\ell+1:s\ell,:) \tag{5}$$

where the notation $M(u:v,:)$ indicates the submatrix of M from rows u to rows v. The shift-invariance property delivers a set of linear equations from which the system matrix A_T can be computed via the solution of a **linear** least-squares problem when $s > n$.

Finding the dimension parameters s (and N) of the Hankel matrix $H_{s,N}$ is a nontrivial problem in general. When only the Markov parameters are given and the knowledge that they stem from a finite-order state-space model, a possible sequential strategy is to select s and N equal to the upperbounds in (4) for presumed orders n and $n+1$, respectively. When the rank of the Hankel matrices for these two selections of s (and N) is identical, the right dimensioning of the Hankel matrix $H_{s,N}$ is found. Otherwise the presumed order is increased by one.

The Stochastic Realization Problem

The output-only identification problem aims at determining a mathematical model from a measured multivariate time series $\{y(k)\}_{k=1}^{N}$ with $y(k) \in \mathbb{R}^{\ell}$. Such a model can be then used for predicting future values of the (output) data from past values.

In the vein of the revival of the work of Kronecker on realizing dynamical systems from its impulse response, Faure and a number of contemporaries like Akaike and Aoki made pioneering contributions to extend this methodology to stochastic processes (Van Overschee and De Moor 1993). These extensions are known as solutions to the stochastic realization problem.

This problem is formulated for $y(k)$ to be a Markovian stochastic process. Reusing the notation in (1) $y(k)$ is assumed to be generated by (1) with the input $u(k) \equiv 0$. The A matrix in (1) is again assumed to be stable. The given data in the early formulations of the stochastic realization problem was the samples of the covariance function

$$R_y(j) = E[y(k)y(k-j)^T]$$

These samples define the strictly positive real spectral density function of $y(k)$:

$$\Phi_y(z) = \sum_{j=-\infty}^{\infty} R_y(j)z^{-j} > 0 \tag{6}$$

Given the samples of the covariance function $R_y(j)$, the stochastic realization problem was to find an innovation model representation of the form

$$\begin{aligned} \hat{x}(k+1) &= A_T\hat{x}(k) + K_T e'(k) \\ \tilde{y}(k) &= C_T\hat{x}(k) + e'(k) \end{aligned} \tag{7}$$

with $e'(k)$ a zero-mean, white-noise input with covariance matrix R_e, the pair (A_T, C_T) observable, and A_T stable, such that the spectral density functions $\Phi_y(z)$ and $\Phi_{\tilde{y}}(z)$ are equal.

The partial similarity between this problem and the minimal realization problem becomes clear when expressing the covariance function samples $R_y(j)$ in terms of the system matrices in (1)–for $u(k) \equiv 0$ as

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_107-1

$$R_y(j) = CA^{j-1}G \quad \text{for} \quad j \neq 0 \tag{8}$$

with the matrices G and $R_y(0)$ derived from the following covariance expressions:

$$E[x(k)x(k)^T] = \Sigma_x : \Sigma_x = A\Sigma_x A^T + KRK^T \tag{9}$$

$$E[x(k+1)y(k)^T] = G : G = A\Sigma_x C^T + KR \tag{10}$$

$$E[y(k)y(k)^T] = R_y(0) : R_y(0) = C\Sigma_x C^T + R \tag{11}$$

Since the spectral density has a two-sided series expansion, there is a so-called forward stochastic realization problem (considering $R_y(j)$ for $j \geq 0$ only) and a backward version. Here we only treat the forward one. Drawing the parallel between the samples of the covariance function $R_y(j)$, as given in (6)–(8) and the Markov parameters in (2), we can use the deterministic tools from realization theory to find a minimal realization (A_T, C_T, G_T).

The *intermediate ST step* in the stochastic realization problem is the construction of a Hankel matrix similar to the matrix $H_{s,N}$ as in the deterministic realization problem but now from the samples of the covariance function $R_y(j)$ in (8).

