Recursive Subspace Identification in a Hilbert Space Framework

Von der Fakultät für Ingenieurwissenschaften, Abteilung Maschinenbau und Verfahrenstechnik der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades

eines

Doktors der Ingenieurwissenschaften

Dr.-Ing.

genehmigte Dissertation

von

Andreas Bathelt
aus
Eisenhüttenstadt

Nessuna humana investigazione si pio dimandara vera scienzia s'essa non passa per le matematiche dimonstrazione.

No human investigation can be called real science if it cannot be demonstrated mathematically.

Keine menschliche Forschung darf sich als wirkliche Wissenschaft bezeichnen, wenn sie nicht den mathematischen Beweis durchlaufen hat.

(Leonardo da Vinci)

Vorwort

Rückblickend ist eine Promotion wahrscheinlich mehr als nur eine einfache wissenschaftliche Arbeit. Dies zeigt sich auch daran, wenn man sich anerkennend alle Mitstreiter vor Augen führt, die in den verschiedenen Stadien der Arbeit vom Anfang bis hin zum Ende ihre Beiträge geleistet haben. Direkt zu Beginn der Arbeit Mitte 2013 zeigt sich dies, als die Weisheit „Sie werden sich verändern." an mich herangetragen wurde, welche sich bis zum Ende bewahrheiten sollte. Am Anfang weiß man noch nicht, wo man hinkommt, und am Ende weiß man nicht mehr, wo man herkam. Das eine, das man aber von Anfang an wissen und einhalten sollte, ist der Grundsatz „Niemals aufgeben und immer weitermachen!".

Am Anfang dieser Arbeit stand nur eine Idee aus meiner Masterarbeit, welche sich mit dem einfachen Satz „Ich möchte Rekursive Subspace Identifikation entwickeln." beschreiben lässt, da es dazu noch nichts zu geben schien. Das ist nun, mit dem heutigen Wissen betrachtet, doch recht mutig gewesen, da bis Mitte 2016 unklar war, welcher theoretische Ansatz für die Lösung dieser Aufgabe geeignet ist. Letztlich ist damit eine Promotion zu einem gewissen Teil auch eine Wette auf sich selbst. Man hat nun das warme Haus des betreuten Lernens verlassen und wirft sich in die eisige Welt der Wissenschaft, in der man, wenn es darauf ankommt, niemanden fragen kann, wie man denn nun das vorliegende Problem lösen soll, da es nur eine Person gibt, die mit dem Problem vertraut ist – und das ist man selbst. Zu Beginn gibt es nur die Hoffnung, klug genug zu sein, um das gesetzte Ziel zu erreichen. Je höher aber das Ziel gesteckt wird, um so größer ist dann auch die Veränderung. Das bekommt man dann allerdings erst am Ende mit, denn unterwegs sind letztlich die treibenden Kräfte einfach nur das Interesse am Ergebnis und das permanente Ausloten der eigenen geistigen Grenzen, d.h. die Frage „Bekomme ich das hin und falls ja, wie?". Das ist im Prinzip der Niemals-aufgeben-Teil einer Promotion.

Der Immer-weitermachen-Teil ist dann nicht ganz so angenehm. Das sind nämlich in den jeweiligen Momenten Tiefschläge, zu denen man entweder herausfindet, dass es ja doch schon etwas in der Welt der Wissenschaft zum Thema gibt, z. B. rekursive Subspace Identifikation für den geschlossenen Kreis mittels des Ansatzes der orthogonalen Zerlegung, oder wenn man sich bis zur ersten wirklich großen Konferenz durchgekämpft hat, glaubt, man hätte es geschafft, und dann doch nur wieder feststellen muss, dass die Denkweise noch nicht das notwendige wissenschaftlich Niveau erreicht hat. Am Ende sind es aber keine Tiefschläge, sondern die entscheidenden Momente, in denen man entweder aufgibt oder bereit ist, seine Grenzen zu verschieben. Es sind die Zeitpunkte, in denen man am nächsten Tag die Motivation wiederfindet und verbissen mit einer Jetzt-erst-recht-Einstellung weitermacht. Da man noch nicht gut genug war, müssen sowohl der eigene Ansatz als auch die Arbeitsweise halt nur noch besser werden als bisher. Das Erste sorgte in der Endkonsequenz dafür, dass sich in dieser Arbeit die Aussage von Herrn da Vinci bewahrheitete, indem am Ende die komplette mathematische Beschreibung eines zu Beginn eher rein technischen Problems stand – das war so nie geplant, da zu Beginn nicht gewußt wurde, wo es hingeht. Das Zweite führte dann auf der zweiten großen Konferenz dazu, dass man beispielsweise von anderen gesagt bekommt, dass der eigene Ansatz zu schwer sei, um ihn zu verstehen.

Damit lässt sich nun das Ergebnis der letzten fünf Jahre, an dem neben dem Promovierenden durch ihre Unterstützung auch noch viele weitere Menschen beteiligt waren, wie folgt zusammenfassen. Es ist als fast sicher zu definieren, dass ich diese Dissertation vor fünf Jahren in die Ecke geworfen hätte, da ich kein Wort verstanden hätte. Abschließend lässt sich somit feststellen, dass eine ziemlich offensichtliche Veränderung stattfand – q. e. d. Vielleicht ist auch ein Teil mehr Ironie im Leben hinzugekommen.

In diesem Sinn gilt daher mein Dank allen, die mich auf dem Weg zum erfolgreichen Abschluß der Arbeit begleitet haben und zur einführend genannten Veränderung beigetragen haben – es hätte schließlich auch anders ausgehen können. Dies sind auf der fachlichen Seite alle, die mit ihren Hinweisen dieses Dokument bis zu seiner letztlichen Fertigstellung noch einmal verbesserten. Hinzu kommt auch die Anerkennung für die halbjährlichen Vortragstrainings – am Ende ist das ziemlich nützlich, da somit die Verteidigung auch nur wieder zu einem „dieser Vorträge" wird. Auch die Unterstützung hinsichtlich verschiedener nicht unwichtiger Kleinigkeiten, wie z. B. der Formatvorlagen für die Dissertation oder Tipps für die Dokumente der Einreichung, durch meinen Weggefährten soll an dieser Stelle nicht ungenannt bleiben. Vor allem gilt aber mein Dank der einen Person, die mir die Möglichkeit zur eigenen Forschung ermöglichte – sowohl finanziell als auch organisatorisch – und mir eigentlich ab etwa Anfang 2014 freie Hand für die Entwicklung der Ideen ließ und nicht permanent mit „Sind Sie schon fertig?" im Rücken saß. Immerhin waren die Konferenzen in Hiroshima, Whistler, Osaka und Melbourne nun nicht gerade Next Door – schon allein die Flüge. Auf der sozialen Seite sind es meine Freunde, bei denen ich mich dann doch das eine oder andere Mal über die Unfairness der Welt, d. h. über die Unfairness einer Promotion, denn das war in den letzten fünf Jahren „meine Welt", lautstark beschweren und so mein inneres Gleichgewicht wieder herstellen konnte.

Kurzfassung

Seit dem Aufkommen modellgestützter Methoden in der Automatisierungstechnik für die Regelung oder Überwachung industrieller Großanlagen, wie z. B. verfahrenstechnische Anlagen oder Walzwerke in der Stahlindustrie, ist die Modellbildung dieser Anlagen wesentlich für die Nutzung solcher Methoden. Hierbei ist im Vergleich zur theoretischen Modellbildung die experimentelle Modellbildung mittels Systemidentifikation ein einfacherer Weg, um an eine mathematische Beschreibung einer Anlage zu gelangen. Obwohl das Anlagenverhalten grundsätzlich nicht-linearer Natur ist, verweilt die Anlage dennoch meist über längere Zeit in einem Arbeitspunkt. Somit ist die Identifikation eines linear parameter-varianten Modelles zur Beschreibung des Gesamtverhaltens eher ungerechtfertigt, wenn der dafür benötigte Aufwand ins Verhältnis zum Nutzen gestellt wird. Es ist einfacher, ein lineares zeit-invariantes Modell für die Beschreibung eines Arbeitspunktes adaptiv anzupassen, d. h. rekursiv zu identifizieren, wenn die Anlage in einen neuen Arbeitspunkt überführt wird. Bezüglich dieser Anforderung bestanden hinsichtlich der Subspace Identifikation folgende Probleme:

- Bestehende Ansätze für die rekursive Subspace Identifikation setzen an der numerischen Implementierung der Methoden an, nutzen damit nicht das grundlegende theoretische Gerüst der Subspace Identifikation und führen letztlich zu speziellen Methoden.

- Trotz der vorhandenen verschiedenen methodischen Ansätze, welche sich für die Identifikation von industriellen Großanlagen eignen, sind deren numerischen Implementierungen eher ungeeignet für einen Einsatz im Rahmen solcher Anlagen.

Ausgehend von dieser Problemstellung befasst sich diese Dissertation mit der Beschreibung eines neuen Ansatzes für die rekursive Subspace Identifikation und der Herleitung zugehöriger Basisalgorithmen für die Identifikationen. Dabei wurden folgende Ergebnisse erzielt:

- Für die Identifikation im offenen Kreis, d. h. im Falle einer ungeregelten Anlage, bei der die Eingänge unabhängig von den Ausgängen sind, wurde durch die Verbindung des Ansatzes der orthogonalen Zerlegung des Anlagenausganges mit der auf der Canonical Correlation Analysis aufsetzenden Identifikationsmethode ein Algorithmus hergeleitet, der wesentlich besser als bisherige Algorithmen das Prozessmodell industrieller Großanlagen auch unter Störeinflüssen bestimmen kann.

- Für die Identifikation im geschlossenen Kreis, d. h. im Falle einer geregelten Anlage, bei der die Eingänge abhängig von den Ausgängen sind, wurde für die Prediktor-basierende Subspace Identifikationsmethode ein Algorithmus hergeleitet, der deren theoretische Beschreibung direkt implementiert und somit bisherige aufwendige Least-squares Verfahren oder eine Identifikation eines Zwischenmodelles unnötig macht.

- Für die rekursive Subspace Identifikation wurde basierend auf dem aus der stochastischen Realisierungstheorie stammenden Coordinate-free Framework eine grundlegende Beschreibung des rekursiven Verhaltens des zu einem Modell äquivalenten Predictor Space hergeleitet. Während dieser Ansatz aufgrund der Eigenschaft des Predictor Spaces die höchstmögliche Kompression von Informationen der Vergangenheit erzielt,

ist er darüber hinaus vor allem allgemeingültig. Damit kann er auf jede beliebige bestehende Subspace Methode angewendet werden und macht Spezialmethoden unnötig.

Der neue Algorithmus für den offenen Kreis erhält seine Leistungsfähigkeit durch die Kombination des allgemein guten Identifikationsverhaltens der Methode auf Basis der Canonical Correlation Analysis mit der sich aus der orthogonalen Zerlegung des Ausganges ergebenden Unterdrückung von Prozessstörungen. Identifikationsstudien, die sowohl auf rein akademischen Beispielen als auch auf dem Tennessee Eastman Process Modell aufsetzen, weisen die Güte des Algorithmus simulativ nach. Wenn Systeme mit entweder autoregressiver moving-average Struktur und exogenem Eingang (deterministisches und stochastisches Subsystem besitzen gleiche Eigenwerte) oder Box-Jenkins Struktur (deterministisches und stochastisches Subsystem besitzen unterschiedliche Eigenwerte) identifiziert werden, sind die durch den Algorithmus erzielten Ergebnisse gleich den Ergebnissen oder minimal besser als die Ergebnisse bekannter Methoden. Hinsichtlich des realistischen Tennessee Eastman Process Modells ist der hier eingeführte Algorithmus bei ungestörtem Prozessausgang besser als vorhandene Identifikationsalgorithmen und bei einer Beeinflussung der Prozessausgänge durch Störungen der einzige Algorithmus, der in Lage ist, das korrekte deterministische Prozessmodell zu bestimmen. Da es sich beim Tennessee Eastman Process um ein Modell handelt, welches entsprechend (nicht-linearer) physikalischer und chemischer Grundprinzipien eine existierende verfahrenstechnische Anlage nachbildet, sind die am Tennessee Eastman Process erzielten Ergebnisse repräsentativ für die Identifikationen solcher Anlagentypen.

Aufgrund des theoretischen Hintergrundes des Ansatzes der orthogonalen Zerlegung kann dieser Ansatz für die Identifikation im geschlossenen Kreis nicht verwendet und folglich auch nicht in den neue Algorithmus für die Prediktor-basierende Subspace Identifikationsmethode integriert werden. Somit beschreibt das ermittelte Modell sowohl das deterministische als auch das stochastische Verhalten des Prozess. Durch die Neuordnung der Daten vor der Berechnung können eine direkte Implementierung der Methode erfolgen und komplexe Least-Squares Ansätze oder eine vorherige Schätzung eine Zwischenmodells vermieden werden. Die Identifikationsergebnisse dieses Algorithmus sind mit minimalen Minderungen gleich denen der bestehenden Algorithmen der Methode. Ein größeres Beispiel unter Nutzung von Modellen, die in der Literatur zur Illustration der Funktionstüchtigkeit der Methoden für die direkte Identifikation im geschlossenen Kreis genutzt wurden, zeigte allerdings, dass die Identifikation der meisten Systemstrukturen im geschlossenen Kreis fehlschlägt. Das betrifft nicht nur den hier eingeführten Algorithmus, sondern alle Methoden für die direkte Identifikation, welche für dieses Beispiel implementiert wurden. Die Klärung dieses Problems ist nicht Bestandteil dieser Arbeit, da deren Schwerpunkt auf der Herleitung des Ansatzes zur rekursiven Subsapce Identifikation liegt.

Der neue Ansatz für die Rekursion beruht auf den Eigenschaften des minimalen Predictor Space. Einerseits verdichtet der minimale Predictor Space alle notwendigen Informationen für die Vorhersage der zukünftigen Entwicklung des Systems, andererseits ist jeder zukünftige minimale Predictor Space ein Unterraum des Raumes, der durch einen beliebigen vergangenen minimalen Predictor Space und die dazwischen liegenden Ein- und Ausgangsdaten aufgespannt wird. Aus der Eigenschaft des zeitlichen Voranschreitens lassen sich unter Beibehaltung der Kompressionseigenschaft die exakte Verbindung eines beliebigen vergangenen minimalen Predictor Space mit einem beliebigen zukünftigen minimalen Predictor Space herstellen und somit eine rekursive Methodik beschreiben, die sich auf beliebige Standardmethoden der Subspace Identifikation anwenden lässt. Simulationen auf Basis akademischer Beispiele veranschaulichen die Funktionsfähigkeit dieses neuen Ansatzes. Außerdem werden

zwei Problemstellungen, die sich aus der Implementierung dieser Methodik ergeben, und entsprechende Lösungsvorschläge erläutert.

Abstract

Since the introduction of model-based methods for control or monitoring of industrial plants, e.g., chemical processes or rolling mills, more and more emphasis is placed on modeling of these processes. In this context, the experimental modeling, i.e., system identification, provides and easier way to achieve this goal than theoretical modeling, i.e., modeling based on first principles. Although the plants are essentially of non-linear nature, they are often operated over long periods of time in one operating point, which exhibits a nearly linear behavior in its proximity. Hence, comparing effort versus reward, it is more rewarding to identify a linear time-invariant model and adapt this model during the change of the operating point, i.e., using a recursive identification scheme, than trying to identify a genuine non-linear model, e.g., a linear parameter-varying model. From this line of thought, the following problems arise if subspace identification is considered:

- Existing methods for recursive subspace identification approach the problem from the numerical point of view and hence do not take the underlying theoretical framework into consideration and, hence, provide purpose-build methods.

- Although there are various methods, which were suitable for the identification of industrial plants, the respective numerical implementations of these methods lack the performance for this intended application.

Based on these problems, this thesis is concerned with the derivation of a new methodological approach to recursive subspace identification and the derivation of related algorithms of base methods. The following results are achieved:

- In terms of open-loop identification, i.e., if the plant is not controlled and there is no feedback or dependency between the output and the input, an algorithm is proposed which combines the approach of the orthogonal decomposition of the system output with the core algorithm of the method based on the canonical correlation analysis. This new algorithm is able to retrieve the model of the plant's process under influences of process disturbances much better than existing algorithms.

- In terms of closed-loop identification, i.e., if the plant is controlled and there is a feedback or dependency between the output and input, an algorithm for the predictor-based subspace identification method is described, which follows exactly the method's theoretical derivation. This algorithm thus avoids the involved least-squares approaches or identifications of intermediate models of previous algorithms.

- In terms of recursive subspace identification, a methodological approach is proposed, which is based on the underlying coordinate-free framework originating from stochastic realization theory. This approach describes the recursion based on the model-equivalent predictor space by using its properties with respect to its progression with time and data compression capabilities. In addition to achieving the highest possible compression of past data, the approach can be implemented in terms of every existing subspace identification method and makes the derivation of purpose-build methods unnecessary.

The proposed algorithm for open-loop identification draws its strength from the fact that the good identification performance of the canonical correlation analysis is applied to data which is free from influences of disturbances. This in turn follows from the integration of the approach of orthogonal decomposition, which decomposes the output data of the system into the deterministic component, i.e., the system-driven data component, and the stochastic component, i.e., the disturbance-driven data component. Identification studies using both academic examples and the Tennessee Eastman Process model show the accuracy of the proposed algorithm. In terms of the academic examples, the quality of the results is equal or slightly better when compared to existing methods. This holds in both cases when system is of autoregressive moving-average structure with exogenous input (dynamics of stochastic and deterministic subsystems are the same) or of Box-Jenkins structure (dynamics of stochastic and deterministic subsystems are disjoint). In terms of the Tennessee Eastman Process model, the algorithm yields results with improved quality in a disturbance-free scenario and outperforms existing methods when disturbances are present. As the Tennessee Eastman Process is a first-principle model of an existing chemical plant, the results achieved on this model are representative for the identifications of such processes.

For reasons of the underlying theoretical relations, it is not possible to merge the orthogonal decomposition approach with the predictor-based subspace identification method. Hence, the proposed algorithm yields a joint stochastic-deterministic model of the system. By re-arranging the data before entering the processing stages of the identification algorithm, the proposes algorithm is able to directly follow the theoretical derivations of the predictor-based subspace identification method. The results of this algorithm are equal to the previously proposed algorithms, where the existing algorithms hold a minimal advantage. Unfortunately, examples using systems with either autoregressive moving-average structure with exogenous input, autoregressive structure with exogenous input, or Box-Jenkins structure under closed-loop conditions resulted in erroneous identifications. Although these very examples have been used in the literature to illustrate the functionality of the respective methods, this problem does not only concern the proposed algorithm but all implementations of methods for the direct approach to closed-loop identification. A solution of this issue is not pursued in this thesis as the main focus is on the derivation of an approach to recursive subspace identification.

The new approach to recursive subspace identification is based on the properties of the minimal predictor space. Firstly, it compresses all past information necessary for the prediction of the future development of the system, and, secondly, any future minimal predictor space is a subspace of the joint space of an arbitrary past minimal predictor space and the respective intermediate data. While retaining the compression property, the exact relation between a future minimal predictor space and a past minimal predictor space is derived by exploiting the property describing the progression with time. Thus, the sought-after recursive methodology is established. This results in an approach which can be implemented in terms of every standard subspace method. Numerical simulations show the proper functioning of the approach in terms of a academic examples. In addition to the theoretical derivation, two practical issues and its possible solutions are addressed.

Contents

List of abbreviations

ARMAX   Autoregressive moving-average with exogenous input (model structure)

ARX   Autoregressive with exogenous input (model structure)

AIC   Akaike's information theoretic criterion

CCA   Canonical correlation analysis (subspace identification method)

CVA   Canonical variate analysis (subspace identification method)

FIR   Finite impulse response (model structure)

FPE   Akaike's final prediction error criterion

IV   Instrumental variable

LTI   Linear time-invariant (system/model type)

LPV   Linear parameter-varying (model structure)

MIMO   Multiple-input multiple-output (system/model type)

MOESP   MIMO/Multivariable output-error state-space model identification (subspace identification method)

N4SID   Numerical algorithms for subspace state-space system identification (subspace identification method)

NIC   Criterion for model order estimation based on Frobenius norm

ORT   Orthogonal decomposition (subspace identification approach)

PARSIM-E   Parsimonious, causal subspace identification method with innovation estimation (subspace identification method)

PBSID   Predictor-based subspace identification (subspace identification method)

SISO   Single input single output (system/model type)

SNR   Signal-to-noise ratio

SSARX   State-space ARX (subspace identification method)

SVC   Singular value criterion, criterion for model order estimation based on 2-norm

SVD   Singular value decomposition

VAF   Coefficient of determination, also called "variance accounted for" criterion

VARX   Vector autoregressive with exogenous input (model structure)

List of symbols

Vectors and matrices

y, u, x	System output, system input, system state (here considered to be stochastic processes)
w, v, e	State noise, measurement noise, innovation (all white-noise processes)
A, B, C, D, K	Matrices of the state-space representation of a system
$y_t^{(\cdot)}$	Data vector of past/future of stochastic process y, where past and future with respect to time t (time point t is included in future) are denoted by replacing $(\cdot)$ with $-$ or $+$
$u_t^{(\cdot)}$	Data vector of past/future of stochastic process u
$p(t)$	Joint random variable of $y(t)$ and $u(t)$
$p_t^{(\cdot)}$	Data vector of past/future of joint stochastic process consisting of y and u
$\mathrm{y_N}(t)$	(Numerical) tail matrix with N columns representing stochastic variable $y(t)$, the columns are the respective numerical values of a sample function of y starting with the element of time t; similar for any other stochastic process, i.e., the input u or the state x
$\mathrm{Y_t^-}$	(Numerical) matrix representing y_t^- and constructed row-wise from tail matrices $\mathrm{y_N}(t)$; similar for $\mathrm{Y_t^+}$, $\mathrm{U_t^-}$, $\mathrm{Y_t^-}$, or $\mathrm{P_t^-}$

Sets and spaces

$\mathcal{Y}_t^{(\cdot)}$	Space spanned by past/future of stochastic process y, where past and future with respect to time t (time point t is included in future) are denoted by replacing $(\cdot)$ with $-$ or $+$
$\mathcal{X}_t^{(\cdot)}$	Space spanned by past/future of stochastic process x (in terms of systems with exogenous input) or the states of the forward and backward model (in terms of stochastic systems)
$\mathcal{U}_t^{(\cdot)}$	Space spanned by past/future of stochastic process u
$\mathcal{P}_t^{(\cdot)}$	Space spanned by past/future of joint stochastic process $p = (y, u)$
$\mathcal{F}_t^+$	Joint space of the future innovations/wandering subspace and input

$\mathcal{H}$	Ambient space; depending on problem (stochastic system or system with exogenous inputs) either $\mathcal{Y}_t^- \vee \mathcal{Y}_t^+ \vee \mathcal{X}_t^- \vee \mathcal{X}_t^+$ or $\mathcal{Y}_t^- \vee \mathcal{Y}_t^+ \vee \mathcal{U}_t^- \vee \mathcal{U}_t^+ \vee \mathcal{X}_t^- \vee \mathcal{X}_t^+$
$\left(\mathcal{Y}_t^{(\cdot)}\right)^{\perp}$	Orthogonal complement of space $\mathcal{Y}_t^{(\cdot)}$ within ambient space
$\mathcal{Y}_{[t,t+k)}^{(\cdot)}$	Space constructed from random variables $y(t)$ through $y(t+k-1)$, where $(\cdot)$ is either $-$ or $+$ denoting its logical assignment to past or future