With the triplet (A_T, C_T, G_T) determined, the innovation model (7) is classically completed via the solution of a Riccati equation in the unknown Σ_x. This Riccati equation results by noting that $R > 0$, and therefore, KRK^T can be written as $KR(R)^{-1}R^T K^T$. This reduces the expression for Σ_x in (9) with the help of (10) and (11) as

$$\Sigma_x = A\Sigma_x A^T + (G - A\Sigma_x C^T)(R_y(0) - C\Sigma_x C^T)^{-1}(G - A\Sigma_x C^T)^T \tag{12}$$

By replacing the triplet (A, C, G) with the found minimal realization (A_T, C_T, G_T) in this Riccati equation, its solution $\Sigma_{x,T}$ enables in the end to define the missing quantities as

$$\begin{aligned} R_e &= R_y(0) - C_T \Sigma_{x,T} C_T^T \\ K_T &= (G_T - A_T \Sigma_{x,T} C_T^T) R_e^{-1} \end{aligned} \tag{13}$$

By the positive realness of $\Phi_y(z)$ and the similar equivalence between the triplets (A_T, C_T, G_T) and (A, C, G), the solution $\Sigma_{x,T}$ is positive definite.

A persistent problem in solving the stochastic realization problem has existed for a long time when using approximate values of the samples $R_y(j)$. This problem is that the estimated power spectrum based on estimates of the triplet (A_T, C_T, G_T) is *no longer positive real*.

An approximate solution overcoming the problem of the loss of positive realness of the estimated power spectrum was provided in the vein of the ST developed in the early 1990s as discussed in the next section.

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_107-1

Combined Deterministic-Stochastic ST

Identification of LTI MIMO Systems in Open Loop

Since the golden 1960s and 1970s of the twentieth century, many attempts have been made to make the insights from deterministic and stochastic realization theory useful for system identification. To mention a few, there are attempts to use the solutions to the deterministic realization problem with measured or estimated impulse response data. One such method is known under the name of the eigensystem realization algorithm (ERA) (Juang and Pappa 1985) and has been used for modal analysis of flexible structures, like bridges, space structures, etc. Although these methods tend to work well in practice for these resonant structures that vibrate (strongly), they did not work well for other type of systems and an input different from an impulse. Extensions to the stochastic realization problem considered the use of finite sample average estimates of the covariance function as an attempt to make the method work with finite data length sequences. As indicated in section "The Stochastic Realization Problem," these approximations of the covariance function tended to violate the positive realness property of the underlying power spectrum.

In the early 1990s of the twentieth century, new breakthroughs were made working directly with the input–output data of an assumed LTI system without the need to first compute the Markov parameters or estimating the samples of covariance functions. Pioneers that contributed to these breakthroughs were Van Overschee and De Moor, introducing the N4SID approach (Van Overschee and De Moor 1994); Verhaegen, introducing the MOESP approach (Verhaegen 1994); and Larimore, presenting ST in the framework of canonical variate analysis (CVA) (Larimore 1990).

These three pioneering contributions considered the identification of the state-space model (1) from the input–output data $\{u(k), y(k)\}_{k=1}^{N}$ recorded in *open loop*. The pair (A, C) was assumed to be observable, and the pair (A, KR) controllable. The innovation noise covariance matrix R was assumed to be positive definite.

The formulation of the *intermediate ST step* from which these three pioneering contributions can be derived (by weighting the result of Theorem 1) and that is at the heart of many more variants is summarized in Theorem 1. This theorem requires two preparations: first the storage of the input and output sequences into (block-) Hankel matrices and relating these Hankel matrices via the model parameters and second to make three observations about the model (1) when presented in the prediction form. This form is obtained by replacing $x(k)$ by $\hat{x}(k)$ and $e(k)$ by $y(k) - C\hat{x}(k) - Du(k)$ and is given by

$$\begin{aligned} \hat{x}(k+1) &= (A - KC)\hat{x}(k) + (B - KD)u(k) + Ky(k) \\ y(k) &= C\hat{x}(k) + Du(k) + e(k) \end{aligned} \tag{14}$$

To compact the notation we make the following substitutions: $\mathcal{A} = (A - KC)$ and $\mathcal{B} = [(B - KD) \;\; K]$.