Operators

$\overline{\mathrm{span}}\{\cdots\}$	Closed Hilbert space spanned by the elements listed in $\{\cdots\}$
$\mathrm{Im}\{\cdot\}$	Range space/image of matrix $(\cdot)$ (spanned by column vectors)
$\mathrm{Im}_{\mathrm{Row}}\{\cdot\}$	Row space of matrix $(\cdot)$ or range space/image of $(\cdot)^{\mathrm{T}}$ (spanned by row vectors)
$\mathrm{Ker}\{\cdot\}$	Nullspace/kernel of matrix $(\cdot)$
$\hat{E}\{a|\mathcal{B}\}$	Orthogonal projection of a onto space $\mathcal{B}$
$\hat{E}_{\|\mathcal{C}}\{a|\mathcal{B}\}$	Oblique projection of a onto space $\mathcal{B}$ along space $\mathcal{C}$; under condition $\mathcal{B} \cap \mathcal{C} = \{0\}$
$\mathcal{A} \vee \mathcal{B}$	Vector sum of $\mathcal{A}$ and $\mathcal{B}$, i.e., joint space of $\mathcal{A}$ and $\mathcal{B}$
$\mathcal{A} + \mathcal{B}$	Direct vector sum of $\mathcal{A}$ and $\mathcal{B}$; $\mathcal{A} \cap \mathcal{B} = \{0\}$
$\mathcal{A} \oplus \mathcal{B}$	Orthogonal direct vector sum of $\mathcal{A}$ and $\mathcal{B}$; $\mathcal{A} \perp \mathcal{B}$
$\mathcal{A} \ominus \mathcal{B}$	Orthogonal complement of $\mathcal{B}$ in $\mathcal{A}$, i.e., $\mathcal{A} - \hat{E}\{\mathcal{A}|\mathcal{B}\}$

Miscellaneous

$\mathcal{A} \perp \mathcal{B}|\mathcal{C}$	Spaces $\mathcal{A}$ and $\mathcal{B}$ are conditional orthogonal given space $\mathcal{C}$; for all elements $a \in \mathcal{A}$ and $b \in \mathcal{B}$, $E\left\{\left(a - \hat{E}\{a|\mathcal{C}\}\right)\left(b - \hat{E}\{b|\mathcal{C}\}\right)^{\mathrm{T}}\right\} = 0$ holds
I_n	Identity matrix; $I_n \in \mathbb{R}^{n \times n}$
$0_{m \times n}$	Zero matrix; $0_{m \times n} \in \mathbb{R}^{m \times n}$
T	Number of data points in the set of recorded data used for identification (or validation)
N	Number of columns in tail matrices $\mathrm{y_N}(\mathrm{t})$
k_p	Length of the past horizon, i.e., number of elements in y_t^-, u_t^-, p_t^-
k_f	Length of the future horizon, i.e., number of elements in y_t^+, u_t^+

1 Introduction

In this chapter, the topic of this thesis will be introduced. Following a general and brief review of system identification in Section 1.1, the object of research of this thesis will be explained in Section 1.2. In this context, the problem as well as the resulting methodical approach, its practical impact, and the main contribution of this work are discussed. A precise formulation of the scope of the work follows later in the third chapter. This formulation cannot be given in a comprehensible way before the review of the field of research – subspace identification – is made in the second chapter. The outline given in Section 1.3 describes the structure of the thesis, which result from the objectives issued for this thesis.

1.1 System identification

Many application in today's world make the existence of a model of the related system necessary. The models in the following are considered to be mathematical descriptions of theses systems. This type of modeling is not limited to certain types of systems. That is, the variety of systems ranges from industrial process, e.g., distillation columns or rolling mills, to the dynamic behavior of car suspensions. In terms of modeling of a system, there are two general approaches to this problem. The first approach is to derive the model from the (nonlinear) physical and chemical principles governing the behavior of the system – known as theoretical modeling. The second approach is to calculate the model from recorded data – known as experimental modeling or system identification. The subject of this thesis resides within experimental modeling.

The systems are usually nonlinear and time-varying. However, as time-variance of most systems results from aging or the change of operating points, the systems behave rather linear in the proximity of these operating points. Hence, linear time-invariant (LTI) models are often used to describe the system behavior. In this context, the system's output can be described by[1]

$$y(t) = P_0(z)u(t) + v_0(t) \; , \tag{1.1}$$

where the output y, the disturbance v_0, and the input u can be either scalars or multi-variable vectors. In the latter case, such a system is called multiple-input multiple-output (MIMO) system. If y, v_0, and u are scalar, a system is called single-input single-output (SISO) system. Hybrid forms are the MISO (multiple-input single-output) or the less common SIMO (single-input multiple-output) systems. The transfer function (scalar case) or transfer matrix (multi-variable case) $P_0(z)$ describes the dynamic behavior of the system.

---

[1]In this thesis, such formulas need to be understood in a symbolic way. The purpose is only the depiction of the connection between the signals and its respective system entities. It is not meant to be an exact mathematical description – which would obviously require the signals to be also in the $\mathcal{Z}$ domain. Thus, $P_0(z)$ is used here in the same symbolic way as in Lindquist and Picci (2015, p. 676), Katayama (2005) or in the style of the equivalent q-transformation used throughout Ljung (2009).

Subspace methods   **Classical methods**

Projection
Singular value
decomposition

Input-output data

Prediction error method
Least-squares

State vector

Transfer matrix

Least-squares

Realization

State space model

State space model

Kalman filter

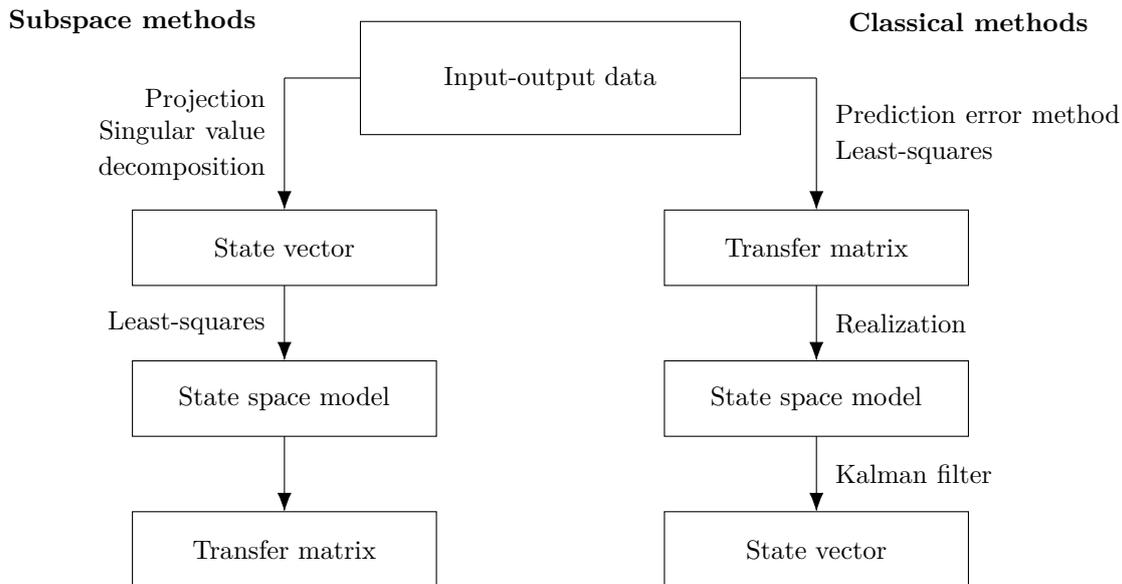Transfer matrix

State vector

Figure 1.1: Approaches to system identification of LTI models (Katayama, 2005; van Overschee and De Moor, 1996)

Given the input and output data of a system, which is excited in a certain way by the input, system identification is concerned with the task of calculating a model of this system representing its controllable and observable properties in the proximity of some point of operation. Usually, a model is restricted to a certain range of closely adjacent operating points or by its intended use. To cover the whole system behavior, nonlinear models are needed.

There are two general approaches to system identification – parametric identification, or classical identification, and subspace identification. The first group of methods estimate the parameters of predefined transfer functions, thus identifying the system in the frequency domain. These types of models are also referred to as black-box models as they only represent the relationship between the input and the output of a system without giving insight into the internal structure of the model. This changes with respect to the model of the latter methods. By estimating the system in the time domain in terms of state-space models, the internal structure of the model is also given. For this reason, these types of models are also called white-box models. The procedures of these two approaches are shown in Figure 1.1. Due to the calculation based on sampled data of the system's input and output, the models of both approaches are always linear discrete-time models. With respect to this data acquisition, it is furthermore distinguished between two modes – open-loop and closed-loop. In open-loop operation, there is no feedback from the output to the input, i.e., there is no controller, and the input is hence independent from output, whereas there is a feedback, i.e., a controller, and hence a dependency of the input on the previous output data, in closed-loop operation. In terms of this thesis, the object of research will be concerned with a certain aspect of subspace identification for both open-loop and closed-loop data acquisition. To point out the differences between classical identification and subspace identification, both approaches will be briefly reviewed in the following.

Classical identification

In terms of classical identification, the model of a system is given by

$$\mathcal{M}: \quad y(t) = P(z, \theta)u(t) + H(z, \theta)e(t) , \tag{1.2}$$

where θ is an explicit parameter vector. The disturbance v_0 is modeled using the white-noise process e and a respective filter $H(z, \theta)$. The relation between the transfer functions $P(z, \theta)$ and $H(z, \theta)$ describes certain model structures, like the ARX (autoregressive with exogenous input) model

$$y(t) = \frac{B(z, \theta_B)}{A(z, \theta_A)}u(t) + \frac{1}{A(z, \theta_A)}e(t) ,$$

the ARXMAX (autoregressive moving average with exogenous input) model

$$y(t) = \frac{B(z, \theta_B)}{A(z, \theta_A)}u(t) + \frac{C(z, \theta_C)}{A(z, \theta_A)}e(t) ,$$

or the most flexible Box-Jenkins model

$$y(t) = \frac{B(z, \theta_B)}{F(z, \theta_F)}u(t) + \frac{C(z, \theta_C)}{D(z, \theta_D)}e(t) .$$

The part of the model given by the transfer function $P(z, \theta)$ is also called deterministic part of the model or deterministic sub-model, while the part described by $H(z, \theta)$ is called stochastic part or stochastic sub-model. In all model structures, the polynomials $A(z, \theta_A)$, $B(z, \theta_B)$, $C(z, \theta_C)$, $D(z, \theta_D)$, and $F(z, \theta_F)$ might have varying degrees, which are respectively fixed to a certain number before the identification starts. Usually, more than one model structure and varying degrees, i.e., different models (defined by the structure of the transfer functions and the degrees of the polynomials), are used during the identification process. Based on certain criteria, the model giving the best fit is used after a validation phase. As the degree of the polynomials is fixed, the parameter vector can be split up according to the various polynomials. In fact, the parameter vector θ contains the coefficients of the polynomials. Hence, this type of system identification is only concerned with the numerical estimation of suitable coefficients of these polynomials, as the model structure is pre-defined and not identified. The most common methods to do so are either the prediction-error method or the instrumental-variable method. Both are covered in the books of Ljung (2009) and Söderström and Stoica (2001), wherein also the entire field of classical identification is addressed. This contains also questions regarding data generation or consistency and convergence of the results, i.e., whether there is a bias within the models and which value the covariance of the parameters has. An overview of the field is given in Åström and Eykhoff (1971), whereas a survey on the types of models is also given in B. Huang and Kadali (2008). Another big group of methods are maximum likelihood methods, which are for example covered in Hannan and Deistler (2012). The more complex the models become, so do the related identifications. In the end, most of the methods result in nonlinear optimization problems.

Subspace identification

Compared to the classical methods, which were introduced roughly 50 years ago, the field of subspace system identification started to develop approximately 25 years ago. The first ideas

can be however traced back to the 1970th, e.g., Akaike (1974). Following from these first ideas, the original problem of subspace identification was the identification of time series or stochastic systems as for example addressed in van Overschee and De Moor (1993), Dahlén and Scherrer (2004), or Tanaka and Katayama (2005). The idea was then extended to the identification of joint stochastic-deterministic systems. In terms of subspace methods, the model is assumed to be given by the state-space representation

$$
\mathcal{M}: \quad
\begin{aligned}
x(t+1) &= A(\theta)x(t) + B(\theta)u(t) + K(\theta)e(t) \,, \\
y(t) &= C(\theta)x(t) + D(\theta)u(t) + e(t) \,,
\end{aligned}
\tag{1.3}
$$

where the parameter vector θ is only used to showcase the fact that the matrices will be determined during the identification. That is, there is no explicit set of parameters whose values will be determined during the course of the identification. This also becomes clear if the connection between state-space models and transfer-function models given by

$$
\begin{aligned}
y(t) &= \left(C(\theta) \left(zI - A(\theta) \right)^{-1} B(\theta) + D(\theta) \right) u(t) + \left(C(\theta) \left(zI - A(\theta) \right)^{-1} K(\theta) + I \right) e(t) \\
&= P(z,\theta)u(t) + H(z,\theta)e(t)
\end{aligned}
\tag{1.4}
$$

is considered. As there are no unique state-space realizations, $P(z,\theta)$ and $H(z,\theta)$ can be described by various state-space realizations. In this context, θ loses its meaning as one unique parameter vector. This also means that the model structure will ideally follow from the identification of system matrices and is consequently identified as well instead of being fixed beforehand. Other advantages are the reduced number of method parameters and the inherent ability to similarly deal with single-input single-output and multiple-input multiple-output systems, as the model is given by a state-space representation. Only the order of the system/model is needed and not the whole structure of the model. However, trade-offs for the aforementioned advantages are an increased estimation variance and a reduced estimation accuracy of zeros and steady-state gains, which is in contrast achieved by classical approaches (Qin, Lin, and Ljung, 2005). Thus, it is more convenient to use classical approaches as far as the identification of SISO systems or simple MIMO systems is concerned. As stated in Katayama (2005), the early success of subspace methods was due to the fact that the numerical implementation is based on the reliable numerical algorithms of the QR decomposition and singular value decomposition (SVD). Thus, nonlinear optimization techniques, which are required to solve the parameter estimation of the classical approaches, are avoided.

In the following, a brief overview of subspace identification is given. In terms of an accurate summary of contributions given to field, a distinction between *approaches*, *methods*, and *algorithms* will be made here and in the remainder of the thesis. Approach refers to a basic concept or template procedure for the identification of a system. Method refers to the exact theoretical description of an identification procedure. Methods can be assigned to the different approaches. Algorithm refers to the respective numerical implementation of a method. It is possible that there are multiple algorithms for one method. Methods and related aspects which are used and referred to in the thesis will be reviewed in detail in Chapter 2. Following Viberg (1995) and Bauer and Ljung (2002), the two basic approaches to the problem of subspace identification are the *realization approach* and the *state-regression approach*. In terms of the later approach, the system matrices are determined by an estimation of the states $x(t)$ and $x(t+1)$, so that by (1.3) a linear regression is described. The former approach is based on an extraction of the system matrices from certain matrices which follow when (1.3) is iterated to yield a description for a contiguous interval of outputs. A special approach limited

to the open-loop case is the orthogonal decomposition (ORT) approach discussed in Picci and Katayama (1996b). By embedding this approach into the algorithms of the methods, a decomposition of the system output y into its deterministic and stochastic components is achieved. Hence, the disjoint identifications of the deterministic and stochastic subsystems of (1.3) is allowed for. By the end of the 1990th, the open-loop problem was covered by the realization method MOESP[2] (Verhaegen, 1993b, 1994; Verhaegen and Dewilde, 1992a), the state-regression method N4SID[3] (van Overschee and De Moor, 1994, 1996), and the state-regression/realization methods CCA[4] (Katayama and Picci, 1999) and its earlier version CVA[5] (Larimore, 1983, 1990). These methods are now considered standard open-loop methods. N4SID is also covered in the monograph by van Overschee and De Moor (1996), where certain aspects of the subspace approach are set into context of non-steady-state Kalman filter estimation. As the name indicates, the CCA method uses the algorithm of the canonical correlation analysis to extract the estimates of the states from the data. Opposed to the usual rank n estimations of the state or observability matrix by N4SID or MOESP, the estimation of the state by a canonical correlation analysis assigns a deeper meaning to this estimation as it will not yield just any rank n state estimate but the estimate with the highest correlation between past and future. A realization method, which uses the principal component analysis (PCA) as the core algorithm and is hence called SIMPCA[6], has been discussed in B. Huang, Ding, and Qin (2005) and Wang and Qin (2002, 2006). The closed-loop problem was first addressed by a joint input-output algorithm of the MOESP method in Verhaegen (1993a). By the beginning of the 2000th, the subspace identification approach was entirely extended to closed-loop identification when the direct-approach methods PARSIM-E[7], SSARX[8], and PB-SID[9] (all state-regression methods) were disclosed in Qin and Ljung (2003a), Jansson (2003) and Chiuso and Picci (2005). The identification by the SSARX and PBSID methods require a pre-estimation of a vector ARX (VARX) model and are, in terms of the basic procedures, similar to each other. The procedure of the PARSIM-E method includes a step-wise estimation of the innovation process e. With either the VARX model or the information of the innovation process, the bias which would otherwise be introduced in the models of open-loop methods by the feedback can be avoided. In addition to these standard methods, statistical properties, such as consistency or asymptotic variance, have been thoroughly analyzed, see, e.g., Peternell, Scherrer, and Deistler (1996), Bauer (2005), Gustafsson (2002), Jansson and Wahlberg (1998), Knudsen (2001), or Chiuso and Picci (2004e). Furthermore, a great variety of methods ranging from the identification in the frequency domain (Akçay, 2011; Hinnen, Verhaegen, and Doelman, 2005; McKelvey, Akçay, and Ljung, 1996; van Overschee, De Moor, et al., 1997), over continuous time identification, e.g., D. Huang and Katayama (2004) and Ohta (2011), to methods for linear parameter-varying (LPV) systems (Favoreel, De Moor, and van Overschee, 1999; van Wingerden and Verhaegen, 2009; Verdult and Verhaegen, 2002) have been proposed over time. Algorithms for the recursive identification are discussed for exmple in Lovera, Gustafsson, and Verhaegen (2000), Mercère, Bako, and Lecuche (2008), and Oku and Kimura (1999). Common to all proposed recursive approaches is the fact that the recursive identification is approached from an algorithmic standpoint without taking the underlying methodological concept of subspace identification into account. That is pronoun-

---

[2]MIMO/Multivariable output-error state-space model identification

[3]Numerical algorithms for subspace state-space system identification

[4]Canonical correlation analysis

[5]Canonical variate analysis

[6]Subspace identification method via principal component analysis

[7]Parsimonious subspace identification, see Qin, Lin, and Ljung (2005), with innovation estimation

[8]State-space ARX

[9]Predictor-based subspace identification

ced by the fact that the basic method which is used within all recursive algorithms is MOESP. This results in recursions which are not based on the previous model or entities equal to the model, like state estimates, but on intermediate data of the identification process.

Due to the rigid theoretical framework of subspace identification, there are only a limit number of genuine methods, which are covered by the mentioned early papers. In later papers, the focus switched increasingly to implementational issues and the discussion of different algorithms for existing methods, e.g., the LPV method of van Wingerden and Verhaegen (2009) is a LPV version of the PBSID method, or the nuclear norm methods proposed in Sugie, Inoue, and Maruta (2017) and Verhaegen and Hansson (2014, 2016) are implementations of the ORT approach and the PBSID method. These papers do hence not contribute from the standpoint of methods, as the core can be traced back to already given method. For this reason, such papers are not mentioned.

The theoretical framework of subspace identification evolves from the stochastic realization theory, which also contributes to the fact that the input and the output of a system are considered to be stochastic processes. A discussion of subspace identification within this framework is made in Katayama (2005), which also covers the ORT approach and the CCA method. An advanced and exhaustive discussion of the fundamentals is given in Lindquist and Picci (2015). In this framework, the procedures of subspace identification are described in terms of data spaces, which carry the information presented by stochastic processes, i.e., the input and output of a system. As, hence, any ties to a generic state-space model are cut, the results are given also in terms of subspaces of these data spaces and not in terms of a particular data vector, i.e., not in terms of a basis of these subspaces. In fact, the identification procedure turns into a question of extracting a certain subspace – called *predictor space* – within an ambient data space. As outlined at the beginning of Lindquist and Picci (1996b), this predictor space is equal to the model in terms of the contained information. This rationale can be extended as to interpret this subspace as an entity equal to the parameter vector θ of the classical identification. This framework for subspace identification, for obvious reasons called *coordinate-free framework*, will be heavily used in this thesis, as it constitutes a firm basis for the questions to be addressed.

1.2 Object of research

The object of research is recursive subspace identification. Although motivated by a purely practical aspect – identification of large industrial processes, like the Tennessee Eastman Process (see Appendix A) – the considerations and research eventually led to a theoretical approach to the problem. This approach is fundamentally different from the previously proposed algorithms, as those algorithms are not derived against the background of the coordinate-free framework, which however will be used here. A more in-depth discussion of the problems and the proposed approach is made in Chapter 3, which in turn can only be comprehensively discussed once the review of subspace identification of Chapter 2 is given. This structure is needed, as the idea does not only result in an approach to recursive subspace identification but also in some modification of existing methods.

1.2.1 Problems and motivation

The motivation of the work results from the question regarding the identification of realistic systems of industrial scale, like chemical plants or rolling mills. In terms of such systems, the identification methods have to cope with the following three challenges:

1. Identification of multi-variable systems. In this context, the number of inputs and outputs can surpass 5 each. For example, the basic configuration of the Tennessee Eastman Process features 12 inputs and 22 outputs. This dimension is common and quite often even surpassed. Depending on the intended use of the model, either a subsystem consisting of certain input-output combinations or, in terms of the inputs and outputs, the whole system is to be identified.

2. Identification of nonlinear systems. The internal physical and chemical principles of the process are given by nonlinear relations[10]. With increasing dimensionality of the process, this becomes also increasingly pronounced. In this context, either an identification of a nonlinear model is required if the whole system behavior is to be covered, or an identification of a linear model is required if only a limited operating range is needed. In terms of the latter case, the methods are however required to yield a sufficiently good linear approximation of the actual nonlinear behavior in the proximity of the operating point.

3. Identification of systems subjected to disturbances. These disturbances stem either from environmental influences, influences of upstream process units, the feedback from downstream process units, or internal disturbances. These disturbances, which are commonly modeled by arbitrary colored-noise sequences, need to be rejected if only the deterministic system behavior is to be modeled.

Whereas the first point and third point are problems which are easy to overcome by any method which includes the ORT approach, a problem defining question arises from the second point. Although LPV methods address the identification of dynamically nonlinear systems, the numerical load to perform such identifications tend to approach an amount too large to handle, if the system becomes multi-variable (in the range mentioned in the first point).

However, the usual operation of such industrial systems, where this becomes a problem, is characterized by

- infrequent changes of operating points, i.e., large industrial plants are usually operated in a full-continuous manner in which the operating points often remain the same over weeks and months,

- slow transition rates, i.e., the change between operating points is a slow process and might take hours, and

- sufficiently accurate approximations of local behavior by MIMO LTI models, i.e., as the systems' nonlinearities are of dynamic nature, small deviations from the operating point will only show a weak nonlinear behavior.

The actual nonlinear behavior of the plants shows itself mostly in terms of the difference between the dynamic behavior of the operating points. Taking these three points into consideration, an identification using LPV models might be an unnecessary effort if an LTI model describes the plant behavior in the proximity of an operating point sufficiently good. This

---

[10]Which are, technically speaking, again just models of observations of relations between physical and chemical values.
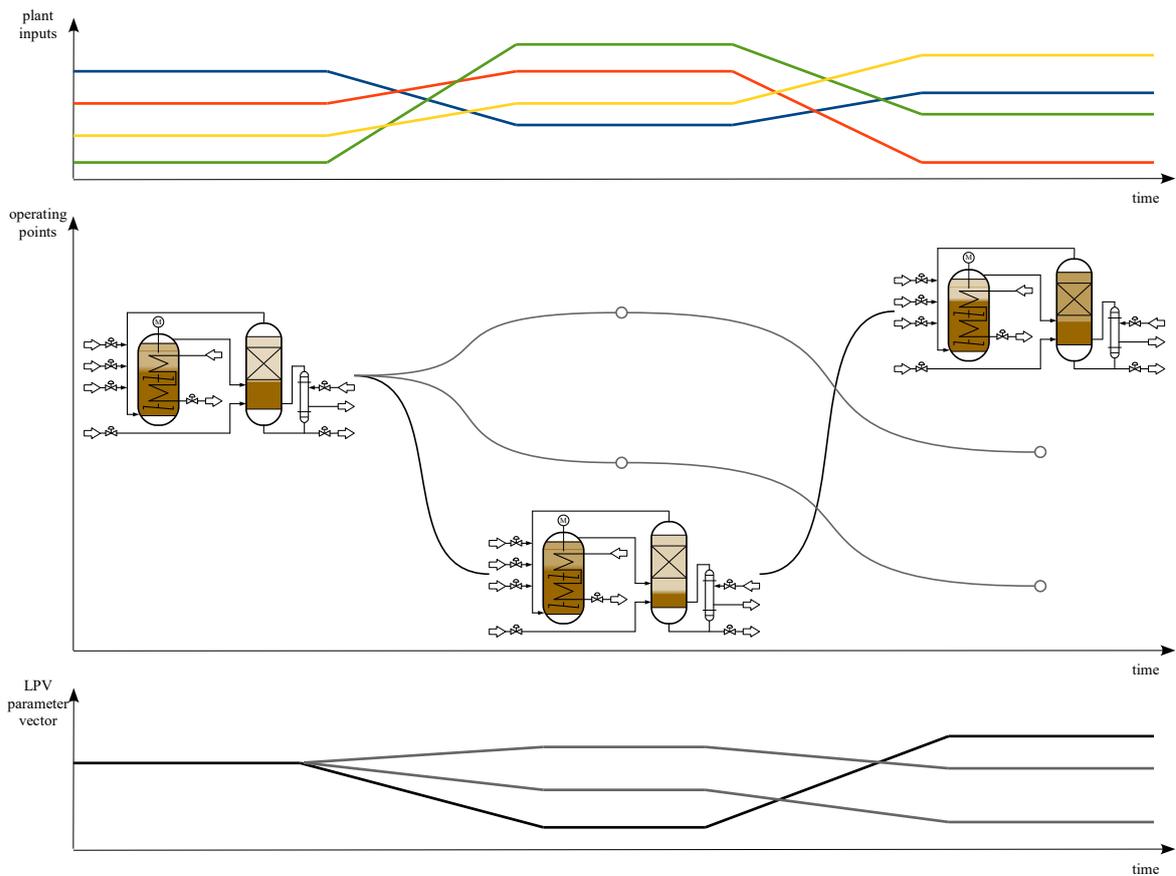
Figure 1.2: Trajectory of plant operating points and representation of possible operating points by a LPV model; the plant trajectory shown in the middle diagram follows from the input trajectories exemplified in the top diagramy whereas the gray plant trajectories can additionally represented by a LPV model, which is governed by the parameter vector shown in the bottom diagram (the black parameter vector represents the actual plant trajectory)

idea is illustrated in Figure 1.2. The upper diagram shows the trajectories of the inputs, the middle diagram shows the trajectory of the system operating point, and the lower diagram shows three possible trajectories of the parameter vector of a LPV model. The actual progression of the operating point of the plant, which is governed by the input, is given by the black operating-point trajectory and the symbolization of the plant during the phase of stationarity. The remaining gray trajectories show possible alternative trajectories of a LPV model of the plant, which is governed by the inputs *and* the parameter vector. The respective parameter-vector trajectories are given in the lower diagram by the gray curves. A LPV model covers hence a much larger range of possible models for the operating points than actually needed at any time during the plant operation. Another problem which should be taken into consideration is aging. Even if the LPV identification is carried out exactly at the present time, the plant behavior changes over the time so that the operating point defined by an input-parameter-vector set might later not conform anymore with the actual operating point. If however a LTI model is used, only the momentary, actual operating point is represented. This model is then updated before or once the next operating point is reached, i.e., the actual progression of the operating point of a plant needs to be traced.

This rationale of adaptive modeling leads to recursive identification and in particular to recursive subspace identification. From the standpoint of updating a model, a recursive procedure as done in terms of classical methods, where the actual parameter vector θ is updated, see Chapter 11 in Ljung (2009, pp. 361–382), is preferred, as the model is directly updated in a manner

$$\mathcal{M}_{t_0} + Z^{[t_0,t_1]} \to \mathcal{M}_{t_1} , \tag{1.5}$$

where $Z^{[t_0,t_1]}$ denotes the intermediate data gathered in the interval $[t_0, t_1]$. In terms of subspace identification, this is equal to an update of the system matrices, or in terms of the coordinate-free framework, to an update of the aforementioned predictor space. Following this idea, recursive subspace identification is approached from the methodological point of view and not as done before from the algorithmic point of view.

1.2.2 Practical impact

A recursive update of much simpler LTI models circumvents the complex and time consuming identification of dynamically nonlinear models. Furthermore, the recursive update does not require a complete re-identification of the system as the data from previous identifications, such as system matrices or the predictor space of the coordinate-free framework, already establishes a data basis. Using a recursive approach, the new model of the plant can be determined while the system itself is transfered into the new operating point. That is, by the time the new operating point is reached, the model of this operating point is also given.

This leads to an easy-to-use approach to the identification of large plants, which might lead to a more frequent use of identification. This in turn facilitates the use of model-based process monitoring approaches (e.g., Ding (2013), which has the same basis as subspace identification) or control performance monitoring approaches (e.g., Jelali (2013), where the determination of a model is a crucial step for the definition of the performance indices). Both improved process monitoring and improved control performance monitoring result in better plant operation and reduced loss of material and wear of the plant[11].

1.2.3 Proposed theoretical method

In terms of the principle given by (1.5), the recursive approach might be done by a simple iteration of the state equation of (1.3) based on the model data of $\mathcal{M}_{t_0}$, which also includes the state estimate $\hat{x}(t_0)$. In this case, the recursion would be given by

$$
\begin{aligned}
\hat{x}(t_1) &= \hat{x}(t_0) + \sum_{k=t_0}^{t_1-1} \hat{A}^{t-1-k}\hat{B}u(k) + \sum_{k=t_0}^{t_1-1} \hat{A}^{t-1-k}\hat{K}e(k) \\
&= \hat{x}(t_0) + \sum_{k=t_0}^{t_1-1} (\hat{A}-\hat{K}\hat{C})^{t-1-k}(\hat{B}-\hat{K}\hat{D})u(k) + \sum_{k=t_0}^{t_1-1} (\hat{A}-\hat{K}\hat{C})^{t-1-k}\hat{K}y(k) ,
\end{aligned}
\tag{1.6}
$$

where the matrices $\hat{A}, \hat{B}, \hat{C}, \hat{D}$, and $\hat{K}$ denote the estimates of the system $\Sigma = (A, B, C, D, K)$. However, the derivation of the recursive approach will be mainly concerned with the question regarding reachability and observability of the resulting model. In a coordinate-based frame-

---

[11]And hence, of course, more profit as the loss of money on plant maintenance is reduced. This is however not the focus or the motivation of this thesis.

work as in (1.6), this can only be achieved by imposing restrictions or assumptions on the system, i.e., on $\Sigma = (A, B, C, D, K)$. Furthermore, the analysis is made in some fixed basis of $\hat{x}(t)$, for which reason it is not clear whether the findings hold also for other bases or if $\hat{x}(t)$ is not minimal. Thus, a coordinate-based approach makes the derivation of a recursive scheme and the analysis of the properties unnecessary complicated. Taking furthermore time-varying systems into account, the question with respect to taking changes of the system within the interval $[t_0, t_1]$ in (1.6) into consideration appears.

As mentioned before, the coordinate-free framework cuts the ties with a model and the hence problems arising from the use of a generic model disappear as well. This framework facilitates the discussion regarding reachability and observability to be made solely in terms of the predictor space (essentially, a space of the state $x(t)$) and the data spaces of the input and the output. Therefore, the approach to recursive subspace identification will be based on the coordinate-free framework. The predictor space $\mathcal{X}_t^{+/-}$ (see definition in Section 2.4) is the respective model entity within this framework as it represents the same information as the system matrices or θ. Thus, the approach to recursive subspace identification is essentially found if a predictor space of time t_1 can be calculated based on a previously determined predictor space $\mathcal{X}_{t_0}^{+/-}$, or if the predictor space $\mathcal{X}_{t_0}^{+/-}$ can be adapted by the intermediate data as

$$\mathcal{X}_{t_0}^{+/-} + Z^{[t_0,t_1]} \to \mathcal{X}_{t_1}^{+/-} \ . \tag{1.7}$$

Such an approach also conforms with the recursive methods of the classical identification, where the parameter vector is updated based on a previous parameter vector θ and the intermediate data.

In terms of application, some of the assumptions imposed for the theoretical derivation of the approach need to be relaxed later on. This concerns mainly the stationarity of the input and output processes. However, the general behavior of the approach remains as the relaxations will not effect the quintessence of the approach. That is, it will facilitate the tracking of changes within a process. The time-varying behavior, as which a nonlinear behavior can also be understood, will only introduce minor deviations from the theoretically derived behavior of the approach. These deviations are however minor in comparison to deviations introduced by disturbances (see, e.g., Figure 6.5, Figure 6.6, Figure 6.7, and Figure 6.8, in Section 6.4.2).

1.2.4 Main contributions of the work

The pursuit of a methodological approach for recursive subspace identification resulted in the following main contributions:

- Derivation of a general approach for recursive subspace identification using the coordinate-free framework
  The description of a recursive scheme for subspace identification is made based on coordinate-free framework and hence approaches the problem from the underlying principles of subspace identification and not in terms of the numerical implementation of one subspace method. The formulation of the recursion follows from the properties of the predictor space which can be updated by new data as it evolves through time. This yields a general approach for recursive subspace identification, which can be implemented in terms of every existing subspace method.

- Derivation of an identification algorithm for the reliable extraction of the deterministic plant behavior from disturbance-affected output data under open-loop conditions

By redefining the procedure of the ORT approach, a combination of this approach with the CCA method (or N4SID method) is made feasible. The resulting combined algorithm of CCA and ORT improves the previously given algorithm of the ORT approach based on the MOESP method. As illustrated by a case study using the Tennessee Eastman Process, the combined algorithm facilitates the correct estimation of the plant model even if the output is subjected to arbitrarily colored noise disturbances of high magnitude. This algorithm can be hence applied in an industrial environment.

- Derivation of an identification algorithm for the direct calculation of the predictor space/state under closed-loop conditions
  The algorithm directly implements the theoretically derived procedure of the PBSID method and omits the pre-estimation of the VARX model or an involved least-squares estimation, which are common for the implementations of the methods of the direct approach of closed-loop identification. It also simplifies the integration of the recursive scheme in conjunction with the PBISD method.

Related to these contributions are 3 publications and a paper in preparation. In Bathelt and Jelali, 2014, the problem of identifying a realistic process model is presented. The actually poor results of this study led to the derivation of the open-loop identification algorithm in Bathelt, Söffker, and Jelali, 2015. Some of the results regarding recursive subspace identification based on the coordinate-free framework are discussed in Bathelt, Söffker, and Jelali, 2017, 2018.

1.3 Outline

This thesis covers the work done to achieve the goal set in the previous section. It thus presents a new way to recursive subspace identification and also some new algorithms for existing methods.

First, in Chapter 2, the field of subspace system identification is reviewed. This review contains the introduction of the coordinate-free framework but also of a number of methods. This method review will lead on the one hand to a thorough understanding of the motivation for the new recursive approach and on the other hand is needed to understand the modification of the existing algorithms and the respective motivation.

In Chapter 3, the actual challenges resulting from the task of identifying industrial processes are identified and the thorough motivation of the work is given. From this motivation, the two main tasks – derivation of the recursive approach and modifications of the algorithms of certain methods – and their conceptual outlines are deduced. It also serves as the end-point of the first part containing the review and motivation and as the lead-in to the second part covering the actual work.

Chapter 4 contains the smaller contribution of the thesis. Here, algorithms of two subspace methods are described. The first algorithm is a modification of the CCA method (Katayama and Picci, 1999), which then will be able to determine the deterministic sub-model $P(z, \theta)$ even if y is a sum of a deterministic and stochastic component, i.e, if the system's output is subjected to disturbances. This is made possible by integrating the idea of the orthogonal decomposition of the ORT approach by Picci and Katayama (1996b) into the core algorithm of the CCA method. Whereas the first algorithm is an open-loop algorithm, the second is an algorithm for the PBSID method (Chiuso and Picci, 2005). This new algorithm facilitates a

recursive use of the PBSID method, a trait which is not possible given the existing algorithms. The derivations of both algorithms is accompanied by respective examples.

To round the derivations of identification algorithms off, an identification study using the realistic process model of the Tennessee Eastman Process is given in Chapter 5. This study showcases the improved results of the combined algorithm of the CCA method and ORT approach as well as the capability of the new algorithm of the PBSID method to yield the same results as the existing algorithms of the PBSID method while dealing with realistic processes.

The main theoretical contribution of the thesis is presented in Chapter 6. Here, the theoretical framework for recursive subspace identification is derived. The derived approach is not a method but a general methodology which can be transferred to any method of subspace identification. In addition to the derivation, which also covers the modeling of stochastic process, i.e., the identification of time series, practical aspects are addressed. The most important aspect is the explanation of an approach which increases the capability of tracking of fast changing systems. The tracking capability as of now is limited by the representation of stochastic processes by data matrices. The second aspect describes a basic concept to keep the basis of the model the same, as it is not possible to guarantee that the resulting model basis after a new recursion cycle is the same as in the previous cycle.

The thesis closes with a summary and suggestions for future work resulting from the work covered in this thesis.

Regarding theorems, lemmas, and propositions adopted from the literature, the proofs of such are not included, if those are – except for the wording – directly taken out of the respective references. The same holds also for theorems, lemmas, and propositions which are a summary of more than one theorem, lemma, or proposition.

2 Review of subspace identification

This chapter provides the introduction to the field of subspace identification. Here, the necessary basic principles and concepts are outlined. The main focus is on the introduction of the coordinate-free framework of Chiuso and Picci (2003) and Lindquist and Picci (1996b, 2015), which the procedures of subspace identification naturally appear from. The method review is limited to the basic facts and methods touched upon in the thesis. Proofs of lemmas and theorems are omitted and can be found in the respective references.

The first two sections of the chapter introduce the basics. The procedure of an identification is briefly described in Section 2.1, which also motivates certain aspects of the following review. Basic definitions of system representations, data spaces and projections are given in Section 2.2. The discussion of the data spaces shows the one-to-one relationship between spaces of stochastic variables and spaces of numerical matrices, which in turn clarifies the easy implementation of subspace method as the theoretical derivations can be given direct numerical implementation. The discussion of the projections is intended to shed some light on the actual meaning of the projection as minimum variance estimations. A fact which will be later referred to when the problem of order estimation is discussed for the case study based on the Tennessee Eastman Process. The basic identification problem and related assumptions of subspace identification are formulated in Section 2.3. Section 2.4 introduces the necessary foundation of the coordinate-free framework referred to during the derivation of the recursive scheme. The following sections provide the method review. This review is limited to the aspects of subspace identification which are necessary for the understanding of the motivation of the work, i.e., to understand the problems of the given methods and algorithms. This review is also the basis for the derivation of the algorithms in Chapter 4, or is referred to in later explanations. First, the basic (coordinate-based) equations are introduced in Section 2.5. Then, open-loop and closed-loop methods will be summarized in Section 2.6 and Section 2.7. The following Section 2.8 and Section 2.9 will briefly summarize the identification of systems by linear parameter-varying models and review previously disclosed methods and approaches related to recursive subspace identification.

2.1 General identification procedure

In order to identify a model of the dynamic behavior of a plant, i.e., the dynamic behavior of the underlying process, the actual identification procedure consists of several steps. This procedure is illustrated in Figure 2.1 and given a more detailed explanation in Ljung (2009, pp. 13–15). First, the inputs and outputs of the process, which will be identified, need to be specified. In this context, inputs and outputs do not refer to the physical process inputs and outputs (in Figure 2.1 marked by arrows) but to system-theoretical inputs and outputs like valves and measurement points. Then follows the data acquisition. During this phase, the process is excited by the specified inputs, and the respective plant response is recorded. This input-output data set is then used for the actual identification. In terms of subspace identification, this identification step consists of the calculation of the model
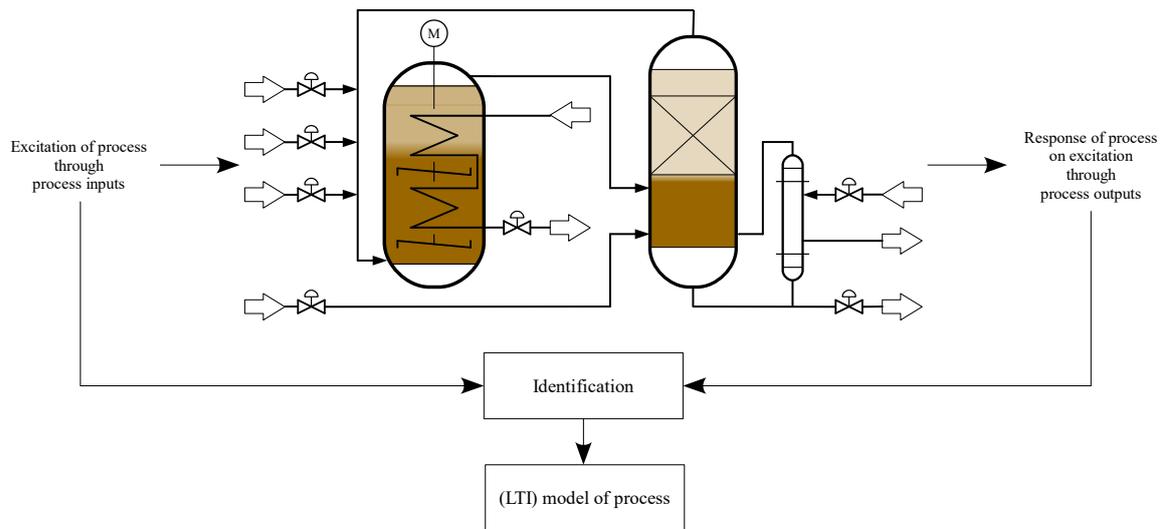
Figure 2.1: Schematic diagram of the identification procedure

and the validation of the model. If the validation of calculated model meets the specified requirements, the model can be used for the intended application, e.g., controller design or plant monitoring. An important aspect, which essentially influences every step, is the existence of a priori knowledge regarding the process. This can influence the identification in terms of the generated excitation signal, e.g., amplitudes of the signal or signal type, the choice of the model, e.g., LTI model or LPV model, or the way the identification is done, i.e., recursive or non-recursive identification.

This brief explanation already shows, that there are several aspects to be taken into account when dealing with system identification or deriving new algorithms, methods, or even approach to certain problems like recursive subspace identification. These aspects concern the components of the methods, like data representation for the derivation, the fundamental assumption, which the derivations are made under, or the basic ideas and procedures of the methods itself. These points will be reviewed in the following to raise the awareness of the problems solved by the algorithms and approach to recursive identification proposed later in this thesis.

2.2 Basic components of identification methods

All methods and related assumptions for the derivation and the use of these methods are based on some common components. These fundamental components will be summarized in this section. First, the different representations of a system or a model in terms of state-space descriptions are discussed. Secondly, the data spaces are introduced. The important point in terms of the data spaces is the fact that the definition as spaces of stochastic processes can be given an equivalent definition based on the sample functions of these stochastic processes. Thirdly, the projections used to decompose the data and hence identify the system are explained. The notation is adopted from Katayama (2005).

2.2.1 System representations

The first problem to be addressed by every identification method is with respect to the representation of the process dynamics, i.e., the model structure. As far as subspace identification for linear time-invariant models is concerned, the structure is given by a generic state-space representation. As outlined in Qin (2006), any LTI system can be represented by three different but equivalent sets of state-space equations. The basic representation is the process form, which is given by

$$x_p(t+1) = Ax_p(t) + Bu(t) + w(t) \,, \tag{2.1a}$$
$$y(t) = Cx_p(t) + Du(t) + v(t) \,, \tag{2.1b}$$

where $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$, $x_p \in \mathbb{R}^n$, $w \in \mathbb{R}^n$, and $v \in \mathbb{R}^p$ denote the input, output, state as well as the state noise and measurement noise. In addition to the white-noise processes w and v, the input, output, and state are also assumed to be stationary zero-mean second-order Gaussian processes[1]. That is,

$$E\{y(t)\} = 0 \,, \qquad E\{y(t)y^{\mathrm{T}}(s)\} = \Lambda_{yy}(t,s) < \infty$$

are fulfilled, where by reason of stationarity Λ_{yy} furthermore depends only on the difference between t and s, i.e., (Katayama, 2005, p. 77)

$$\Lambda_{yy}(t,s) = \Lambda_{yy}(t-s,0) = \Lambda_{yy}(t-s) = \Lambda_{yy}(l) = E\{y(t+l)y^{\mathrm{T}}(t)\} \,. \tag{2.2}$$

The restriction to Gaussian processes simplifies the following derivations as conditions, which otherwise would appear, can be omitted. Under this restriction, the innovation process $e \in \mathbb{R}^p$, defined by

$$e(t) = y(t) - E\{y(t) \mid u(s), \, y(s-1); \, s \le t\} \,, \tag{2.3}$$

is also Gaussian and automatically fulfills certain assumption (cf. Hannan and Deistler, 2012 or Chiuso and Picci, 2004e, pp. 278–279). The innovation process is the causal prediction error or conditional innovation of y given u up to the present time t (Picci and Katayama, 1996b) and combines the influences of w and v. Based on the Kalman filter theory, the system can be represented by using this innovation process in terms of the innovation form (see, e.g., Katayama, 2005, Chapter 5)

$$x_i(t+1) = Ax_i(t) + Bu(t) + Ke(t) \,, \tag{2.4a}$$
$$y(t) = Cx_i(t) + Du(t) + e(t) \,. \tag{2.4b}$$

The model of the system is usually given in terms of this innovation form. Note that the state $x_i(t)$ of the innovation form is not equal to the state $x_p(t)$ of the process form, although the output is the same (see, e.g., Katayama, 2005, p. 121). If this, $x_p(t) = x_i(t)$, would hold true, the innovation could be defined by

$$e(t) = y(t) - Cx_p(t) - Du(t) = v(t) \,,$$

---

[1]In this case, the entities may more precisely be denoted as stochastic processes $u\colon T \times \Omega \to \mathbb{R}^m$, $y\colon T \times \Omega \to \mathbb{R}^p$, $x\colon T \times \Omega \to \mathbb{R}^n$, $w\colon T \times \Omega \to \mathbb{R}^n$, and $v\colon T \times \Omega \to \mathbb{R}^p$, where Ω and T denote the sample space and the index set (time). In terms of the identification problem at hand, the index set contains discrete time points, i.e., $T \ne \mathbb{R}$ but $T \subset \mathbb{R}$, where it holds with proper inclusion.

which would erroneously state equivalence between v and e. Based on the innovation form of the system representations, the innovation process is also given by

$$e(t) = y(t) - Cx_i(t) - Du(t) ,$$

which yields the predictor form

$$x_i(t+1) = A_K x_i(t) + B_K u(t) + Ky(t) , \qquad (2.5a)$$
$$y(t) = Cx_i(t) + Du(t) + e(t) , \qquad (2.5b)$$

where

$$A_K = A - KC , \qquad B_K = B - KD .$$

This representation is used to define the state in terms of the past observations of the output y and the input u.