Let the Hankel matrix with the "future" part $\{y(k)\}_{k=p+1}^{N}$ be defined as

$$Y_f = \begin{bmatrix} y(p+1) & y(p+2) & \cdots & y(N-f+1) \\ y(p+2) & & & \\ \vdots & & \ddots & \\ y(p+f) & & \cdots & y(N) \end{bmatrix} \tag{15}$$

for the dimensioning parameters p and f selected such that

$$p \geq f > n$$

In a similar way we define the Hankel matrices U_f and E_f from the input $u(k)$ and the innovation $e(k)$, respectively. Then with the definition of the (block-)Toeplitz matrix T_u from the quadruple of system matrices (A, B, C, D) as

$$T_u = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & & 0 \\ CAB & CB & & 0 \\ & & \ddots & \\ CA^{f-1}B & CA^{f-2}B & \cdots & D \end{bmatrix}$$

and similarly the definition of the Toeplitz matrix T_e from the quadruple of system matrices (A, K, C, I), we can relate the data Hankel matrices Y_f and U_f as

$$\begin{aligned} Y_f &= O_f \left[\hat{x}(p+1) \cdots \hat{x}(N-f+1)\right] + T_u U_f + T_e E_f \\ &= O_f \hat{X}_f + T_u U_f + T_e E_f \end{aligned} \tag{16}$$

Based on the prediction form (14), **3**, key observations are made to support the rational of the intermediate step summarized in Theorem 1:

O1: The standard assumption that the transfer function from $e(k)$ to $y(k)$ is minimum phase leads to the fact that matrix $\mathcal{A}$ is stable. Therefore, there exists a finite integer p such that

$$\mathcal{A}^p \approx 0$$

O2: The state-pace model of (14) has inputs $u(k)$ and $y(k)$. Grouping both together into the new vector $z(k) = \begin{bmatrix} u(k) \\ y(k) \end{bmatrix}$ enables to express the state $\hat{x}(k+p)$ as

$$\hat{x}(k+p) = \mathcal{A}^p \hat{x}(k) + \sum_{j=1}^{p} \mathcal{A}^{j-1} \mathcal{B} z(k+p-j)$$

for $k \geq 1$. With the assumption that $\mathcal{A}^p \approx 0$ and the definition of the input-output data vector sequence $Z(k) = \left[z(k)^T \cdots z(k+p-1)^T\right]^T$, we have the following approximation of the state:

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_107-1

$$\hat{x}(k+p) \approx \left[\mathcal{A}^{p-1}\mathcal{B} \cdots \mathcal{B}\right] Z(k) = \mathcal{L}^z Z(k)$$

As such the state sequence $\hat{X}_f$ in (16) can be approximated by $\mathcal{L}^z Z_p = \mathcal{L}^z \left[Z(1)\ Z(2) \cdots Z(N-f-p+1)\right]$.

O3: The (approximate) knowledge of the row space of the state sequence in $\hat{X}_f$ makes that the unknown system matrices $(\mathcal{A}, \mathcal{B}, C, D, K)$ appear (approximately) linearly in the model (14).

The intermediate ST step to retrieve a matrix with relevant subspaces is summarized in the following theorem taken from Peternell et al. (1996).

Theorem 1 (Peternell et al. 1996) *Consider the model* (1) *with all stochastic processes assumed to be ergodic and with the input $u(k)$ to be statistically uncorrelated from the innovation $e(\ell)$ for all k, ℓ. Consider the following least-squares problem:*

$$\left[\hat{L}_N^u\ \hat{L}_N^z\right] = \arg\min_{L^u, L^z} \|Y_f - \left[L^u\ L^z\right] \begin{bmatrix} U_f \\ Z_p \end{bmatrix} \|_F^2 \tag{17}$$

with $\|.\|_F^2$ denoting the Frobenius norm of a matrix, then

$$\lim_{N \to \infty} \hat{L}_N^z = \mathcal{O}_f \mathcal{L}_z + \mathcal{O}_f \mathcal{A}^p \Delta_z$$

with Δ_z a bounded matrix.