2.2.2 Data spaces and its numerical counterparts

The second problem is the representation and ordering of the recorded data. This varies depending on the different identification approaches, e.g., subspace identification or classical identification. The derivation of the methods is simplified by considering the plant input and output to be stochastic processes and the recorded data to be random variables. It will however follow that there is a rather simple one-to-one relationship between the actual numerical data and the stochastic processes as well as random variables.

In order to identify the system, the stochastic processes u and y are divided around some arbitrarily-chosen "present" time t into a future and past segment. The present time t belongs by definition to the future segment. The lengths of the segments, i.e., the past and future horizons, will be denoted by k_p (past) and k_f (future) or by k if both horizons are assumed to have the same length. This assumption will be made throughout the following discussions, unless the horizon lengths are explicitly denoted by k_p and k_f. The initial time point will be denoted by $t_0 = t - k_p$. In terms of the theoretical discussion of the methods, $t_0 \to -\infty$ will be assumed and k only denotes the (finite) length of the future horizon. Data vectors containing the respective past and future random variables of the processes y and u are defined and denoted by

$$y_t^- = \begin{bmatrix} y(t_0) \\ \vdots \\ y(t-2) \\ y(t-1) \end{bmatrix} , \qquad y_t^+ = \begin{bmatrix} y(t) \\ y(t+1) \\ \vdots \\ y(t+k-1) \end{bmatrix} ; \qquad (2.6)$$

u_t^-, u_t^+, and e_t^+ are defined similarly. Using

$$p(t) = \begin{bmatrix} u(t) \\ y(t) \end{bmatrix} , \qquad (2.7)$$

the joint process (y, u) is defined. The past of the joint process is formed according to (2.6) and denoted by p_t^-. If the length of these data vectors deviates from the length of the actual horizons or the length of the horizon is emphasized, the respective interval is explicitly given

by a subscript. The data vectors are then for example denoted by $y^+_{[t,t+k)}$. Notations like y^-_{t+1} or y^+_{t+1} are the logical extensions of the construction pattern of the data vectors if the "present" time is $t+1$.

Based on these data vectors, the respective data spaces are defined (see, e.g., Katayama, 2005). Taking for example a finite number of output random variables $y(i)$, $i = 1, \cdots N$, the set

$$\mathcal{H} = \left\{ a + \sum_{k=1}^{N} A_k y(k) \,\middle|\, a \in \mathbb{R}^p,\ A_k \in \mathbb{R}^{p \times p} \right\} \tag{2.8}$$

of linear combinations can be formed. Defining furthermore the inner product

$$\langle a, b \rangle_{\mathcal{H}} = E\{a^{\mathrm{T}} b\} = \mathrm{trace}\left(E\{ab^{\mathrm{T}}\} \right) , \tag{2.9}$$

where a, $b \in \mathcal{H}$, the norm

$$\|a\|_{\mathcal{H}} = \sqrt{E\{a^{\mathrm{T}} a\}} = \sqrt{E\{\|a\|_2^2\}} \tag{2.10}$$

is induced. By completing $\mathcal{H}$ with this inner product and this norm, it becomes a finite-dimensional Hilbert space – a finite-dimensional closed vector space with inner product. The construction of a space $\mathcal{Y}$ spanned by an infinite number of elements of y is similar. Equal to (2.8), the set of all finite linear combinations is constructed and an inner product and its induced norm are defined. Then the set of elements $\mathcal{Y}$ is completed by the limits of all Cauchy sequences constructed from elements of $\mathcal{Y}$. Thus, $\mathcal{Y}$ becomes an infinite-dimensional closed Hilbert space. This construction is denoted by

$$\mathcal{Y} = \overline{\mathrm{span}}\{\cdots, y(t-1), y(t), y(t+1), \cdots\} . \tag{2.11}$$

As only zero-mean random variables are considered, a constant element a as in (2.8) is omitted for the construction of $\mathcal{Y}$. The operator $\overline{\mathrm{span}}\{\cdot\}$ likewise defines a finite-dimensional closed Hilbert space, if the number of the listed elements is finite.

Thus, the spaces of the inputs and outputs are given by

$$\begin{aligned} \mathcal{Y} &= \overline{\mathrm{span}}\{\cdots, y(t-1), y(t), y(t+1), \cdots\} , \\ \mathcal{U} &= \overline{\mathrm{span}}\{\cdots, u(t-1), u(t), u(t+1), \cdots\} , \end{aligned} \tag{2.12}$$

and can be interpreted as structures carrying all information contained in the processes (see discussion following Lemma 2.2). The spaces spanned by the past or the future of the respective processes are denoted by

$$\begin{aligned} \mathcal{Y}_t^- &= \overline{\mathrm{span}}\{\cdots, y(t-2), y(t-1)\} , \\ \mathcal{Y}_t^+ &= \overline{\mathrm{span}}\{y(t), y(t+1), \cdots\} \ ; \end{aligned} \tag{2.13}$$

with equal definitions for $\mathcal{U}_t^-$ and $\mathcal{U}_t^+$. These past and future spaces are hence spanned by the elements of u_t^-, u_t^+, y_t^-, and y_t^+, which can also be written as

$$\mathcal{Y}_t^- = \overline{\mathrm{span}}\{y_t^-\} , \qquad \mathcal{Y}_t^+ = \overline{\mathrm{span}}\{y_t^+\} .$$

The space of the joint past $\mathcal{P}_t^-$ is subsequently spanned by p_t^-. Furthermore, spaces spanned by single random variables, e.g., $y(t)$ or $u(t)$, are denoted by $\mathcal{Y}_t$ or $\mathcal{U}_t$. The superscripts $-$ or $+$ denoting past or future are dropped and the subscript denotes the time point of the space.

Similar notations are used for space spanned by elements of a certain interval, e.g., $[t, t+k)$. Those spaces are denoted by

$$
\begin{aligned}
\mathcal{Y}_{[t,t+k)}^{-} &= \overline{\mathrm{span}}\{y(t), y(t+1), \cdots, y(t+k-1)\} \ , \ \text{or} \\
\mathcal{Y}_{[t,t+k)}^{+} &= \overline{\mathrm{span}}\{y(t), y(t+1), \cdots, y(t+k-1)\} \ .
\end{aligned}
\tag{2.14}
$$

The assignment to past or future depends on the context these spaces are used in.

The joint space of two vector spaces $\mathcal{A}$ and $\mathcal{B}$ is denoted by the vector sum

$$
\mathcal{A} \vee \mathcal{B} \ .
\tag{2.15}
$$

If those two spaces do not share elements except for 0, i.e., $\mathcal{A} \cap \mathcal{B} = \{0\}$, the joint space can also be denoted by the direct vector sum

$$
\mathcal{A} + \mathcal{B} \ .
\tag{2.16}
$$

If furthermore $\mathcal{A} \perp \mathcal{B}$, the joint space is given by the orthogonal direct vector sum

$$
\mathcal{A} \oplus \mathcal{B} \ .
\tag{2.17}
$$

The joint past $\mathcal{P}_t^{-}$ is hence also defined by a vector sum based on $\mathcal{Y}_t^{-}$ and $\mathcal{U}_t^{-}$. It is given by

$$
\mathcal{P}_t^{-} = \overline{\mathrm{span}}\{p_t^{-}\} = \mathcal{Y}_t^{-} \vee \mathcal{U}_t^{-} \ .
\tag{2.18}
$$

Regarding the numerical representations of the random variables $y(t)$ and $u(t)$, the concept of tail matrices as discussed in Lindquist and Picci (2015, pp. 508–513) is used. This concept is based on the assumed stationarity of y and u and the ergodic theorem, see, e.g., Katayama (2005) and Rozanov and Feinstein (1967). Hence, the tail matrices

$$
\begin{aligned}
\mathrm{u_N(t)} &= \begin{bmatrix} \mathrm{u(t)} & \mathrm{u(t+1)} & \cdots & \mathrm{u(t+N-1)} \end{bmatrix} \triangleq u(t) \ , \\
\mathrm{y_N(t)} &= \begin{bmatrix} \mathrm{y(t)} & \mathrm{y(t+1)} & \cdots & \mathrm{y(t+N-1)} \end{bmatrix} \triangleq y(t) \ ,
\end{aligned}
\tag{2.19}
$$

where $\mathrm{u(i)}$ and $\mathrm{y(i)}$, $\mathrm{i} = \mathrm{t}, t+1, \cdots, t+N-1$ denote the numerical values of a sample function of the processes, are representations of $y(t)$ and $u(t)$. Consequently,

$$
\begin{aligned}
\mathrm{Y_t^{-}} &= \begin{bmatrix} \mathrm{y_N^T(t_0)} & \cdots & \mathrm{y_N^T(t-2)} & \mathrm{y_N^T(t-1)} \end{bmatrix}^{\mathrm{T}} \triangleq y_t^{-} \ , \\
\mathrm{Y_t^{+}} &= \begin{bmatrix} \mathrm{y_N^T(t)} & \mathrm{y_N^T(t+1)} & \cdots & \mathrm{y_N^T(t+k-1)} \end{bmatrix}^{\mathrm{T}} \triangleq y_t^{+}
\end{aligned}
\tag{2.20}
$$

correspond to y_t^{-} and y_t^{+}; the representations of u_t^{-} and u_t^{+} are formed likewise. Note the different notations for stochastic values, e.g., $y(t) \in L^2(\Omega, \mathcal{A}, P)$, and elements of sample functions, e.g., $\mathrm{y(t)} \in \mathbb{R}^m$, or tail matrices, e.g., $\mathrm{y_N(t)} \in \mathbb{R}^{m \times N}$. Governed by the ergodic theorem, an estimation of ensemble covariance of $y(t+l)$ and $y(t)$ is given in terms of the sample covariance by

$$
\hat{\Lambda}_{yy}(l) = \frac{1}{N} \sum_{k=0}^{N-1} \mathrm{y(t+l+k)y^T(t+k)} = \frac{1}{N} \mathrm{y_N(t+l)y_N^T(t)} \ .
\tag{2.21}
$$

Herein, the meaning of the tail matrices as representations of $y(t + l)$ and $y(t)$ becomes evident. For $N \to \infty$ the sample covariance converges to the ensemble covariance. Hence, the respective numerical equivalents of the data spaces $\mathcal{Y}_t^-$ or $\mathcal{Y}_t^+$ are then given by the row spaces of the above defined matrices Y_t^- or Y_t^+. The respective inner product is given by

$$\langle y_N(t + l), y_N(t) \rangle = \text{trace}\left(\frac{1}{N} y_N(t + l) y_N^T(t)\right) . \tag{2.22}$$

The connection between the spaces $\mathcal{Y}_t^-$ or $\mathcal{Y}_t^+$ with $\text{Im}_{\text{Row}}\{Y_t^-\}$ or $\text{Im}_{\text{Row}}\{Y_t^+\}$ is established by means of the ensemble covariance (cf. (2.9)) and the sample covariance. The spaces spanned by y_t^- and y_t^+ and the row spaces of Y_t^- and Y_t^+ are isometrically isomorph, if $N \to \infty$ (Lindquist and Picci, 1996a and Lindquist and Picci, 2015, p. 512).

2.2.3 Projections
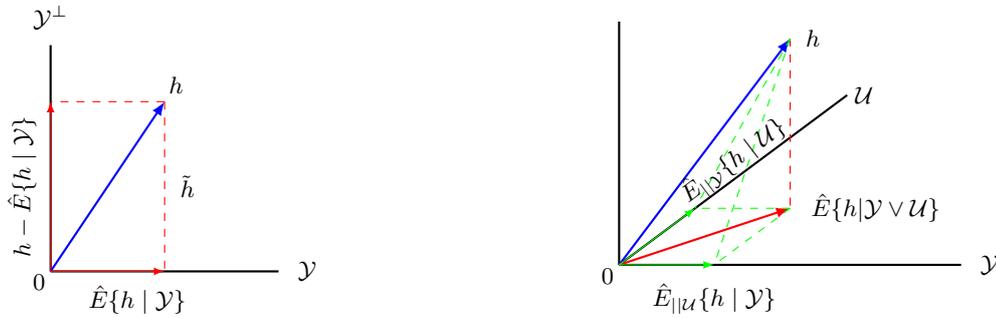
Figure 2.2: Illustrations of the orthogonal projection (left) and the oblique (or parallel) projection (right) (Katayama, 2005)

Operations which play a major role in terms the data decomposition necessary for the identification are the orthogonal and oblique projections. For the discussion of these projections, let

$$\mathcal{H} = \mathcal{U}_t^- \vee \mathcal{U}_t^+ \vee \mathcal{Y}_t^- \vee \mathcal{Y}_t^+$$

be the ambient space and

$$\mathcal{Y} = \overline{\text{span}}\{y(i) | i = 1, \cdots N\} \subset \mathcal{H}$$

an arbitrary subspace of $\mathcal{H}$.

Lemma 2.1. *(Katayama, 2005) Let $\mathcal{H}$ and $\mathcal{Y}$ be defined as above and $\tilde{h}$ be an element in $\mathcal{H}$. Then $\tilde{h}$ is orthogonal to $\mathcal{Y}$ if and only if the conditions*

$$E\{\tilde{h}\} = 0 , \qquad E\{\tilde{h} y^T(i)\} = 0 , i = 1, ..., N \tag{2.23}$$

hold.

The conditions of this lemma are equal to the conditions for the solution of the least-squares problem

$$\min_{\hat{y} \in \mathcal{Y}} \|h - \hat{y}\|_{\mathcal{H}}^2 , \tag{2.24}$$

where $h \in \mathcal{H}$. The solution is obviously given if $h - \hat{y} \perp \mathcal{Y}$, i.e., if the remainder $\tilde{h} = h - \hat{y}$ of the estimation of h by $\hat{y}$ is orthogonal to $\hat{y}$. In terms of the inner product as defined by (2.9), this would be expressed by

$$\langle \tilde{h}, \hat{y} \rangle_{\mathcal{H}} = \mathrm{trace}\left(E\{\tilde{h}\hat{y}^{\mathrm{T}}\}\right) = 0 \, , \tag{2.25}$$

which is equal to the second condition of (2.23). As $\tilde{h}$ minimizes (2.24) and by (2.10),

$$\|h - \hat{y}\|_{\mathcal{H}}^2 = \|\tilde{h}\|_{\mathcal{H}}^2 = \mathrm{trace}\left(E\left\{\tilde{h}\tilde{h}^{\mathrm{T}}\right\}\right) \tag{2.26}$$

is the smallest achievable value, and the respective estimation of $\hat{y}$ yields hence a minimum variance estimate. The calculation of $\hat{y}$ is given by an orthogonal projection of h onto $\mathcal{Y}$.

Lemma 2.2. *(Katayama, 2005) The orthogonal projection of h onto $\mathcal{Y} = \overline{\mathrm{span}}\{y(i)|i = 1, \cdots N\}$, denoted by $\hat{E}\{h|\mathcal{Y}\}$, is defined by the conditional expectation as*

$$\hat{E}\{h|\mathcal{Y}\} = E\{h|y(i),\ i = 1, \cdots N\} = \Sigma_{hy}\Sigma_{yy}^{-1}y \, , \tag{2.27}$$

where

$$y = \left[y^{\mathrm{T}}(1) \ \cdots \ y^{\mathrm{T}}(N)\right]^{\mathrm{T}} \, .$$

Strictly mathematically written, the conditional expectation of (2.27) is given by

$$E\{h|y(i),\ i = 1, \cdots N\} = E\{a|\mathcal{F}_y\} \, ,$$

where $\mathcal{F}_y = \sigma\{y(i);\ i = 1, \cdots, N\}$ is the σ-algebra generated by $y(i),\ i = 1, \cdots, N$. This σ-algebra contains all information conveyed by $y(i),\ i = 1, \cdots, N$ (Katayama, 2005, p. 114). Thus, (2.27) further implies the aforementioned interpretation of the data spaces as structures holding all information contained in the stochastic processes. The respective orthogonal projection onto the orthogonal complement $\mathcal{Y}^{\perp}$ is

$$\hat{E}\left\{h\left|\mathcal{Y}^{\perp}\right.\right\} = h - \hat{E}\left\{h\left|\mathcal{Y}\right.\right\} = h - \Sigma_{hy}\Sigma_{yy}^{-1}y \, . \tag{2.28}$$

In left part of Figure 2.2, these two projections are shown. In terms of spaces, the orthogonal complement of $\mathcal{Y}$ in $\mathcal{H}$ is also expressed by

$$\mathcal{H} \ominus \mathcal{Y} = \mathcal{H} - \hat{E}\{\mathcal{H} \mid \mathcal{Y}\} = \mathcal{Y}^{\perp} \, . \tag{2.29}$$

This is the inverse operation of the orthogonal direct vector sum

$$\mathcal{H} = \mathcal{Y} \oplus \mathcal{Y}^{\perp} \, .$$

However, the operator $\ominus$ does not commute, i.e.,

$$\mathcal{H} \ominus \mathcal{Y} \neq \mathcal{Y} \ominus \mathcal{H} \, .$$

Equal to $\mathcal{Y}$, let a space $\mathcal{U}$ be given by

$$\mathcal{U} = \overline{\mathrm{span}}\{u(k)|k = 1, \cdots M\} \subset \mathcal{H} \, .$$

Based on the orthogonal projection, the more general oblique projection is defined according to the following lemma.

Lemma 2.3. *(Katayama, 2005) Given two spaces $\mathcal{Y} = \overline{\text{span}}\{y(i)|i = 1, \cdots N\}$ and $\mathcal{U} = \overline{\text{span}}\{u(k)|k = 1, \cdots M\}$ (with $\mathcal{U} \cap \mathcal{Y} = \{0\}^2$). Then*

$$\hat{E}\{h|\mathcal{Y} \vee \mathcal{U}\} = \hat{E}\{h|\mathcal{Y} + \mathcal{U}\} = \hat{E}_{||\mathcal{U}}\{h|\mathcal{Y}\} + \hat{E}_{||\mathcal{Y}}\{h|\mathcal{U}\} , \tag{2.30}$$

where

$$\hat{E}_{||\mathcal{U}}\{h|\mathcal{Y}\} = \Sigma_{hy|u}\Sigma_{yy|u}^{-1}y , \quad \hat{E}_{||\mathcal{Y}}\{h|\mathcal{U}\} = \Sigma_{hu|y}\Sigma_{uu|y}^{-1}u \tag{2.31}$$

are called oblique projection of h onto $\mathcal{Y}$ along $\mathcal{U}$ and oblique projection of h onto $\mathcal{U}$ along $\mathcal{Y}$. The matrices $\Sigma_{hy|u}$, $\Sigma_{yy|u}$, $\Sigma_{hu|y}$, and $\Sigma_{uu|y}$ are the conditional covariance matrices *defined by*

$$\begin{aligned}
\Sigma_{hy|u} &= E\left\{\hat{E}\left\{h\left|\mathcal{U}^\perp\right.\right\}\hat{E}\left\{y\left|\mathcal{U}^\perp\right.\right\}^{\mathrm{T}}\right\} = \Sigma_{hy} - \Sigma_{hu}\Sigma_{uu}^{-1}\Sigma_{uy} , \\[2mm]
\Sigma_{yy|u} &= E\left\{\hat{E}\left\{y\left|\mathcal{U}^\perp\right.\right\}\hat{E}\left\{y\left|\mathcal{U}^\perp\right.\right\}^{\mathrm{T}}\right\} = \Sigma_{yy} - \Sigma_{yu}\Sigma_{uu}^{-1}\Sigma_{uy} , \\[2mm]
\Sigma_{hu|y} &= E\left\{\hat{E}\left\{h\left|\mathcal{Y}^\perp\right.\right\}\hat{E}\left\{u\left|\mathcal{Y}^\perp\right.\right\}^{\mathrm{T}}\right\} = \Sigma_{hu} - \Sigma_{hy}\Sigma_{yy}^{-1}\Sigma_{yu} , \\[2mm]
\Sigma_{uu|y} &= E\left\{\hat{E}\left\{u\left|\mathcal{Y}^\perp\right.\right\}\hat{E}\left\{u\left|\mathcal{Y}^\perp\right.\right\}^{\mathrm{T}}\right\} = \Sigma_{uu} - \Sigma_{uy}\Sigma_{yy}^{-1}\Sigma_{yu} ,
\end{aligned} \tag{2.32}$$

where

$$y = \begin{bmatrix} y^{\mathrm{T}}(1) \cdots y^{\mathrm{T}}(N) \end{bmatrix}^{\mathrm{T}} , \qquad u = \begin{bmatrix} u^{\mathrm{T}}(1) \cdots u^{\mathrm{T}}(M) \end{bmatrix}^{\mathrm{T}} .$$

If the spaces $\mathcal{U}$ and $\mathcal{Y}$ are orthogonal to each other the oblique projections of (2.31) become orthogonal projections. This follows, as $\Sigma_{yu} = 0$ in the case of orthogonality. The illustration of the oblique projections in the right part of Figure 2.2 shows the decomposition of one orthogonal projection onto the joint space of $\mathcal{Y}$ and $\mathcal{U}$ into two oblique projections onto the respective spaces $\mathcal{Y}$ and $\mathcal{U}$. In fact, the condition $\mathcal{U} \cap \mathcal{Y} = \{0\}$ becomes also evident. In terms of identifications, the decomposition of the future outputs into a component given by the past inputs and outputs (cf. predictor representation of a system or Kalman filter) and a component given by the future inputs follows from the decomposition of an orthogonal projection into two oblique projections. From the results of these projections, the states and system matrices are subsequently estimated.

Based on the above defined orthogonal projection, the concept of *conditional orthogonality* will be introduced. This property of two vectors or spaces with respect to a third is crucial for the definition of a space, which will contain the state estimate. In Figure 2.3, the idea of conditional orthogonality is illustrated.

Definition 2.1. *(Katayama, 2005) Suppose that $\hat{y} \in \mathcal{Y}$ and $\hat{u} \in \mathcal{U}$ satisfy the orthogonality condition*

$$E\left\{\left(\hat{y} - \hat{E}\{\hat{y}|\mathcal{Z}\}\right)\left(\hat{u} - \hat{E}\{\hat{u}|\mathcal{Z}\}\right)^{\mathrm{T}}\right\} = 0 , \qquad \mathcal{Z} \subset \mathcal{H} . \tag{2.33}$$

---

[2]For this choice of spaces, the condition is fulfilled if for all k the condition $k > i$ holds, and the system operates in open-loop. That is, there is no causal connection between the $y(i)$ and $u(k)$.
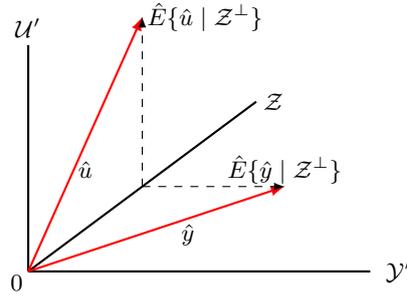
Figure 2.3: Conditional orthogonality of two vectors $\hat{y} \in \mathcal{Y} = \mathcal{Y}' \vee \mathcal{Z}$ and $\hat{u} \in \mathcal{U} = \mathcal{U}' \vee \mathcal{Z}$ given a vector space $\mathcal{Z}$ (Katayama, 2005)

Then, $\hat{y}$ and $\hat{u}$ are conditionally orthogonal with respect to $\mathcal{Z}$. If this orthogonality condition furthermore holds for all $\hat{y} \in \mathcal{Y}$ and $\hat{u} \in \mathcal{U}$, $\mathcal{Y}$ and $\mathcal{U}$ are conditionally orthogonal with respect to $\mathcal{Z}$, which is denoted by

$$\mathcal{Y} \perp \mathcal{U} | \mathcal{Z} . \tag{2.34}$$

In terms of Figure 2.3, the two spaces are formed as $\mathcal{Y} = \mathcal{Y}' \vee \mathcal{Z}$ and $\mathcal{U} = \mathcal{U}' \vee \mathcal{Z}$ so that

$$\hat{E}\left\{ \mathcal{Y} \,\middle|\, \mathcal{Z}^{\perp} \right\} = \mathcal{Y}' \perp \mathcal{U}' = \hat{E}\left\{ \mathcal{U} \,\middle|\, \mathcal{Z}^{\perp} \right\} .$$

2.3 Assumptions and identification problem

In addition to considering the system to be linear and time-invariant, there are additional assumptions which set the scope for the derivation and application of identification methods. These assumptions need to be met in order to accurately identify a system and are given in the following.

Assumption 2.1. *(e.g., Jansson and Wahlberg, 1998; Knudsen, 2001; Qin, 2006)*

A1 The eigenvalues of $A - KC$ are strictly inside the unit circle, i.e., the stochastic subsystem is of minimal phase.

A2 The system is minimal, i.e., (A, C) is observable and $\left(A, \begin{bmatrix} B & K \end{bmatrix} \right)$ is reachable (or controllable)[3].

A3 The innovation process e or the state noise w and measurement noise v are zero-mean, white-noise processes with bounded second moments, i.e.,[4]

$$E\{e(t)e(s)\} = \Sigma_{ee}\delta_{ts} < \infty \tag{2.35a}$$

---

[3] This reachability (or controllability) condition allows the identification of Box-Jenkins as well as ARX/ARMAX systems, as for the Box-Jenkins case the reachability (or controllability) of (A, B) excludes the existence of states that are purely driven by the stochastic subsystem (Knudsen, 2001).

[4] δ_{ts} denotes the Kronecker delta defined by

$$\delta_{ts} = \begin{cases} 1, & t = s \\ 0, & t \neq s \end{cases} .$$

$$E\left\{\begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w^{\mathrm{T}}(s) & v^{\mathrm{T}}(s) \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S^{\mathrm{T}} & R \end{bmatrix} \delta_{ts} < \infty \qquad (2.35b)$$

A4 The richness condition

$$(\mathcal{Y}_t^+ \vee \mathcal{U}_t^+) \cap (\mathcal{Y}_t^- \vee \mathcal{U}_t^-) = \{0\} \qquad (2.36)$$

is fulfilled, which is the same as the assumption that the spectrum of the joint process (y, u) is bounded away from zero (Lindquist and Picci, 2015, p. 695). This includes the requirement of a persistently exciting input (see Ljung, 2009), which in this case is persistently exciting of any required order as

$$\mathcal{U}_t^- \cap \mathcal{U}_t^+ = \{0\} \, . \qquad (2.37)$$

This is a general condition necessary for the unique identification of the model parameter and in particular for subspace identification to accurately decompose the data spaces (see condition for the decomposition of an orthogonal projection into oblique projections in Lemma 2.3 and the discussion following that lemma).

A5 In the open-loop case, the input u is orthogonal to the innovation process e or the state noise w and measurement noise v. In terms of the closed-loop case, this assumption does not apply.

Regarding the numerical representation of the data spaces, which are then also of finite dimension, a condition equivalent to (2.37) is given in Moonen, De Moor, et al. (1989). Taking the finite dimensionality into account,

$$\mathrm{Im}_{\mathrm{Row}}(\mathrm{x_N}(\mathrm{t_0})) \cap \mathrm{Im}_{\mathrm{Row}}(\mathrm{U_t^-}) \cap \mathrm{Im}_{\mathrm{Row}}(\mathrm{U_t^+}) = \{0\}$$

is required to hold. Furthermore, if the input signal has only a finite-order persistence of excitation, it needs to fulfill certain conditions (which sometimes also depend on the identification algorithm). These requirements are analyzed in Jansson and Wahlberg (1998), Chui and Maciejowski (2005) and Willems, Rapisarda, et al. (2005). General requirements as well as conditions for the horizons as given in Chui and Maciejowski (2005) are summarized (and simplified) by the following proposition.

Proposition 2.1. *(Chui and Maciejowski, 2005) Let the length of the future and past horizons be denoted by k_p and k_f and the system be stable and reachable. Then, the input signal must be at least persistently exciting of order $k_f + k_p + n$, where*

- *k_p is not smaller than 3 times the order of the system ($k_p \geq 3n$),*

- *$k_f - 1$ is not smaller than the system's observability index.*

As analyzed in Chiuso and Picci (2004c) and Chiuso and Picci (2004d), it is also necessary that the spectrum of the input should be as flat as possible and should not posses zeros that are equal to the system's eigenvalues (see Chiuso and Picci, 2004d, Section 4). Furthermore, the input is required to not exhibit a spectral density smaller than the spectral density of the stochastic system (Chiuso and Picci, 1999, p. 241). If this requirement is violated the variance of the results in this frequency range increases (see also Section 3 in Chiuso and Picci, 2004b).

With these conditions fulfilled, the identification tasks can be formulated as follows.

Problem 2.1. *Given the input-output data* $\{y(t), u(t), t = 0, 1, ..., T\}$ *of a system, determine* *A, B, C, D, and* $\Sigma_{v,w}$ *or* K, *and* Σ_{ee} *of the linear time-invariant system given by* (2.1) *or* (2.4) *up to a global similarity transform.*

The value of T follows from the definition of the tail matrices as $T \geq k_p + k_f + N - 2$.

2.4 Coordinate-free framework

In this section, the underlying theoretical framework of subspace identification is reviewed and summarized. This framework was introduced and discussed in Chiuso and Picci (2003) and Lindquist and Picci (1996b, 2015). Instead of working with a generic model, the identification procedure is based on the relations between the above introduced spaces. Due to this avoidance of a model, which would fix the identification to a certain coordinate system or basis, this framework is called *coordinate-free framework*. The goal is to extract a subspace of the past $\mathcal{P}_t^-$ containing all information of the past needed for the prediction of the future. This subspace is hence called *predictor space*. Once it is extracted from $\mathcal{P}_t^-$, the model is essentially obtained (Akaike, 1974, p. 669, Lindquist and Picci, 1996b). The remainder of the identification process is then only concerned with the determination of a basis of this space and the subsequent realization of the system model. The idea of the predictor space, i.e., some subspace of the past containing the necessary information for prediction, was first discussed by Akaike in Akaike (1974) for stochastic systems.

These stochastic systems, i.e., systems without exogenous inputs, will be the starting point of the following discussion, as the fundamental ideas of the definition of the space of the state, or the state itself, become more evident in this case. Following these introductory explanations, the discussion moves on to systems with exogenous inputs. In fact, the definition of the predictor space for systems with exogenous inputs is a mere extension of the results given for stochastic systems. The difference concerns only the then arising conditional dependency on the (future) inputs. The summary given in this section is an excerpt of the thorough discussion of linear stochastic systems in Lindquist and Picci (2015).

2.4.1 Stochastic systems

Let a stochastic system be given by

$$x(t+1) = Ax(t) + Bw(t) , \tag{2.38a}$$
$$y(t) = Cx(t) + Dw(t) , \tag{2.38b}$$

where w is a assumed to be a second-order white-noise process with unity variance and

$$\begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} Q & S \\ S^{\mathrm{T}} & R \end{bmatrix} . \tag{2.39}$$

This stochastic system, in terms of the output also known as time-series model, should be given a realization in terms of

$$x(t+1) = Ax(t) + Ke(t) ,$$
$$y(t) = Cx(t) + e(t) .$$

This is essentially the special case of the filtering problem of Wiener and Kolmogorov (see, e.g., Chapter 4.1 in Lindquist and Picci, 2015, pp. 103–110). In terms of the problem at hand, i.e., finding a causal estimation of a second stochastic process x based on the knowledge of y, the filtering problem reads[5]

$$\hat{x}_-(t) := E\{x(t) \mid \mathcal{F}_t^y\} . \tag{2.41}$$

As $\mathcal{F}_t^y = \sigma\{y(t-1), y(t-2), \cdots\}$ is the σ-algebra generated by the past up to $t-1$, i.e., the present output $y(t)$ is not included, $\hat{x}_-(t)$ is the one-step ahead predictor of $x(t)$. The respective acausal estimation problem is given by

$$\hat{x}(t) := E\{x(t) \mid \mathcal{F}^y\} ,$$

where $\mathcal{F}_t^y = \sigma\{\cdots, y(t-1), y(t), y(t+1)\cdots\}$ is the σ-algebra generated by the entire process y. The former problem was given a solution by the well-known Kalman filter in Kalman (1960). This yields two types of models for the process y; the forward innovation model (based on the above one-step ahead predictor) and the backward innovation model. The (stationary) forward innovation model (see Katayama, 2005, pp. 128–130 or Lindquist and Picci, 2015, pp. 175–179) is

$$\hat{x}_-(t+1) = A\hat{x}_-(t) + Ke_-(t) , \tag{2.42a}$$
$$y(t) = C\hat{x}_-(t) + e_-(t) , \tag{2.42b}$$

where the innovation is defined by

$$e_-(t) = y(t) - \hat{E}\{y(t)|\mathcal{Y}_t^-\} .$$

The stationary Kalman gain is given by

$$K = \left(\overline{C}^{\mathrm{T}} - A\Sigma_{\hat{x}\hat{x}}C^{\mathrm{T}}\right)\left(\Lambda(0) - C\Sigma_{\hat{x}\hat{x}}C^{\mathrm{T}}\right)^{-1} ,$$

where $\Sigma_{\hat{x}\hat{x}} = E\{\hat{x}_-(t)\hat{x}_-^{\mathrm{T}}(t)\}$ satisfies the algebraic Ricatti equation

$$\Sigma_{\hat{x}\hat{x}} = A\Sigma_{\hat{x}\hat{x}}A^{\mathrm{T}} + \left(\overline{C}^{\mathrm{T}} - A\Sigma_{\hat{x}\hat{x}}C^{\mathrm{T}}\right)\left(\Lambda(0) - C\Sigma_{\hat{x}\hat{x}}C^{\mathrm{T}}\right)^{-1}\left(\overline{C} - C\Sigma_{\hat{x}\hat{x}}A^{\mathrm{T}}\right)$$

and

$$\overline{C} = E\{y(t)x^{\mathrm{T}}(t+1)\} = C\Pi A^{\mathrm{T}} + S^{\mathrm{T}} ,$$
$$\Lambda(0) = E\{y(t)y^{\mathrm{T}}(t)\} = C\Pi C^{\mathrm{T}} + R ,$$
$$\Pi = E\{x(t)x^{\mathrm{T}}(t)\} .$$

---

[5]Note that in this context the following two assumptions need to be fulfilled (Lindquist and Picci, 2015, p. 105):

- The processes x and y are jointly stationary and their distribution function is absolutely continuous with known spectral density matrix.

- The observations are made since $t_0 = -\infty$.

The (stationary) backward innovation model (see Katayama, 2005, pp. 131–134 or Lindquist and Picci, 2015, pp. 179–183) is

$$\hat{x}_+(t-1) = A^{\mathrm{T}}\hat{x}_+(t) + \overline{K}^{\mathrm{T}}e_+(t) \, , \qquad (2.43\mathrm{a})$$

$$y(t) = \overline{C}\hat{x}_+(t) + e_+(t) \, , \qquad (2.43\mathrm{b})$$

where the backward innovation is defined by

$$e_+(t) = y(t) - \hat{E}\{y(t)|\mathcal{Y}_{t+1}^+\} \, .$$

The stationary backward Kalman gain is given by

$$\overline{K}^{\mathrm{T}} = \left(C^{\mathrm{T}} - A^{\mathrm{T}}\overline{\Sigma}_{\hat{x}\hat{x}}\overline{C}^{\mathrm{T}}\right)\left(\Lambda(0) - \overline{C}\Sigma_{\hat{x}\hat{x}}\overline{C}^{\mathrm{T}}\right)^{-1} \, ,$$

where $\overline{\Sigma}_{\hat{x}\hat{x}} = E\{\hat{x}_+(t)\hat{x}_+^{\mathrm{T}}(t)\}$ satisfies the algebraic Ricatti equation

$$\overline{\Sigma}_{\hat{x}\hat{x}} = A^{\mathrm{T}}\overline{\Sigma}_{\hat{x}\hat{x}}A + \left(C^{\mathrm{T}} - A^{\mathrm{T}}\overline{\Sigma}_{\hat{x}\hat{x}}\overline{C}^{\mathrm{T}}\right)\left(\Lambda(0) - \overline{C}\Sigma_{\hat{x}\hat{x}}\overline{C}^{\mathrm{T}}\right)^{-1}\left(C - \overline{C}\Sigma_{\hat{x}\hat{x}}A\right) \, .$$

Based on the associated processes y, $\hat{x}_-$, and $\hat{x}_+$ of the stochastic system (2.38), let

$$\begin{aligned}
\mathcal{Y}_t^- &= \overline{\mathrm{span}}\{y(s)|s < t\} \, , & \mathcal{X}_t^- &= \overline{\mathrm{span}}\{\overline{x}(s)|s \leq t\} \, , \\
\mathcal{Y}_t^+ &= \overline{\mathrm{span}}\{y(s)|s \geq t\} \, , & \mathcal{X}_t^+ &= \overline{\mathrm{span}}\{\overline{x}(s)|s \geq t\} \, ,
\end{aligned} \qquad (2.44)$$

be the respective past and future spaces of the output and the joint state. Regarding the spaces $\mathcal{X}_t^-$ and $\mathcal{X}_t^+$ of this joint state, it should be noted that they

1. are defined as overlapping subspaces, i.e., both contain the element $\overline{x}(t)$, and

2. contain both $\hat{x}_-$ and $\hat{x}_+$, i.e., the joint state of both the forward *and* backward innovation model.

Furthermore, the ambient space is given by

$$\mathcal{H} = \mathcal{Y}_t^- \vee \mathcal{X}_t^- \vee \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ \, . \qquad (2.45)$$

Perpendicular intersection

To define a predictor space $\mathcal{X}_t$ and analyze its properties, the concept of *perpendicular intersection* of two subspaces is needed. The concept of perpendicular intersection is based on the conditional orthogonality of two subspaces, e.g., $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$, and the question of how to enlarge these spaces while retaining conditional orthogonality. The relations stated in this paragraph will become important for the definitions of observability and constructibility of a predictor space. The following discussion introduces a basic mathematical concept and is not an analysis of system entities. That is, the denotation of subspaces $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$ does not carry any deeper meaning than just "some subspaces" and $\mathcal{X}_t$ is any small or large subspace of the ambient space $\mathcal{H}$.

Lemma 2.4. *(Lindquist and Picci, 2015) Assuming that the three subspaces $\mathcal{Y}_t^-$, $\mathcal{Y}_t^+$, and $\mathcal{X}_t$ fulfill*

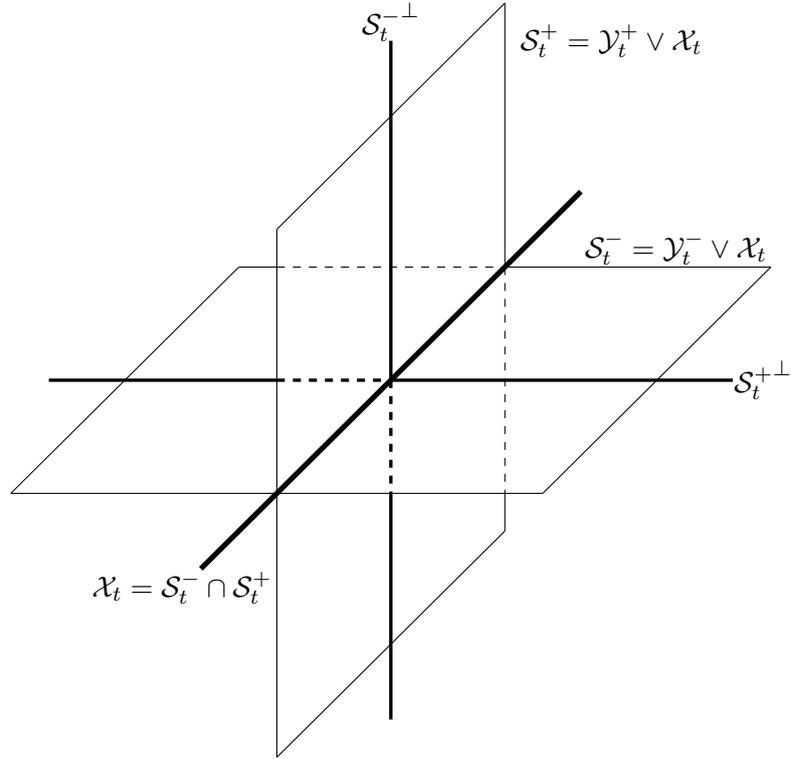$$\mathcal{Y}_t^- \perp \mathcal{Y}_t^+|\mathcal{X}_t \, ,$$

Figure 2.4: Perpendicular intersection (Lindquist and Picci, 2015)

then

$$(i) \quad \mathcal{Y}_t^- \cap \mathcal{Y}_t^+ \subset \mathcal{X}_t \tag{2.46a}$$

$$(ii) \quad (\mathcal{Y}_t^- \vee \mathcal{X}_t) \perp (\mathcal{Y}_t^+ \vee \mathcal{X}_t)|\mathcal{X}_t \tag{2.46b}$$

$$(iii) \quad (\mathcal{Y}_t^- \vee \mathcal{X}_t) \cap (\mathcal{Y}_t^+ \vee \mathcal{X}_t) = \mathcal{X}_t \tag{2.46c}$$

hold.

The points of this lemma give instructions on how the subspaces $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$ can be enlarged. Setting[6]

$$\mathcal{S}_t^- := \mathcal{Y}_t^- \vee \mathcal{X}_t \,, \qquad \mathcal{S}_t^+ := \mathcal{Y}_t^+ \vee \mathcal{X}_t \,, \tag{2.47}$$

these two subspaces behave as follows.

Proposition 2.2. *(Lindquist and Picci, 2015) The following conditions are equivalent.*

$$(i) \quad \mathcal{S}_t^- \perp \mathcal{S}_t^+ | \mathcal{S}_t^- \cap \mathcal{S}_t^+ \tag{2.48a}$$

$$(ii) \quad \hat{E}\{\mathcal{S}_t^+ | \mathcal{S}_t^-\} = \mathcal{S}_t^- \cap \mathcal{S}_t^+ \tag{2.48b}$$

$$(iii) \quad \hat{E}\{\mathcal{S}_t^- | \mathcal{S}_t^+\} = \mathcal{S}_t^- \cap \mathcal{S}_t^+ \tag{2.48c}$$

$$(iv) \quad \hat{E}\{\mathcal{S}_t^+ | \mathcal{S}_t^-\} = \hat{E}\{\mathcal{S}_t^- | \mathcal{S}_t^+\} \tag{2.48d}$$

Subspaces fulfilling these conditions are called perpendicular intersecting. See Figure 2.4 for an illustration of these conditions.

---

[6]If these definitions would be interpreted in terms of system entities, the state, given by $\mathcal{X}_t$, would be internal (see next paragraph).

Theorem 2.1. *(Lindquist and Picci, 2015) Let the ambient data space $\mathcal{H}$ be composed of $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ as $\mathcal{S}_t^- \vee \mathcal{S}_t^+ = \mathcal{H}$. Then, the following conditions are equivalent.*

$$(i) \quad \mathcal{S}_t^- \text{ and } \mathcal{S}_t^+ \text{ intersect perpendicular} \tag{2.49a}$$

$$(ii) \quad (\mathcal{S}_t^+)^\perp \subset \mathcal{S}_t^- \text{ or, equivalently } (\mathcal{S}_t^-)^\perp \subset \mathcal{S}_t^+ \tag{2.49b}$$

$$(iii) \quad \mathcal{H} = (\mathcal{S}_t^-)^\perp \oplus (\mathcal{S}_t^- \cap \mathcal{S}_t^-) \oplus (\mathcal{S}_t^+)^\perp \tag{2.49c}$$

From the above discussion, the previously raised question is answered.

Theorem 2.2. *(Lindquist and Picci, 2015) Let two subspaces $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$ be defined so that $\mathcal{Y}_t^- \vee \mathcal{Y}_t^+ = \mathcal{H}$, $\mathcal{Y}_t^- \perp \mathcal{Y}_t^+ | \mathcal{X}_t$ and $\mathcal{Y}_t^- \subset \mathcal{S}_t^-$, $\mathcal{Y}_t^- \subset \mathcal{S}_t^+$ hold. Then $\mathcal{S}_t^- \perp \mathcal{S}_t^+ | \mathcal{X}_t$ holds if and only if*

$$\mathcal{S}_t^- \subset \mathcal{Y}_t^- \vee \mathcal{X}_t , \qquad \mathcal{S}_t^+ \subset \mathcal{Y}_t^+ \vee \mathcal{X}_t . \tag{2.50}$$

This theorem states that $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$ can be enlarged by any subspace of $\mathcal{X}_t$ without loosing conditional orthogonality. In the special case that (2.47) holds, the two spaces $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ intersect perpendicular

Definition of Markovian splitting subspace

As it will turn out by the following summary, the above equations already have a deeper meaning than just the illustration of the perpendicular intersection. Now, the denotations are coupled with the actual entities of the system.