The theorem delivers the matrix $\hat{L}_N^z$ via the solution of a convex linear least-squares problem that has asymptotically (in the number of measurements N) the extended observability matrix $\mathcal{O}_f$ as its column space and that has asymptotically (in the number of measurements **as well as** in the dimension parameter p) the matrix $\mathcal{L}^z$ as its row space. Based on the expression of the state sequence $\hat{X}_f$ given in the observation O2 above, the estimate of the row space of $\mathcal{L}^z$ delivers an estimate of the row space of the state sequence $\mathcal{X}_f$. The observation O3 then shows that this intermediate step allows to derive an estimate of the system matrices $[A, B, C, D, K]$ (up to a similarity transformation) via a linear least-squares problem.

Towards Understanding the Statistical Properties

Many ST variants for system identification using data recorded in open loop have been developed since the early 1990s of the twentieth century. These variants mainly differ in the use of weighting matrices $\mathcal{W}_\ell$ and $\mathcal{W}_r$ in the product $\mathcal{W}_\ell \hat{L}_N^z \mathcal{W}_r$ prior to computing the subspaces of interest. The effect on the accuracy and the statistical properties of the estimated model by these weighting matrices is yet not fully understood as is that of the dimensioning parameters p and f in the definition of the data Hankel matrices Y_f, U_f, Z_p. Only for very specific restrictions results have been achieved. For example, in Bauer and Ljung (2002), it has been shown that when the input $u(k)$ in (1) is either non-present or zero-mean white noise, as well as when the system order n of the underlying system to be known and letting in addition to the dimension parameter p and the number of data points N the dimension parameter f go to infinity, that the weighting matrices selected to represent the CVA approach (Larimore 1990) yield an optimal *minimum variance*

DOI 10.1007/978-1-4471-5102-9_107-1

estimate. A framework for analyzing the statistical properties like consistency and asymptotic distribution of the estimates determined by the class of STs that were discovered in the 1990s is given in Bauer (2005).

The minimum variance property of the estimates by the CVA approach (Larimore 1990) is theoretically not yet proven for more generic and practically relevant experimental conditions. For these cases, the choices of the different weighting matrices, the dimensioning parameters f, p, as well as selecting the system order are often diverted to user. Despite this fact, practical evidence has shown that STs are able to accurately identify state-space models for LTI MIMO systems under industrially realistic circumstances. As such they are by now accepted and widely used as a common engineering tool in various areas, such as model-based control, fault diagnostics, etc. Further they generally provide excellent initial estimates to the nonlinear parametric optimization methods in prediction error or maximum likelihood estimation methods.

Identification of LTI MIMO Systems in Closed Loop

The least-squares problem (17) in Theorem 1 leads to biased estimates when using input-output data that is recorded in a closed-loop identification experiment. This is because of the correlation between the measurable input and the innovation sequence. A number of solutions have been developed to overcome this problem. We refer to the paper van der Veen et al. (2013) for an overview of a number of these rescues. A simple and performant rescue is described here based on the work in Chiuso (2010). The *intermediate ST step* in order to avoid biased estimates is to estimate a high-order vector autoregressive models with exogenous inputs, a so-called VARX model:

$$\min_{\Theta} \sum_{k=1}^{N-p} \| y(k+p) - \Theta Z(k) - D u(k+p) \|_2^2 \tag{18}$$

Using the result on the approximation of the state vector $\hat{x}(k+p)$ in observation O2, it can be shown that the solution $\hat{\Theta}$ of (18) is an approximation of the parameter vector:

$$\hat{\Theta} = \left[\widehat{\mathcal{C}\mathcal{A}^{p-1}\mathcal{B}} \cdots \widehat{\mathcal{C}\mathcal{B}} \right]$$