Definition 2.2. *(Lindquist and Picci, 2015) Let $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$ be the past and future spaces of y, and $\mathcal{X}_t$ an arbitrary subspace of the ambient space $\mathcal{H}$ defined by (2.45), i.e., $\mathcal{X}_t \subset \mathcal{H}$. Then $\mathcal{X}_t$ is called* splitting subspace *if it has the property*

$$\mathcal{Y}_t^- \perp \mathcal{Y}_t^+ \mid \mathcal{X}_t \tag{2.51}$$

i.e., if the past and future become conditionally orthogonal given $\mathcal{X}_t$. It is furthermore a Markovian splitting subspace *if*

$$(\mathcal{Y}_t^- \vee \mathcal{X}_t^-) \perp (\mathcal{Y}_t^+ \vee \mathcal{X}_t^+) \mid \mathcal{X}_t , \tag{2.52}$$

which includes the Markov property (where $\mathcal{X}_t^-$ and $\mathcal{X}_t^+$ are the respective past and future of $\mathcal{X}_t$)

$$\mathcal{X}_t^- \perp \mathcal{X}_t^+ \mid \mathcal{X}_t . \tag{2.53}$$

The Markov property can be also stated by

$$E\{\overline{x}_{t+1} \mid \overline{x}_s; s \leq t\} = E\{\overline{x}_{t+1} \mid \overline{x}_t\} .$$

The splitting property of $\mathcal{X}_t$ is also given by the following relations.

Lemma 2.5. *(Lindquist and Picci, 2015) Let $\mathcal{X}_t$ be a (Markovian) splitting subspace for $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$, then*

$$\hat{E}\{\lambda | \mathcal{Y}_t^- \vee \mathcal{X}_t\} = \hat{E}\{\lambda | \mathcal{X}_t\} \, \forall \, \lambda \in \mathcal{Y}_t^+ , \tag{2.54a}$$

$$\hat{E}\{\lambda | \mathcal{Y}_t^+ \vee \mathcal{X}_t\} = \hat{E}\{\lambda | \mathcal{X}_t\} \, \forall \, \lambda \in \mathcal{Y}_t^- . \tag{2.54b}$$

For a proof, see Proposition 2.4.2 in Lindquist and Picci (2015). This lemma states that $\mathcal{X}_t$ contains all necessary information of the past to predict the future and all necessary information of the future to predict past, i.e., it serves as a "memory" (Lindquist and Picci, 2015, p. 220). A Markovian splitting subspace is hence synonymic with a predictor space.

The definition of a splitting subspace can be given in terms of perpendicular intersection.

Theorem 2.3. *(Lindquist and Picci, 2015) Let $(\mathcal{S}_t^-, \mathcal{S}_t^+)$ be a pair of perpendicular intersecting subspaces with $\mathcal{Y}_t^- \subset \mathcal{S}_t^-$ and $\mathcal{Y}_t^+ \subset \mathcal{S}_t^+$. Then, $\mathcal{X}_t \subset \mathcal{H}$ is a splitting subspace if and only if*

$$\mathcal{X}_t = \mathcal{S}_t^- \cap \mathcal{S}_t^+ \tag{2.55}$$

and

$$\mathcal{X}_t = \hat{E}\{\mathcal{S}_t^- | \mathcal{S}_t^+\} = \hat{E}\{\mathcal{S}_t^+ | \mathcal{S}_t^-\} . \tag{2.56}$$

Furthermore,

$$\hat{E}\{\lambda | \mathcal{S}_t^-\} = \hat{E}\{\lambda | \mathcal{X}_t\} \; \forall \; \lambda \in \mathcal{S}_t^+ , \tag{2.57a}$$

$$\hat{E}\{\lambda | \mathcal{S}_t^+\} = \hat{E}\{\lambda | \mathcal{X}_t\} \; \forall \; \lambda \in \mathcal{S}_t^- . \tag{2.57b}$$

Based thereon the Markov property is rephrased by using $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ as extensions of the past and future spaces of y.

Theorem 2.4. *(Lindquist and Picci, 2015) The subspace $\mathcal{X}_t$ is a Markovian splitting subspace if and only if it can be given a* scattering pair $(\mathcal{S}_t^-, \mathcal{S}_t^+)$ *satisfying*

$$\mathcal{S}_{t-1}^- \subset \mathcal{S}_t^- , \qquad \mathcal{S}_{t+1}^+ \subset \mathcal{S}_t^+ . \tag{2.58}$$

A unique scattering pair of $\mathcal{X}_t$ is given by

$$\mathcal{S}_t^- := \mathcal{Y}_t^- \vee \mathcal{X}_t^- , \qquad \mathcal{S}_t^+ := \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ . \tag{2.59}$$

The pair $(\mathcal{S}_t^-, \mathcal{S}_t^+)$ is called scattering pair by reason of (2.55). Furthermore, the ambient space is given by

$$\mathcal{H} = \mathcal{S}_t^- \vee \mathcal{S}_t^+ . \tag{2.60}$$

As the properties of a Markovian splitting subspace are now given, the last remaining question is regarding the calculation or extraction of the respective Markovian splitting subspaces contained in $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$. For that matter, a result of the conditional orthogonality is needed.

Lemma 2.6. *(Lindquist and Picci, 2015) For any subspaces $\mathcal{A}$ and $\mathcal{B}$,*

$$\mathcal{A} \perp \mathcal{B} | \hat{E}\{\mathcal{B} | \mathcal{A}\} . \tag{2.61}$$

Now, the respective Markovian splitting subspaces are the results of rather simple calculations.

Proposition 2.3. *(Lindquist and Picci, 2015) The subspaces*

$$\mathcal{X}_t^{+/-} = \overline{\mathrm{span}}\{x_-(t)\} = \hat{E}\{\mathcal{Y}_t^+ | \mathcal{Y}_t^-\} , \qquad \mathcal{X}_t^{-/+} = \overline{\mathrm{span}}\{x_+(t)\} = \hat{E}\{\mathcal{Y}_t^- | \mathcal{Y}_t^+\} \tag{2.62}$$

are the respective Markovian splitting subspaces contained in $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$.

Both $\mathcal{X}_t^{+/-}$ and $\mathcal{X}_t^{-/+}$ are the predictor spaces contained in $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$. In this context, it was tacitly assumed that the superior Markovian splitting subspace $\mathcal{X}_t$ ($= \mathcal{X}_t^{+/-} \vee \mathcal{X}_t^{-/+}$) is *internal*. This attributes to the more general fact that

$$\mathcal{H} = \mathcal{Y}_t^- \vee \mathcal{X}_t^- \vee \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ = \mathcal{Y}_t^- \vee \mathcal{Y}_t^+ , \tag{2.63}$$

i.e., $\mathcal{X}_t \subset \mathcal{Y}_t^- \vee \mathcal{Y}_t^+$, and also

$$\mathcal{S}_t^- = \mathcal{Y}_t^- \vee \mathcal{X}_t , \qquad \mathcal{S}_t^+ = \mathcal{Y}_t^+ \vee \mathcal{X}_t . \tag{2.64}$$

As both $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ need to contain both the forward *and* backward predictor space (cf. (2.55)), which are neither completely contained in $\mathcal{Y}_t^-$ nor $\mathcal{Y}_t^+$, the definitions of $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ can not be made alone in terms of equality to $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$.

Observability, constructibility, and minimality

Observability, constructibility and hence minimality of predictor spaces/Markovian splitting subspaces can now be given in terms of the decomposition of $\mathcal{X}_t$ with respect to $\mathcal{Y}_t^-$ and $\mathcal{Y}_t^+$. Using (see Lemma 2.2.6 in Lindquist and Picci, 2015)

$$\mathcal{A} = \hat{E}\{\mathcal{B}|\mathcal{A}\} \oplus (\mathcal{A} \cap \mathcal{B}^\perp) , \tag{2.65}$$

a Markovian splitting subspace can be decomposed as

$$\mathcal{X}_t = \hat{E}\{\mathcal{Y}_t^+|\mathcal{X}_t\} \oplus \left(\mathcal{X}_t \cap (\mathcal{Y}_t^+)^\perp\right) , \tag{2.66a}$$

$$\mathcal{X}_t = \hat{E}\{\mathcal{Y}_t^-|\mathcal{X}_t\} \oplus \left(\mathcal{X}_t \cap (\mathcal{Y}_t^-)^\perp\right) . \tag{2.66b}$$

These two decompositions give rise to a rather illustrative definitions of observability and constructibility. As the part of $\mathcal{X}_t$ intersecting with $(\mathcal{Y}_t^+)^\perp$ is obviously not part of $\mathcal{Y}_t^+$, it cannot retrieved or calculated from the observation of the future outputs, i.e., $\mathcal{X}_t \cap (\mathcal{Y}_t^+)^\perp$ defines the unobservable part of $\mathcal{X}_t$. Likewise, $\mathcal{X}_t \cap (\mathcal{Y}_t^-)^\perp$ defines the unconstructible part of $\mathcal{X}_t$. Hence, the observable subspace of $\mathcal{X}_t$ is given by $\hat{E}\{\mathcal{Y}_t^+|\mathcal{X}_t\}$ and the constructible subspace of $\mathcal{X}_t$ by $\hat{E}\{\mathcal{Y}_t^-|\mathcal{X}_t\}$. Thus, observability and constructibility are defined as follows.

Definition 2.3. *(Lindquist and Picci, 2015) A Markovian splitting subspace $\mathcal{X}_t$ is said to be observable if $\mathcal{X}_t \cap (\mathcal{Y}_t^+)^\perp = 0$ and to be constructible if $\mathcal{X}_t \cap (\mathcal{Y}_t^-)^\perp = 0$. If both conditions hold, it is furthermore minimal.*

This definition can be also seen from the standpoint of the spaces

$$\mathcal{N}_t^- = \mathcal{Y}_t^- \cap (\mathcal{Y}_t^+)^\perp , \qquad \mathcal{N}_t^+ = \mathcal{Y}_t^+ \cap (\mathcal{Y}_t^-)^\perp ,$$

which are called junk spaces of the past and future. By reason of forming the intersection of $\mathcal{Y}_t^-$ with the orthogonal complement of the future $\mathcal{Y}_t^+$, $\mathcal{N}_t^-$ is the subspace of the past containing no information on the future. By virtue of similar reasoning, $\mathcal{N}_t^+$ is the subspace of the future containing no information on the past. This can be also seen from Figure 2.4, as $\left(\mathcal{S}_t^+\right)^\perp = \mathcal{N}_t^-$ and $\left(\mathcal{S}_t^-\right)^\perp = \mathcal{N}_t^+$. A similar reasoning holds for $\mathcal{X}_t \cap (\mathcal{Y}_t^+)^\perp$ and $\mathcal{X}_t \cap (\mathcal{Y}_t^-)^\perp$.

Theorem 2.5. *(Lindquist and Picci, 2015) Let $(\mathcal{S}_t^-, \mathcal{S}_t^+)$ be the scattering pair of some Markovian splitting subspace $\mathcal{X}_t$. Then, this Markovian splitting subspace is observable if and only if*

$$\mathcal{S}_t^+ = \mathcal{Y}_t^+ \vee \left(\mathcal{S}_t^-\right)^\perp \tag{2.67}$$

and constructible if and only if

$$\mathcal{S}_t^- = \mathcal{Y}_t^- \vee \left(\mathcal{S}_t^+\right)^\perp . \tag{2.68}$$

The connection between the definitions of observability and constructibility given in terms of the junk spaces or in terms of the scattering pairs as by Theorem 2.5 follows from the decomposition of the ambient space with

$$\mathcal{H} = \mathcal{S}_t^+ \oplus \left(\mathcal{S}_t^+\right)^\perp = \left(\mathcal{Y}_t^+ \vee \left(\mathcal{S}_t^-\right)^\perp\right) \oplus \left(\mathcal{S}_t^+\right)^\perp = \mathcal{Y}_t^+ \vee \left(\mathcal{S}_t^-\right)^\perp \vee \left(\mathcal{S}_t^+\right)^\perp .$$

This is equal to (performing the orthogonal complement within $\mathcal{H}$ on both sides of the equation)

$$\left(\mathcal{Y}_t^+\right)^\perp \cap \left(\mathcal{S}_t^- \cap \mathcal{S}_t^+\right) = \{0\} .$$

Taking (2.55) into account, this equation describes the condition for observability of a Markovian splitting subspace given by Definition 2.3. The condition for constructibility of Theorem 2.5 can be related to the definition of constructibility in Definition 2.3 by a similar reasoning.

Evolution of the predictor space in time

To see how a Markovian splitting subspace evolves through time, define the wandering subspace

$$\mathcal{W} = \mathcal{S}_{t+1}^- \ominus \mathcal{S}_t^- . \tag{2.69}$$

Theorem 2.6. *(Lindquist and Picci, 2015) The evolution (forward in time) of a Markovian splitting subspace $\mathcal{X}_t$ is governed by*

$$\begin{aligned}
\mathcal{X}_{t+1}^{+/-} &\subset \mathcal{X}_t^{+/-} \oplus \mathcal{W}_t , \\
\mathcal{Y}_t &\subset \mathcal{X}_t^{+/-} \oplus \mathcal{W}_t .
\end{aligned} \tag{2.70}$$

Similarly, the evolution backward in time is given by

$$\begin{aligned}
\mathcal{X}_{t-1} &\subset \mathcal{X}_t^{-/+} \oplus \mathcal{W}_{t-1}^b , \\
\mathcal{Y}_{t-1} &\subset \mathcal{X}_t^{-/+} \oplus \mathcal{W}_{t-1}^b ,
\end{aligned}$$

where

$$\mathcal{W}_b = \mathcal{S}_t^+ \ominus \mathcal{S}_{t+1}^+ .$$

2.4.2 Systems with exogenous inputs

In the following, the extension of the coordinate-free framework to the case of systems with exogenous inputs will be discussed. The general system structure is shown in Figure 2.5. The two processes w and e_u are mutually orthogonal white-noise processes (cf. Section 2.3). In

this context, the predictor space of the joint system $P(z)$ & $H(z)$ is sought. This system can be decomposed according to $P(z)$ and $H(z)$ into the deterministic subsystem

$$x_d(t+1) = Ax_d(t) + Bu(t) \,, \tag{2.71a}$$
$$y_d(t) = Cx_d(t) + Du(t) \,, \tag{2.71b}$$

and the stochastic subsystem

$$x_s(t+1) = Ax_s(t) + Gw(t) \,, \tag{2.72a}$$
$$y_s(t) = Cx_s(t) + Jw(t) \,, \tag{2.72b}$$

where the joint output is

$$y(t) = y_d(t) + y_s(t) \,. \tag{2.73}$$

The remaining transfer functions are either the generation filter $C(z)$ of the input or the feedback transfer function $F(z)$ and should not be included in the predictor space. Hence, there are two major points, which need to be accounted for by the derivations:

1. The predictor space cannot be defined by taking only y into consideration. This would lead to a predictor space and later to a state, which does not only represent the dynamic behavior of the system $P(z)$ & $H(z)$ but also the dynamic behavior of $C(z)$ (and $F(z)$).

2. The intrinsic feedback introduced by $F(z)$ needs to be taken into consideration. If $F(z) = 0$, the system operates in open-loop.

Based on the last point, the following discussion is split in half. Whereas the first part of the discussion summarizes the general case with feedback, the later part considers the case without feedback.

Without knowledge of the system structure, the absence of feedback can be defined by the *feedback-free condition*

$$\mathcal{Y}_t^- \perp \mathcal{U}_t^+ | \mathcal{U}_t^- \,. \tag{2.74}$$

This condition simply states that the residual spaces resulting from the orthogonal complement of $\mathcal{U}_t^-$ in both $\mathcal{Y}_t^-$ and $\mathcal{U}_t^+$ are orthogonal if the system operates under open-loop conditions. Those spaces are in fact spanned by the past of w and the future of e_u. This condition is hence in line with assumption A5 given in Section 2.3.

In addition to

$$\mathcal{X}_t \subset \mathcal{Y}_t^- \vee \mathcal{U}_t^- \vee \mathcal{Y}_t^+ \vee \mathcal{U}_t^+ \,, \tag{2.75}$$
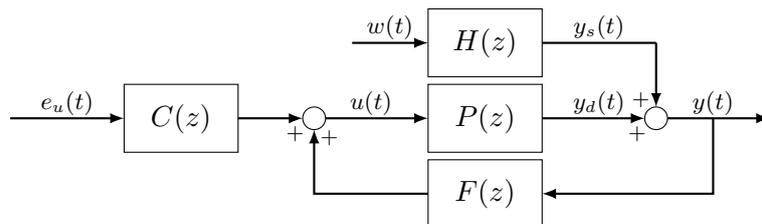

Figure 2.5: General structure of a system with exogenous inputs and feedback (Lindquist and Picci, 2015, p. 677)

defining a predictor space as an internal predictor space, a predictor space is furthermore called *causal* if

$$\mathcal{X}_t \subset \mathcal{Y}_t^- \vee \mathcal{U}_t^- \ . \tag{2.76}$$

This later condition clearly contains internalization of a predictor space. The space $\mathcal{X}_t$ is again contained in both the past space $\mathcal{X}_t^-$ and the future space $\mathcal{X}_t^+$ of the predictor space. The ambient space is now defined by

$$\mathcal{H} = \mathcal{Y}_t^- \vee \mathcal{U}_t^- \vee \mathcal{X}_t^- \vee \mathcal{Y}_t^+ \vee \mathcal{U}_t^+ \vee \mathcal{X}_t^+ \ . \tag{2.77}$$

General case with feedback

The general case for systems with exogenous inputs includes the feedback of system output. The definition of the predictor space given for this case describes the general procedure for identifying systems that operate under both closed-loop and open-loop conditions, i.e., if there is a controller or not. In this context, the richness condition of the assumptions of Section 2.3 is equivalent to

$$\left(\mathcal{U}_t^- \vee \mathcal{W}_t^- \right) \cap \left(\mathcal{U}_t^+ \vee \mathcal{W}_t^+ \right) = \{0\} \ , \tag{2.78}$$

which again includes the richness condition of u. The spaces $\mathcal{W}_t^-$ and $\mathcal{W}_t^+$ are the past and future of the conditional wandering process of $\mathcal{S}_t^-$, which is defined by

$$\mathcal{W}_t = \mathcal{S}_{t+1}^- \ominus \left(\mathcal{S}_t^- \vee \mathcal{U}_t \right) \ . \tag{2.79}$$

This conditional wandering process is not equal to the innovation process, as the wandering process would be defined by (cf. (2.3))

$$w(t) = \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} - \hat{E} \left\{ \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} \middle| \mathcal{P}_t^- \vee \mathcal{X}_{t-1}^- \vee \mathcal{U}_t \right\} \ .$$

The spaces $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ are defined similar to stochastic systems with

$$\mathcal{S}_t^- = \mathcal{P}_t^- \vee \mathcal{X}_t^- \ , \qquad \mathcal{S}_t^+ = \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ \ .$$

Furthermore, the space of the observable and not observable future inputs is defined as

$$\mathcal{F}_t^+ = \mathcal{U}_t^+ \vee \mathcal{W}_t^+ \ . \tag{2.80}$$

As a consequence of (2.78), it obeys

$$\mathcal{S}_t^- \cap \mathcal{F}_t^+ = \{0\} \ . \tag{2.81}$$

The following discussion of the coordinate-free framework for systems with exogenous inputs is similar to the discussion given for stochastic systems.

Definition 2.4. *(Lindquist and Picci, 2015) Let $\mathcal{P}_t^-$ be the joint past of the output y and input u, $\mathcal{Y}_t^+$ the future space of the output y, and $\mathcal{X}_t$ an arbitrary subspace of the ambient space $\mathcal{H}$ defined by (2.77), i.e., $\mathcal{X}_t \subset \mathcal{H}$. Then, $\mathcal{X}_t$ is called* oblique (i.e., conditional) splitting subspace *if it has the property*

$$\mathcal{P}_t^- \perp \mathcal{Y}_t^+ \mid \mathcal{X}_t \vee \mathcal{F}_t^+ \ , \tag{2.82}$$

i.e., if the past and future become conditionally orthogonal given $\mathcal{X}_t \vee \mathcal{F}_t^+$. It is furthermore an oblique Markovian splitting subspace *if*

$$(\mathcal{P}_t^- \vee \mathcal{X}_t^-) \perp (\mathcal{Y}_t^+ \vee \mathcal{X}_t^+) \mid \mathcal{X}_t \vee \mathcal{F}_t^+ , \tag{2.83}$$

which includes the oblique (i.e., conditional) Markov property (where $\mathcal{X}_t^-$ and $\mathcal{X}_t^+$ are the respective past and future of $\mathcal{X}_t$)

$$\mathcal{X}_t^- \perp \mathcal{X}_t^+ \mid \mathcal{X}_t \vee \mathcal{F}_t^+ . \tag{2.84}$$

In the case of exogenous inputs and feedback, the splitting property of $\mathcal{X}_t$ can be given only conditional to $\mathcal{F}_t^+$. That is, as (2.82) states, the joint past $\mathcal{P}_t^-$ does not contain any information regarding the future of y, only if information of the past, represented by $\mathcal{X}_t$, *and* the future inputs $\mathcal{F}_t^+$ are known. This is essentially the consequence of the two points mentioned at the beginning of the section. Without $\mathcal{F}_t^+$ in condition (2.82), $\mathcal{X}_t$ would be a Markovian splitting subspace containing also parts of the input process, as $\mathcal{Y}_t^+$ is also coupled with $\mathcal{U}_t^+$, which is in general not orthogonal to $\mathcal{U}_t^-$.

Lemma 2.7. *(Lindquist and Picci, 2015) Let $\mathcal{X}_t$ be a (Markovian) splitting subspace for $\mathcal{P}_t^-$ and $\mathcal{Y}_t^+$, then*

$$\hat{E}\{\lambda \mid \mathcal{P}_t^- \vee \mathcal{X}_t \vee \mathcal{F}_t^+\} = \hat{E}\{\lambda \mid \mathcal{X}_t \vee \mathcal{F}_t^+\} \ \forall \ \lambda \in \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ \tag{2.85}$$

or

$$\hat{E}_{\|\mathcal{F}_t^+}\{\lambda \mid \mathcal{P}_t^- \vee \mathcal{X}_t\} = \hat{E}_{\|\mathcal{F}_t^+}\{\lambda \mid \mathcal{X}_t\} \ \forall \ \lambda \in \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ . \tag{2.86}$$

The conditions and definition of an oblique Markovian splitting subspace are summed up by the next theorem.

Theorem 2.7. *(Lindquist and Picci, 2015) Let the space $\mathcal{F}_t^+$ be defined as in (2.80) and the spaces $\mathcal{S}_t^-$ and $\mathcal{S}_t^+$ be defined, such that*

$$(i) \quad \mathcal{Y}_t^+ \subset \mathcal{S}_t^+ , \qquad\qquad \mathcal{P}_t^- \subset \mathcal{S}_t^- , \qquad \mathcal{S}_t^- \cap \mathcal{F}_t^+ = \{0\} \tag{2.87a}$$

$$(ii) \quad \mathcal{S}_{t-1}^- \subset \mathcal{S}_t^- , \qquad\qquad \mathcal{S}_{t+1}^+ \subset \mathcal{S}_t^+ \tag{2.87b}$$

$$(iii) \quad \mathcal{S}_t^- \perp \mathcal{S}_t^+ \mid (\mathcal{S}_t^- \cap \mathcal{S}_t^+) \vee \mathcal{F}_t^+ \tag{2.87c}$$

Then, $\mathcal{X}_t$ is an oblique Markovian splitting subspace if and only if

$$\mathcal{X}_t = \mathcal{S}_t^- \cap \mathcal{S}_t^+ . \tag{2.88}$$

Furthermore,

$$\hat{E}_{\|\mathcal{F}_t^+}\{\mathcal{S}_t^+ \mid \mathcal{S}_t^-\} = \hat{E}_{\|\mathcal{F}_t^+}\{\mathcal{S}_t^+ \mid \mathcal{X}_t\} , \tag{2.89}$$

where the minimal spaces fulfilling the conditions are given by

$$\mathcal{S}_t^- = \mathcal{Y}_t^- \vee \mathcal{U}_t^- \vee \mathcal{X}_t^- = \mathcal{P}_t^- \vee \mathcal{X}_t^- , \qquad \mathcal{S}_t^+ = \mathcal{Y}_t^+ \vee \mathcal{X}_t^+ . \tag{2.90}$$

Let now the oblique Markovian splitting subspace be internal and causal, i.e., $\mathcal{X}_t^- \subset \mathcal{P}_t^-$, which in particular means $\mathcal{X}_t \subset \mathcal{P}_t^-$. Hence,

$$\mathcal{S}_t^- = \mathcal{P}_t^- . \tag{2.91}$$

Condition (2.81) is consequently also fulfilled. Under these assumptions the conditional wandering process of $\mathcal{S}_t^-$ is equal to the innovation process e. An oblique Markovian splitting subspace can then be defined as follows.

Theorem 2.8. *(Lindquist and Picci, 2015) The subspace*

$$\mathcal{X}_t^{+/-} = \hat{E}_{\|\mathcal{F}_t^+}\{\mathcal{Y}_t^+ \mid \mathcal{P}_t^-\} \tag{2.92}$$

is the minimal oblique Markovian splitting subspace or the minimal oblique predictor space contained in $\mathcal{P}_t^-$.

Note that this definition directly includes minimality of $\mathcal{X}_t^{+/-}$ and hence its observability and constructibility. Minimality means here that $\mathcal{X}_t^{+/-}$ contains no subspace which also obeys the conditions of an oblique Markovian splitting subspace as given by Theorem 2.7. In fact, this is the only definition that can be given for minimality as meaningful definitions for observability and constructibility cannot be specified for the general case.

The condition (2.81) is rather strong and might be violated under feedback (Lindquist and Picci, 2015, p. 688). Hence an oblique Markovian splitting subspace might also be defined based on a weaker condition as a one-step-ahead oblique Markovian splitting subspace.

Definition 2.5. *(Lindquist and Picci, 2015) A subspace $\mathcal{X}_t$ is a one-step-ahead oblique Markovian splitting subspace if*

$$\mathcal{S}_t^- \cap \mathcal{U}_t = \{0\} \tag{2.93}$$

and

$$\hat{E}_{\|\mathcal{U}_t}\{\mathcal{Y}_t \vee \mathcal{X}_{t+1}|\mathcal{P}_t^- \vee \mathcal{X}_t\} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t \vee \mathcal{X}_{t+1}|\mathcal{X}_t\} \ . \tag{2.94}$$

The condition for the one-step-ahead oblique Markovian splitting subspace is essentially equal to the calculations within a state-space model. The above defined one-step-ahead oblique Markovian splitting subspace is equivalent to an oblique Markovian splitting subspace if the strong condition (2.81) holds. Based on the one-step-ahead oblique Markovian splitting subspace, $\mathcal{X}_t^{+/-}$ is defined as follows.

Proposition 2.4. *(Lindquist and Picci, 2015) The oblique predictor space is given by the vector sum*

$$\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{\infty} \mathcal{X}_t^h \tag{2.95}$$

with

$$\mathcal{X}_t^h = \hat{E}_{\|\mathcal{U}_t}^{\mathcal{P}_t^-}\left\{ \hat{E}_{\|\mathcal{U}_{t+1}}^{\mathcal{P}_{t+1}^-}\left\{ \cdots \hat{E}_{\|\mathcal{U}_{t+h}}^{\mathcal{P}_{t+h}^-}\{\mathcal{Y}_{t+h}\}\right\}\right\} \ , \tag{2.96}$$

where $\hat{E}_{\|\mathcal{U}_{t+i}}^{\mathcal{P}_{t+i}^-}\{\cdot\}$, $i = 0, \cdots h$ is a short-hand notation for $\hat{E}_{\|\mathcal{U}_{t+i}}\{\cdot \mid \mathcal{Y}_{t+i}^- \vee \mathcal{U}_{t+i}^-\}$.

Special case without feedback – The feedback-free oblique predictor space

The relations for the general case with feedback can be simplified if there is no feedback. This feedback-free case is equal to the open-loop case, i.e., if there is no controller. In the absence of feedback, the definition of $\mathcal{F}_t^+$ becomes

$$\mathcal{F}_t^+ = \mathcal{U}_t^+ \oplus \mathcal{W}_t^+ \ , \tag{2.97}$$

which simplifies the definition of the oblique Markovian splitting subspace. This follows as now

$$\mathcal{S}_t^- \vee \mathcal{F}_t^+ = (\mathcal{S}_t^- \vee \mathcal{U}_t^+) \oplus \mathcal{W}_t^+ \tag{2.98}$$

and, furthermore, Lemma 17.4.4 in Lindquist and Picci (2015, pp. 696–697) states that

$$\hat{E}\{\mathcal{A}|\mathcal{B} \oplus \mathcal{C}\} = \hat{E}\{\mathcal{A}|\mathcal{X} \oplus \mathcal{C}\} \quad \Rightarrow \quad \hat{E}\{\mathcal{A}|\mathcal{B}\} = \hat{E}\{\mathcal{A}|\mathcal{X}\} , \tag{2.99}$$

if $\mathcal{X} \subset \mathcal{B}$[7]. Hence, the existence of the wandering subspace or the innovation process does not need to be taken explicitly into account.

Theorem 2.9. *(Lindquist and Picci, 2015) Let $\mathcal{P}_t^-$ be the joint past of the input u and output y, $\mathcal{Y}_t^+$ the future space of the output y, and $\mathcal{X}_t$ an arbitrary subspace of the ambient space $\mathcal{H}$ defined by (2.77), i.e., $\mathcal{X}_t \subset \mathcal{H}$, satisfying the feedback-free condition*

$$\left(\mathcal{P}_t^- \vee \mathcal{X}_t^-\right) \perp \mathcal{Y}_t^+ | \mathcal{U}_t^- .$$

Then, $\mathcal{X}_t$ is called feedback-free oblique splitting subspace if

$$\mathcal{P}_t^- \perp \mathcal{Y}_t^+ \mid \mathcal{X}_t \vee \mathcal{U}_t^+ . \tag{2.100}$$

It is furthermore a feedback-free oblique Markovian splitting subspace if

$$(\mathcal{P}_t^- \vee \mathcal{X}_t^-) \perp (\mathcal{Y}_t^+ \vee \mathcal{X}_t^+) \mid \mathcal{X}_t \vee \mathcal{U}_t^+ . \tag{2.101}$$

As before, under the assumption of the richness condition,

$$\hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+|\mathcal{P}_t^-\} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+|\mathcal{X}_t\} ,$$
$$\hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{S}_t^+|\mathcal{S}_t^-\} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{S}_t^+|\mathcal{X}_t\} .$$

With the above definitions, it is not yet possible to establish a connection to the concept of perpendicular intersection, which however facilitates the analysis of minimality. In order to align the derivation for open-loop systems with exogenous inputs with the setting of perpendicular intersection, let the usual scattering pair $(\mathcal{S}_t^-, \mathcal{S}_t^+)$ be replaced by the extended scattering pair

$$\mathcal{S}_t^{e,-} = \mathcal{S}_t^- \vee \mathcal{U}_t^+ , \qquad \mathcal{S}_t^{e,+} = \mathcal{S}_t^+ \vee \mathcal{U}_t^+ . \tag{2.102}$$

Based on the extended scattering pair $(\mathcal{S}_t^{e,-}, \mathcal{S}_t^{e,+})$, the condition

$$\mathcal{S}_t^{e,-} \perp \mathcal{S}_t^{e,+} | \mathcal{S}_t^{e,-} \cap \mathcal{S}_t^{e,+} \tag{2.103}$$

of perpendicular intersection can be again stated, and, in particular,

$$\mathcal{S}_t^{e,-} \cap \mathcal{S}_t^{e,+} = \mathcal{X}_t \vee \mathcal{U}_t^+ \tag{2.104}$$

follows. Then, the conditions for observability and constructibility and hence minimality, are along the lines of the definitions given for stochastic systems.

---

[7]Which is the case for $\mathcal{S}_t^- \vee \mathcal{U}_t^-$ as $\mathcal{X}_t \subset \mathcal{S}_t^-$.

Theorem 2.10. *(Lindquist and Picci, 2015) Let $(\mathcal{S}_t^{e,-}, \mathcal{S}_t^{e,+})$ be the extended scattering pair of some oblique Markovian splitting subspace $\mathcal{X}_t$. Then, this oblique Markovian splitting subspace is observable if and only if*

$$\mathcal{S}_t^{e,+} = \left(\mathcal{S}_t^{e,-}\right)^{\perp} \vee \mathcal{Y}_t^+ \vee \mathcal{U}_t^+ \tag{2.105}$$

and constructible if and only if

$$\mathcal{S}_t^{e,-} = \left(\mathcal{S}_t^{e,+}\right)^{\perp} \vee \mathcal{P}_t^- \vee \mathcal{U}_t^+ . \tag{2.106}$$

The calculation of the (feedback-free) minimal oblique Markovian splitting subspace follows directly from Theorem 2.8 under consideration of (2.98) and (2.99).

Theorem 2.11. *(Lindquist and Picci, 2015) The subspace*

$$\mathcal{X}_t^{+/-} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+ \mid \mathcal{P}_t^-\} \tag{2.107}$$

is the minimal oblique Markovian splitting subspace or the minimal predictor space contained in $\mathcal{P}_t^-$.

Evolution of the predictor space in time

The following theorem will be crucial for the derivation of the main theorem of recursive subspace identification. It gives the description of the evolution of $\mathcal{X}_t^{+/-}$ in time. It is assumed that the oblique Markovian splitting space is internal.

Theorem 2.12. *(Lindquist and Picci, 2015) Let $\mathcal{E}_t$ be the space generated by the innovation*

$$e(t) = y(t) - \hat{E}\{y(t) \mid \mathcal{P}_t^- \vee \mathcal{U}_t\} ,$$

and $\mathcal{X}_t^{+/-}$ the (minimal internal) oblique Markovian splitting space of time t. Then, the following inclusions hold:

$$\mathcal{X}_{t+1}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{U}_t \vee \mathcal{Y}_t , \tag{2.108a}$$

$$\mathcal{Y}_t \subset \left(\mathcal{X}_t^{+/-} \vee \mathcal{U}_t\right) \oplus \mathcal{E}_t , \tag{2.108b}$$

$$\mathcal{E}_t \subset \mathcal{X}_t^{+/-} \vee \mathcal{U}_t \vee \mathcal{Y}_t . \tag{2.108c}$$

This result holds for both the case with feedback as well as for the feedback-free case. Note that the first equation is the coordinate-free representation of the state equation of the predictor system (2.5). The subspace inclusions are equal to the ones given in Theorem 2.6 as the first inclusion could be also noted for the general case (non-internal oblique Markovian splitting space) by

$$\mathcal{X}_{t+1}^{+/-} \subset \left(\mathcal{X}_t^{+/-} \vee \mathcal{U}_t\right) \oplus \mathcal{W}_t .$$

The additional $\mathcal{U}_t$ stems from the existence of the exogenous input.

2.5 Basic equations

Although the coordinate-free framework is a versatile basis and mighty tool for the description and analysis of subspace methods, a framework based on the system description gives a slightly better understanding regarding the meaning of the projections. Due to the then introduced assumption of a generic model, this framework becomes coordinate-based and is mostly used for derivations in terms of numerical calculations. This coordinate-based framework gives the basic equations that describe the relationship between the data vectors u_t^-, u_t^+, y_t^-, and y_t^+. These basic equations follow from the description of the future outputs y_t^+. In terms of an iteration of the innovation form (2.4), the future outputs are hence described by

$$
\begin{aligned}
y_t^+ &= \mathcal{O}_k \mathcal{C}_\infty p_t^- + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ \\
&= \mathcal{O}_k x(t) + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ .
\end{aligned}
\tag{2.109}
$$

In this equation, Ψ_k^d and Ψ_k^s are Toeplitz matrices of the Markov parameters of the deterministic and stochastic subsystems, where

$$
\Psi_k^d =
\begin{bmatrix}
D & 0 & 0 & \cdots & 0 \\
CB & D & 0 & \cdots & 0 \\
CAB & CB & D & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & 0 \\
CA^{k-2}B & CA^{k-3}B & CA^{k-4}B & \cdots & D
\end{bmatrix} ,
$$

and Ψ_k^s is formed likewise. Furthermore, $\mathcal{O}_k$ and $\mathcal{C}_\infty$ denote the extended observability matrix and the reversed and extended reachability matrix for an infinite past horizon. They are defined by

$$
\mathcal{O}_k = \begin{bmatrix} C^{\mathrm{T}} & (CA)^{\mathrm{T}} & \cdots & (CA^{k-1})^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} ,
$$

$$
\mathcal{C}_\infty = \begin{bmatrix} \cdots & A_K^2 \begin{bmatrix} B_K & K \end{bmatrix} & A_K \begin{bmatrix} B_K & K \end{bmatrix} & \begin{bmatrix} B_K & K \end{bmatrix} \end{bmatrix} .
$$

The structure of the reversed reachability matrix results from the structure of the past data vector p_t^-. Based on the observations of u and y, the calculation of the state $x(t)$ follows with

$$
x(t) = \sum_{i=-\infty}^{t} A_K^{t-i} B_K u(i) + \sum_{i=-\infty}^{t} A_K^{t-i} K y(i) .
\tag{2.110}
$$

If p_t^- is based on a finite past horizon, the calculation of the state $x(t)$ also includes an initial state $x(t_0)$, i.e.,

$$
x(t) = A_K^k x(t_0) + \sum_{i=t_0}^{t} A_K^{t-i} B_K u(i) + \sum_{i=t_0}^{t} A_K^{t-i} K y(i) .
\tag{2.111}
$$

In this case, the future outputs are given by

$$
\begin{aligned}
y_t^+ &= \mathcal{O}_k (A_K^k x(t_0) + \mathcal{C}_k p_t^-) + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ \\
&= \mathcal{O}_k x(t) + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ .
\end{aligned}
\tag{2.112}
$$

Only if the system is of ARX type, which implies that A_K is nilpotent for a certain choice of k, i.e., $A_K^k = 0$, the term $A_K^k x(t_0)$ vanishes and the finite past is equal to the infinite past. As the description of $x(t)$ by (2.110) can only be made based on observable data vectors, the description of the future outputs by (2.109) or (2.112) is based on both the predictor form for the description of the state $x(t)$, i.e., the past of the system, and on the innovation form for the description of the future outputs y_t^+, i.e., the future of the system. Another representation of the future outputs y_t^+ is given if only the predictor form of the system is used. This yields

$$
\begin{aligned}
y_t^+ &= \mathcal{O}_k \mathcal{C}_\infty p_t^- + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ \\
&= \mathcal{O}_k^K x(t) + \Psi_k^{d,K} u_t^+ + \Psi_k^{s,K} y_t^+ + e_t^+ \ .
\end{aligned}
\tag{2.113}
$$

This representation will be referred to in terms of closed-loop identification. Note that now $\mathcal{O}_k^K$, $\Psi_k^{d,K}$, and $\Psi_k^{s,K}$ are also build from the system matrices of the predictor form, i.e., based on A_K, B_K, K, C, and D. As $\Psi_k^{s,K}$ is moreover used in conjunction with the future outputs, the main diagonal of $\Psi_k^{s,K}$ is zero.

2.6 Open-loop identification

In the following section methods for open-loop identification are summarized. These methods are to be used when there is no feedback from the outputs to the inputs, i.e., if there is no controller. It should be noted that not all methods are based on the coordinate-free framework. Some approaches use the coordinate-based framework resulting from numerical relations governed by the system. For the sake of brevity, the discussion is limited to methods used in the remainder of the thesis. Other methods, which will be referred to, are briefly summarized in terms of their basic ideas and respective references.

In terms of the following review, the distinction between approaches, methods, and algorithms will remain as given in the introduction:

- Approach: basic concept or template procedure of an identification
- Method: exact theoretical description of an identification
- Algorithm: numerical implementation of a method

Based on the underlying principles of the methods, they are divided into two main approaches. These two approaches are called *realization approach* and *state-regression approach* (cf. Viberg, 1995, Bauer and Ljung, 2002, where similar classifications are made). Although the methods – whose number is already limited due to the fixed theoretical framework – might be different, the core operations are either of the realization approach or of the state-regression approach.

2.6.1 Realization and state-regression approach

Realization approach

Except for implementation-related differences, all realization methods conform with the following procedure:

1. Determination of the extended observability matrix $\mathcal{O}_k$ of a high-order model or any equivalent intermediate entity comprising the extended observability matrix from (2.109)

2. Reduction of the model to order n and if necessary extraction of the extended observability matrix by means of a singular value decomposition; see Theorem 2.14

3. Determination of the Toeplitz matrix Ψ_k^d or intermediate entity comprising Ψ_k^d

4. Determination (upto a similarity transformation) of $\hat{A}$, $\hat{C}$ from the reduced-order observability matrix and $\hat{B}$, $\hat{D}$ from the Toeplitz matrix or respective intermediate entity using the estimate of the observability matrix of the reduced-order model

5. Estimation of the Kalman gain K and innovation variance Σ_{ee}

Although an estimation of the Kalman gain is not explicitly considered in the standard methods, an algorithm for the estimation of the Kalman gain K and innovation variance Σ_{ee} based on the realization framework is outlined in Knudsen (2002).

State-regression approach

Likewise, except for implementation-related differences, all state-regression methods (abbreviated by state methods) conform with the following procedure:

1. Estimation of states $x(t)$ and $x(t+1)$ of a high-order model or any entities comprising these state from (2.109)

2. Reduction of the model to order n and if necessary extraction and estimation of the states $x(t)$ and $x(t+1)$ by means of a singular value decomposition; see Theorem 2.14

3. Estimation (upto a similarity transformation) of the system matrices by solving

$$\min_{A,B,C,D} \left\| \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \right\|_{\mathcal{H}}^2 \tag{2.114}$$

4. Estimation of the covariance matrices Q, R, S from the residuals

$$\begin{bmatrix} \hat{w}(t) \\ \hat{e}(t) \end{bmatrix} = \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} - \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \tag{2.115}$$

as

$$E \left\{ \begin{bmatrix} \hat{w}(t) \\ \hat{e}(t) \end{bmatrix} \begin{bmatrix} \hat{w}^{\mathrm{T}}(t) & \hat{e}^{\mathrm{T}}(t) \end{bmatrix} \right\} = \begin{bmatrix} \hat{Q} & \hat{S} \\ \hat{S}^{\mathrm{T}} & \hat{R} \end{bmatrix} \tag{2.116}$$

5. Estimation of the Kalman gain K and the innovation variance Σ_{ee}

The estimate of the Kalman gain is given by

$$\hat{K} = (\hat{A}P\hat{C}^{\mathrm{T}} + \hat{S})(\hat{C}P\hat{C}^{\mathrm{T}} + \hat{R})^{-1} , \tag{2.117}$$

where P is the stationary error covariance of $P(t) = E\{(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^{\mathrm{T}}\}$ satisfying the algebraic Riccati equation (Katayama, 2005, pp. 282–285, p. 292)

$$P = \hat{A}P\hat{A}^{\mathrm{T}} + \hat{Q} - (\hat{A}P\hat{C}^{\mathrm{T}} + \hat{S})(\hat{C}P\hat{C}^{\mathrm{T}} + \hat{R})^{-1}(\hat{A}P\hat{C}^{\mathrm{T}} + \hat{S})^{\mathrm{T}} \ . \tag{2.118}$$

Similar to Knudsen (2002), an approach for the determination of the Kalman gain $\hat{K}$ and the innovation variance $\hat{\Sigma}_{ee}$ without calculating the solution of the algebraic Riccati equation is outlined in Katayama (2010). These algorithms can however not guarantee that the forward innovation model is of minimal phase, i.e., $\lambda_{\max}(A - KC) < 1$ is not guaranteed.

Core calculations

Both the realization and state approaches are essentially based on two core calculations. The first is the extraction of either the state or observability matrix from (2.109). This extraction is part of the optimal predictor, which in turn is the implementation of Theorem 2.9. At this point of the identification procedure, the order of the model is still equal to the number of rows in y_t^+, which is why the model is called high-order model.

Theorem 2.13. *(Katayama and Picci, 1999; van Overschee and De Moor, 1996) Under the assumption $\mathcal{U}_t^- \cap \mathcal{U}_t^+ = \{0\}$, i.e., under the assumption of persistence of excitation, the optimal prediction $\hat{y}_t^+$ of y_t^+ based on the data*

p_t^- *and u_t^+ is given by*

$$\hat{E}\{y_t^+ | \mathcal{P}_t^- \vee \mathcal{U}_t^+\} = \mathcal{O}_k \mathcal{C}_\infty p_t^- + \Psi_k^d u_t^+ = \hat{y}_t^+ \ , \tag{2.119}$$

where

$$\hat{E}_{\|\mathcal{U}_t^+}\{y_t^+ | \mathcal{P}_t^-\} = \mathcal{O}_k \mathcal{C}_\infty p_t^- = \mathcal{O}_k x(t) \ , \tag{2.120a}$$

$$\hat{E}_{\|\mathcal{P}_t^-}\{y_t^+ | \mathcal{U}_t^+\} = \Psi_k^d u_t^+ \ . \tag{2.120b}$$

Regarding an orthogonal decomposition of the data as done by some methods, the following corollary characterizes the resulting implications. A similar statement, focusing primarily on the implication in terms of the state estimation, was made in Katayama (2005) and Katayama and Picci (1999).