Then using this solution $\hat{\Theta}$ and O1 above leads to the following "subspace revealing matrix" (cf. Fig. 1):

$$\begin{bmatrix} \widehat{\mathcal{C}\mathcal{A}^{p-1}\mathcal{B}} & \widehat{\mathcal{C}\mathcal{A}^{p-2}\mathcal{B}} & \cdots & \widehat{\mathcal{C}\mathcal{A}^{p-f}\mathcal{B}} & \cdots & \widehat{\mathcal{C}\mathcal{B}} \\ 0 & \widehat{\mathcal{C}\mathcal{A}^{p-1}\mathcal{B}} & & \widehat{\mathcal{C}\mathcal{A}^{p-f+1}\mathcal{B}} & \cdots & \widehat{\mathcal{C}\mathcal{A}\mathcal{B}} \\ \vdots & & \ddots & & & \\ 0 & 0 & \cdots & \widehat{\mathcal{C}\mathcal{A}^{p-1}\mathcal{B}} & \cdots & \widehat{\mathcal{C}\mathcal{A}^{f-1}\mathcal{B}} \end{bmatrix} \tag{19}$$

As in the open-loop case of section "Identification of LTI MIMO Systems in Open Loop," column and row weighting matrices as well as changing the size of the subspace revealing matrix (19) can be used to influence the accuracy of the estimates (Chiuso 2010). The subspace of interest of

Encyclopedia of Systems and Control
DOI 10.1007/978-1-4471-5102-9_107-1

this weighted subspace revealing matrix is its row space that is an approximation of that of the state sequence $\hat{X}_f$ as in (16), now extended to make the size compatible to the weighted version of (19). Similarly as in the open-loop case, knowledge of this subspace turns the estimation of the system matrices $[A, B, C, D, K]$ (up to a similarity transformation) into a linear least-squares problem. The statistical asymptotic properties of this closed-loop ST and the treatment of the dimensioning parameters have also been studied in Chiuso (2010). Here, the result is proven that the asymptotic variance of any system invariant of the model estimated via the above closed-loop ST is a nonincreasing function of the dimensioning parameter f when the input $u(k)$ to the plant is generated by an LQG controller with a white-noise reference input.

Beyond LTI Systems

The summarized discrete-time ST methodology has been extended in various ways. A number of important extensions including representative papers are towards continuous-time systems (van der Veen et al. 2013), using frequency-domain data (Cauberghe 2006) or for different classes of nonlinear systems, like block-oriented Wiener and/or Hammerstein and linear parameter-varying systems (van Wingerden and Verhaegen 2009). ST for linear time-varying systems with changing dimension of the state vector is treated in Verhaegen and Yu (1995), and finally we mention the developments to make ST recursive (van der Veen et al. 2013).

Summary and Future Directions

Subspace techniques aim at simplifying the system identification cycle and make it more user-friendly. Still a number of challenges persist in improving on this general goal. A critical one is the "optimal" selection of the weighting matrices and the dimensioning parameters p and f of the subspace revealing matrix. Optimality here can be expressed, e.g., by the minimality of the variance of the estimates but could also be viewed more generally in relationship with the use of the model, e.g., in terms of the performance of a model-based closed-loop design. A profound theoretical framework is necessary to fully automate the selection of the weighting matrices and dimensioning and order indices. This would substantially contribute to fully automated identification procedures for doing system identification (for linear systems).

A second challenge is to better integrate ST with robust controller design. This requires the assessment of the model quality and the selection of an optimal input. Particular to the integration of ST to control design is the striking similarity of data equations used in ST and model predictive control. The challenge is to further exploit this similarity to develop data-driven model predictive control methodologies that are robust w.r.t. the identified model uncertainty.

One interesting development in ST is the use of regularization via the nuclear norm in order to improve the model order selection with respect to, e.g., SVD-based ST in Liu and Vandenberghe (2010).

A final challenge is to extend ST for LTI systems to other classes of dynamic systems, such as nonlinear, hybrid, and large-scale systems.

DOI 10.1007/978-1-4471-5102-9_107-1

Cross-References

- ▶ Innovation Model
- ▶ Linear Time Invariant
- ▶ Linear Least Squares
- ▶ Markov Parameters
- ▶ Nuclear Norm
- ▶ State Space Models
- ▶ SVD
- ▶ Subspace
- ▶ Similarity Transformation

Recommended Readings

The recommended readings for further study are the books that appeared on the topic of subspace identification. In the books Verhaegen and Verdult (2007) and Katayama (2005), the topic of subspace identification is treated in a wider context for classroom teaching at the MSc level since more elaborate topics relevant in the understanding of ST are treated, such as key results from linear algebra, linear least squares, and Kalman filtering. The book Van Overschee and De Moor (1996) is focused on subspace identification only and also emphasizes the success of ST on various applications. All these books provide access to numerical implementations for getting hands-on experience with the methods. The integration of subspace methods with other identification approaches is done in the toolbox (Ljung 2007).

There also exist a number of overview articles. An overview of the early developments of ST since the 1990s of the twentieth century is given in Viberg (1995). Here also the link between ST for identifying dynamical systems and the signal processing application of direction-of-arrival problems was clearly made. A more recent overview article is van der Veen et al. (2013). In this article also reference is made to the statistical analysis and closed-loop application of ST.

Many papers have appeared reporting successful application of subspace methods in practical applications. We refer to the book Van Overschee and De Moor (1996) and the overview paper van der Veen et al. (2013).

Bibliography

Bauer D (2005) Asymptotic properties of subspace estimators. Automatica 41(3):359–376

Bauer D, Ljung L (2002) Some facts about the choice of the weighting matrices in larimore type of subspace algorithms. Automatica 38(5):763–773

Cauberghe B, Guillaume P, Pintelon R, Verboven P (2006) Frequency-domain subspace identification using {FRF} data from arbitrary signals. J Sound Vib 290(3–5):555–571

Chiuso A (2010) Asymptotic properties of closed-loop cca-type subspace identification. IEEE-TAC 55(3):634–649

Jansson M, Wahlberg B (1996) A linear regression approach to state-space subspace systems. Signal Process 52:103–129

DOI 10.1007/978-1-4471-5102-9_107-1

Juang J-N, Pappa RS (1985) Approximate linear realizations of given dimension via Ho's algorithm. J Guid Control Dyn 8(5):620–627

Katayama T (2005) Subspace methods for system identification. Springer, London

Kronecker L (1890) Algebraische reduktion der schaaren bilinearer formen. S.B. Akad. Berlin, pp 663–776

Larimore W (1990) Canonical variate analysis in identification, filtering, and adaptive control. In: Proceedings of the 29th IEEE conference on decision and control, 1990, Honolulu, vol 2, pp 596–604

Liu Z, Vandenberghe L (2010) Interior-point method for nuclear norm approximation with application to system identification. SIAM J Matrix Anal Appl 31(3):1235–1256

Ljung L (2007) The system identification toolbox: the manual. The MathWorks Inc., Natick. 1st edition 1986, 7th edition 2007

Peternell K, Scherrer W, Deistler M (1996) Statistical analysis of novel subspace identification methods. Signal Process 52(2):161–177

Schutter BD (2000) Minimal state space realization in linear system theory: an overview. J Comput Appl Math 121(1–2):331–354

van der Veen GJ, van Wingerden JW, Bergamasco M, Lovera M, Verhaegen M (2013) Closed-loop subspace identification methods: an overview. IET Control Theory Appl 7(10):1339–1358

Van Overschee P, De Moor B (1993) Subspace algorithms for the stochastic identification problem. Automatica 29(3):649–660

Van Overschee P, De Moor B (1994) N4sid: subspace algorithms for the identification of combined deterministic-stochastic systems. Automatica 30(1):75–93

Van Overschee P, De Moor B (1995) A unifying theorem for three subspace system identification algorithms. Automatica 31(12):1853–1864

Van Overschee P, De Moor B (1996) Identification for linear systems: theory – implementation – applications. Kluwer Academic Publisher Group, Dordrecht

van Wingerden J, Verhaegen M (2009) Subspace identification of bilinear and LPV systems for open and closed loop data. Automatica 45(2):372–381

Verhaegen M (1994) Identification of the deterministic part of mimo state space models given in innovations form from input–output data. Automatica 30(1):61–74

Verhaegen M, Verdult V (2007) Filtering and identification: a least squares approach. Cambridge University Press, Cambridge/New York

Verhaegen M, Yu X (1995) A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. Automatica 31(2):201–216

Viberg M (1995) Subspace-based methods for the identification of linear time-invariant systems. Automatica 31(12):1835–1851