Corollary 2.1. *(Bathelt, Söffker, and Jelali, 2015) By decomposing $\hat{E}\{y_t^+ \mid \mathcal{U}_t^+ \vee \mathcal{P}_t^-\}$ into two orthogonal projections according to*

$$\begin{aligned}
\hat{E}\{y_t^+ &\mid \mathcal{P}_t^- \vee \mathcal{U}_t^+\} \\
&= \hat{E}\left\{y_t^+ \left| \hat{E}\left\{\mathcal{P}_t^- \left| (\mathcal{U}_t^+)^\perp\right.\right\}\right.\right\} + \hat{E}\{y_t^+ | \mathcal{U}_t^+\} \\
&= \tilde{\Pi}_k \hat{E}\left\{p_t^- \left| (\mathcal{U}_t^+)^\perp\right.\right\} + \tilde{\Psi}_k u_t^+ \ ,
\end{aligned} \tag{2.121}$$

the resulting operator $\tilde{\Psi}$ is non-causal (unless u is white noise) and

$$\hat{E}\left\{y_t^+ \left| \hat{E}\left\{\mathcal{P}_t^- \left| (\mathcal{U}_t^+)^\perp\right.\right\}\right.\right\} = \mathcal{O}_k \hat{E}\left\{x(t) \left| (\mathcal{U}_t^+)^\perp\right.\right\} \ . \tag{2.122}$$

By absence of causality of $\tilde{\Psi}_k$, it is referred to the absence of a lower triangular structure, which implies causality (cf. (2.109) and (2.112)).

The second core calculation is the order reduction based on a singular value decomposition. Although this can be also expressed in terms of an operator theoretical framework (see Lindquist and Picci, 2015, pp. 33–35), it is more comprehensible if it is discussed in terms of the actual data. Let therefore $\mathcal{O}_k x(t)$ or its numerical equivalent $\mathcal{O}_k \mathrm{x_N}(t)$ be separated from (2.109) by the projection (2.120a).

Theorem 2.14. *Let the singular value decomposition of $\mathcal{O}_k \mathrm{x_N}(t)$ be*

$$\mathcal{O}_k \mathrm{x}_N(t) = U\Sigma V^{\mathrm{T}} = \begin{bmatrix} U_n & U_r \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_r \end{bmatrix} \begin{bmatrix} V_n^{\mathrm{T}} \\ V_r^{\mathrm{T}} \end{bmatrix} , \tag{2.123}$$

where U_n and V_n are the matrices constructed from the left and right singular vectors corresponding to the first n singular values, i.e., corresponding to Σ_n. Then, the rank n approximation of $\mathcal{O}_k \mathrm{x}_N(t)$ is given by

$$\arg \min_{\mathrm{rank}(\Gamma)=n} \|\mathcal{O}_k \mathrm{x}_N(t) - \Gamma\|_2 = U_n \Sigma_n V_n^{\mathrm{T}} . \tag{2.124}$$

A rank n estimate of the observability matrix is hence given by

$$\hat{\mathcal{O}}_k = U_n \Sigma_n^{1/2} . \tag{2.125}$$

Equally, a rank n estimate of the state is given by

$$\hat{\mathrm{x}}_N(t) = \Sigma_n^{1/2} V_n^{\mathrm{T}} . \tag{2.126}$$

Proof. For a proof of the rank n approximation, see De Moor (1993), van Overschee and De Moor (1996), pp. 110–111, or Katayama (2005) and the references therein. As explained by Lemma 2.9 in Katayama (2005),

$$\mathrm{Im}(\mathcal{O}_k) = \mathrm{Im}(U), \quad \mathrm{Im}_{\mathrm{Row}}\left(\mathrm{x_N}(t)\right) = \mathrm{Im}_{\mathrm{Row}}(V^{\mathrm{T}}) .$$

This yields, for some matrix $T_1 \in \mathbb{R}^{k_f p \times n}$,

$$\mathcal{O}_k = UT_1 ,$$

and, taking a basis transformation into account, a rank n estimate of $\mathcal{O}_k$ as in (2.125). The same holds for the rank n estimation of $\mathrm{x}_N(t)$ by (2.126). □

Standard methods

In the following, the standard methods of the subspace approach for the identification of systems operating under open-loop conditions are summarized. The remainder of methods discussed in the literature are only versions and modifications of this set of basic methods.

MOESP. The standard method in terms of the realization approach is MOESP (MIMO/Multivariable output-error state-space model identification). The MOESP method is based on the numerical relations between the input and output data governed by the system and does not use the coordinate-free framework of subspace identification. These methods extracts the sought-after entities directly from the data. See the review in the following.

N4SID. The standard method of the state approach is the N4SID method (numerical algorithms for subspace state-space system identification). As outlined in van Overschee and De

Moor (1993) for time series, the basis is the interpretation of the projections of Theorem 2.13 in the context of a state estimation by a non-steady state Kalman filter. This was extended to systems with exogenous inputs in van Overschee and De Moor (1994, 1996). This interpretation of the optimal predictor provides valuable insight into the meaning of the projections. A thorough explanation of this rationale is given in van Overschee and De Moor (1996, pp. 100–106). Regarding the implementation of the coordinate-free framework, N4SID uses parts of it, like Theorem 2.13. In terms of the implementation, the derivations follow numerical rationales. In fact, the first of two numerical implementation of the N4SID approach uses also realization-approach techniques for the calculation of the system matrices. However, this way of calculating the model also facilitates the exact identification of systems that are not of ARX-type (Bauer, 2005). The second numerical implementation follows exactly the general state approach, which, for a finite past, results in a slightly incoherent estimation of $x(t)$ and $x(t+1)$.

CCA. The CCA method (calculation based on the canonical correlation analysis) was derived completely against the background of the above outlined coordinate-free framework of subspace identification. Based on the underlying canonical correlation analysis, a realization-based and state-based implementation can be given. See the review in the following.

2.6.2 Methods

Multivariable output-error state-space model identification

The group of identification algorithms summarized under MOESP are introduced and analyzed in Verhaegen and Dewilde (1992a), Verhaegen and Dewilde (1992b), Verhaegen (1993b), and Verhaegen (1994). The principle of the underlying approach is the QR decomposition of the input-output data of the system which has to be identified. Based on the basic algorithm called ordinary MOESP (Verhaegen and Dewilde, 1992a), several instrumental variable (IV) algorithms were proposed. Of these algorithms, the PO-MOESP algorithm (instrumental variables made from past inputs and past outputs) is dominantly used. In terms of the ordinary MOESP algorithm, where $\Psi_k^s = I$ is assumed, the observability matrix is estimated from an orthogonal projection of the output onto the orthogonal complement of the input, i.e.,

$$\hat{E}\left\{Y_t^+ \,\middle|\, \left(\mathcal{U}_t^+\right)^\perp\right\} = \mathcal{O}_k \hat{E}\left\{x_N(t) \,\middle|\, \left(\mathcal{U}_t^+\right)^\perp\right\} + E_t^+ \, . \qquad (2.127)$$

As for example outlined in Verhaegen and Dewilde (1992b) or De Moor (1993), the estimation of $\mathcal{O}_k$ using the singular value decomposition of the left hand-side of (2.127) will not be affected by the disturbance E_t^+ as long as E_t^+ is white noise, which is guaranteed by the assumption. In terms of the PO-MOESP algorithm, where the assumption of $\Psi_k^s = I$ is dropped, the observability matrix is extracted from a projection onto the complementary past $P_t^- \ominus U_t^+$. Here, this complementary past serves as the instrument. The estimation of Ψ_k^d is basically given for all algorithms by the orthogonal projection onto the future inputs

$$\hat{E}\left\{Y_t^+ \,\middle|\, \mathcal{U}_t^+\right\} = \mathcal{O}_k \hat{E}\{x_N(t) | \mathcal{U}_t^+\} + \Psi_k^d U_t^+ \, . \qquad (2.128)$$

To extract Ψ_k^d, the additional term $\mathcal{O}_k \hat{E}\{\mathrm{x}_\mathrm{N}(\mathrm{t})|\mathcal{U}_t^+\}$ needs to be removed. This is done by a left-multiplication with $(\mathcal{O}_k^\perp)^\mathrm{T}$ as

$$\mathrm{Im}(\mathcal{O}_k^\perp) \perp \mathrm{Im}(\mathcal{O}_k) = \mathrm{Ker}\left(\left(\mathcal{O}_k^\perp\right)^\mathrm{T}\right) \tag{2.129}$$

This yields

$$\left(\mathcal{O}_k^\perp\right)^\mathrm{T} \hat{E}\left\{\mathrm{Y}_\mathrm{t}^+ \middle| \mathcal{U}_t^+\right\} = \left(\mathcal{O}_k^\perp\right)^\mathrm{T} \Psi_k^d \mathrm{U}_\mathrm{t}^+ . \tag{2.130}$$

In terms of the singular value decomposition (2.123), an estimate of $(\mathcal{O}_k^\perp)^\mathrm{T}$ is given by U_r^T. As the estimates of A and C are known from $\hat{O}_k$, $(\mathcal{O}_k^\perp)^\mathrm{T}\Psi_k^d$ becomes linear in B and D. Hence, the estimates of B and D can be extracted using a least-squares approach; refer to Verhaegen and Dewilde (1992a) and Katayama (2005).

Methods based on canonical correlation analysis

The methods discussed in Katayama and Picci (1999) and Larimore (1990) use the canonical correlation analysis applied to the past data and the future data to calculate a state estimate. In Katayama and Picci (1999), it is shown that the obliqued projection

$$\hat{E}_{\|\mathcal{U}_t^+}\{y_t^+|\mathcal{P}_t^-\} = \Sigma_{y_t^+ p_t^- |u_t^+}\Sigma_{p_t^- p_t^- |u_t^+}^{-1} p_t^- = \mathcal{O}_k x(t) \tag{2.131}$$

can be calculated by a (conditional) canonical correlation analysis of the respective values. Defining the Cholesky factors of the conditional covariance matrices

$$\Sigma_{y_t^+ y_t^+ |u_t^+} = L_f L_f^\mathrm{T} ,$$
$$\Sigma_{p_t^- p_t^- |u_t^+} = L_p L_p^\mathrm{T} ,$$

the singular value decomposition of the normalized conditional cross-covariance $\Sigma_{y_t^+ p_t^- |u_t^+}$

$$L_f^{-1}\Sigma_{y_t^+ p_t^- |u_t^+}L_p^{-\mathrm{T}} = U\Sigma V^\mathrm{T} \approx U_n \Sigma_n V_n^\mathrm{T} , \tag{2.132}$$

is calculated. The reduction to the n dominant singular values gives the state estimate of a system of order n with

$$x(t) = \mathcal{C}_k \Sigma_{p_t^- p_t^- |u_t^+}^{-1} p_t^- = \Sigma_n^{1/2} V_n^\mathrm{T} L_p^{-1} p_t^- . \tag{2.133}$$

Refer to Katayama (2005) and Katayama and Picci (1999) for the numerical implementation of the resulting state-approach algorithm of the CCA method and the associated estimations of the states $x(t)$ and $x(t+1)$. Likewise, an estimate of the observability matrix can be calculated. This leads to a realization-approach algorithm. As additionally outlined, the canonical vector $V^\mathrm{T}\Sigma_{p_t^- p_t^- |u_t^+}^{-1/2}(p_t^-|\mathcal{U}_t^{+\perp})$ cannot be used to identify the system as it yields a non-causal representation (Katayama, 2005; Katayama and Picci, 1999). In Jansson and Wahlberg (1996), the CCA-based calculation of the oblique projection was derived using a weighted optimization approach with a special choice of the weighting matrix.

The algorithm introduced in Larimore (1983) as canonical variate analysis (CVA) is based on a modification of the canonical correlation analysis. Introducing the weighted prediction error

$$E\{\|y_t^+ - \hat{y}_t^+\|_{\Theta^{-1}}\} = E\{(y_t^+ - \hat{y}_t^+)^\mathrm{T}\Theta^{-1}(y_t^+ - \hat{y}_t^+)\} ,$$
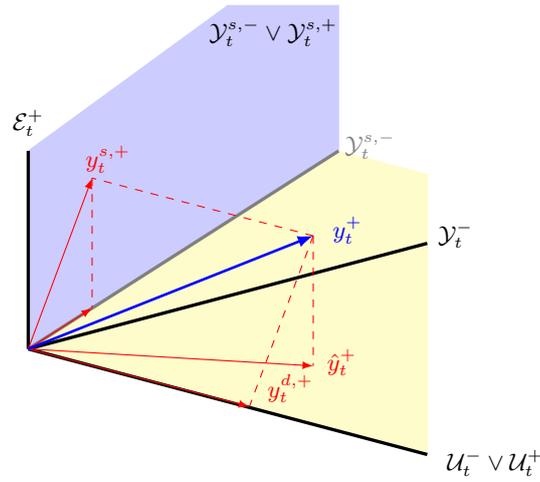
Figure 2.6: Principle of the orthogonal decomposition; $\mathcal{Y}_t^{s,-}$ and $\mathcal{Y}_t^{s,+}$ are the past and future spaces of the stochastic component of y

the calculation of the normalized cross-covariance is converted into

$$\Theta^{-1}\Sigma_{y_t^+ p_t^-}\Sigma_{p_t^- p_t^-}^{-\mathrm{T}/2} = U\Sigma V^{\mathrm{T}} \approx U_n\Sigma_n V_n^{\mathrm{T}} \ .$$

Using the resulting quasi-canonical vector $V^{\mathrm{T}}\Sigma_{p_t^- p_t^-}^{-1/2}p_t^-$ (see description of canonical correlation analysis in Katayama (2005, pp. 203–207)) as an auxiliary state, the system matrices are identified. Note that here the conditional covariances are not used and thus the covariance between u_t^- and u_t^+ is not taken into account, restricting this algorithm to white inputs. In Larimore (1990), this was addressed and the conditional CVA leading to a singular value decomposition of the form

$$\Theta^{-1}\Sigma_{y_t^+ p_t^- |u_t^+}L_p^{-\mathrm{T}/2} = U\Sigma V^{\mathrm{T}} \approx U_n\Sigma_n V_n^{\mathrm{T}}$$

was proposed.

Orthogonal decomposition

The standard methods do not a priori assume a certain structure of the system/model, as the information regarding the structure will ideally also be recovered from the data during the identification. Thus, a generic joint stochastic-deterministic model (basically an ARMAX structure) is used. By means of pole-zero-cancellations, any model structure can be recovered. This assumption of a joint model can be restrictive if the deterministic and stochastic subsystems have disjoint dynamics, i.e., if the noise is arbitrarily colored. An approach which cuts the connection between both subsystems and facilitates the identification of disjoint models is the orthogonal decomposition (ORT) approach proposed in Picci and Katayama (1996b). The theoretical basis of the ORT approach is the assumption of the absence of feedback. In this case, the output process y of a system can be decomposed as follows; see also Figure 2.6.

Theorem 2.15. *In the absence of feedback, the orthogonal projection of $y(t)$ onto $\mathcal{U}$ is causal, so that*

$$y_d(t) = \hat{E}\{y(t) \,|\, \mathcal{U}\} = \hat{E}\{y(t) \,|\, \mathcal{U}_{t+1}^-\} \tag{2.134}$$

is the deterministic component of $y(t)$, driven by the exogenous input u. The corresponding stochastic component is given by

$$y_s(t) = \hat{E}\{y(t) \,|\, \mathcal{U}^\perp\} \tag{2.135}$$

and is driven by the innovation process

$$
\begin{aligned}
e(t) &= y_s(t) - \hat{E}\{y_s(t) \,|\, \mathcal{Y}_t^{s,-}\} \\
&= y(t) - \hat{E}\{y(t) \,|\, \mathcal{U}_{t+1}^- \vee \mathcal{Y}_t^-\} \,,
\end{aligned}
\tag{2.136}
$$

where $\mathcal{Y}_t^{s,-}$ is the past of the stochastic component y_s. The future data vector of the output y_t^+ is hence decomposed according to

$$
\begin{aligned}
y_t^{d,+} &= \hat{E}\{y_t^+ \,|\, \mathcal{U}_t^- \vee \mathcal{U}_t^+\} \,, \\
y_t^{s,+} &= y_t^+ - \hat{E}\{y_t^+ \,|\, \mathcal{U}_t^- \vee \mathcal{U}_t^+\} \,.
\end{aligned}
\tag{2.137}
$$

Proof. The causality of the projection (2.134) follows from Theorem 9.1 in Katayama (2005), whereas (2.136) is shown in Proposition 3.1 in Picci and Katayama (1996b). □

By performing two separate identifications based on $y_t^{d,+}$ and $y_t^{s,+}$, the system is split into two independent subsystems according to

$$
\begin{bmatrix} x_d(t+1) \\ x_s(t+1) \end{bmatrix} = \begin{bmatrix} A_d & 0 \\ 0 & A_s \end{bmatrix} \begin{bmatrix} x_d(t) \\ x_s(t) \end{bmatrix} + \begin{bmatrix} B_d \\ K_s \end{bmatrix} \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} ,
\tag{2.138a}
$$

$$
y(t) = \begin{bmatrix} C_d & C_s \end{bmatrix} \begin{bmatrix} x_d(t) \\ x_s(t) \end{bmatrix} + \begin{bmatrix} D_d \\ I \end{bmatrix} \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} .
\tag{2.138b}
$$

This structure is equal to the Box-Jenkins structure of the classical system identification; see Ljung (2009, p. 87). The state estimation as outline in Katayama (2005) is given by

$$\mathcal{X}_t^{d,+/-} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^{d,+} |\mathcal{U}_t^-\} \,, \tag{2.139}$$

where

$$\mathcal{Y}_t^{d,+} = \hat{E}\{\mathcal{Y}_t^+ |\mathcal{U}_t^- \vee \mathcal{U}_t^+\} \tag{2.140}$$

follows from Theorem 2.15. If both subsystems share modes, the identified model is no longer minimal. This is expressed in terms of the predictor space by the following result.

Proposition 2.5. *(Picci and Katayama, 1996b) Let the (orthogonal) predictor spaces of the respective components y_d and y_s of y be defined by*

$$\mathcal{X}_t^{d,+/-} = \overline{\mathrm{span}}\{x_d(t)\} \,, \qquad \mathcal{X}_t^{s,+/-} = \overline{\mathrm{span}}\{x_s(t)\} \,.$$

Then, the predictor space as defined by (2.107) is related to these spaces by

$$\mathcal{X}_t^{+/-} \subset \mathcal{X}_t^{d,+/-} \oplus \mathcal{X}_t^{s,+/-} \,. \tag{2.141}$$

Implementations of the identification algorithm for the deterministic part are given in Chiuso and Picci (2004a) and Katayama (2005), whereas the stochastic part can be identified by any

method for the identification of stochastic systems, e.g., Tanaka and Katayama (2006, 2007) and van Overschee and De Moor (1993). Given the estimation based on finite past data, the stochastic component will be perturbed by a residual term, which arises from the orthogonal projection (2.134). An algorithm taking this deviation into account is presented in Picci and Katayama (1996a).

The merit of this approach is the feasibility of the sole identification of the deterministic subsystem. Assuming the noise process has different dynamics, identifications using N4SID or PO-MOESP will result in models, whose number of eigenvalues/poles, i.e., the order, exceeds that of the actual deterministic subsystem. A system reduction of these models to the order of the deterministic subsystem will not give the actual deterministic eigenvalues as the additional eigenvalues of the stochastic subsystem are not cut off but intermixed with the deterministic eigenvalues. It is also advantageous that the bases of the models for the deterministic and stochastic subsystems can be selected independently, e.g., choosing the a stochastic balanced basis for the stochastic model without interfering with the basis of the deterministic system (Chiuso and Picci, 1999). The PI-MOESP algorithm of the MOESP family can also be used to extract the deterministic subsystem. The IV idea is however different.

2.6.3 Order estimation

One of the few parameters needed for the identification by subspace methods is the order of the system or model. If the order is however not known beforehand, it has to be estimated based on the available data. There are two approaches discussed in the literature. The first approach is based on the singular values resulting form the SVD used for the model reduction and extraction of the state or observability matrix. One way is to check for the number of dominant singular values and set the order to this number, as for example suggested in terms of the N4SID method in van Overschee and De Moor (1996). The second approach is the use of order estimation criteria, which are introduced in Bauer (2001). Those are based on the Frobenius norm or the 2-norm of the residuals of the rank n approximations of the state or the observability matrix. These residuals are the remaining singular values $\hat{\sigma}_i$, $i = n + 1, \cdots, \min(k_f p, k_p(m + p))$ of the SVD after a certain order is chosen. The first criterion, called NIC, is defined by

$$NIC(n) = \sum_{j=n+1}^{\min(k_f p, k_p(m+p))} \hat{\sigma}_j^2 + C(T)d(n)/T \qquad (2.142)$$

whereas the SVC (singular value criterion) is given by

$$SVC(n) = \hat{\sigma}_{n+1}^2 + C(T)d(n)/T \ , \qquad (2.143)$$

where T is the length of the recorded data, $C(T)$ is a penalty term, which is based on the length of the recorded data and the horizons k_f and k_p, and $d(n)$ is the number of independent model parameters. The value of $C(T)$ should lie in the range $\log T$ to $k_f k_p \log T$. For systems with exogenous inputs, the value of $d(n)$ is given in Hannan and Deistler (2012) with

$$d(n) = np + n(m + p) + mp \ . \qquad (2.144)$$

The order is then chosen according to that value of n minimizing $NIC(n)$ or $SVC(n)$. The consistency of the criteria is also analyzed in Bauer (2001).

An order estimation approach for CCA-based methods can be derived from the discussions in Fujikoshi and Veitch (1979), where the dimensionality of canonical correlation analysis is discussed[8]. Adapted to subspace identification, the measure calculated for each order n is

$$C_n = (N-1) \left(\frac{\hat{\sigma}_{n+1}^2}{1 - \hat{\sigma}_{n+1}^2} + \cdots + \frac{\hat{\sigma}_{k_f p}^2}{1 - \hat{\sigma}_{k_f p}^2} \right) - 2(k_f p - n)(k_p m + k_p p - n) \;, \qquad (2.145)$$

where N is the column number of the tail matrices, and $\hat{\sigma}_i$, $i = n+1, \cdots, \min(k_f p, k_p(m+p))$ are the individual singular values of the SVD of the model reduction and extraction of the state or the observability matrix. The order is then chosen equal to that n giving the smallest C_n. The condition for the use of this estimation method is $k_f p \le k_p m + k_p p$.

2.7 Closed-loop identification

If there is a controller and hence a feedback from the output to the input, the previously explained open-loop methods can not (theoretically sound) be used with respective data, as there is now a correlation between the disturbances and the future inputs. As the discussion of the coordinate-free framework explained, a correct estimation of the predictor space/state is hence not possible. Hence, separate methods are needed for the closed-loop case. As above, the following explanation is limited to the methods referred to in the remainder of the thesis.

2.7.1 General problem

The general issue in a closed-loop environment is the occurrence of correlations of future inputs with future innovations (actually with past innovations as well). Applying open-loop methods despite the data was collected under closed-loop conditions, the model will exhibit a bias. As now $\Sigma_{e_t^+ u_t^+} = 0$ does no longer hold, the obliqued projections of y_t^+ onto $\mathcal{P}_t^-$ and $\mathcal{U}_t^+$ yield

$$\begin{aligned}
\hat{E}_{\|\mathcal{U}_t^+}\{y_t^+ \mid \mathcal{P}_t^-\} &= \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{O}_k x(t) + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ \mid \mathcal{P}_t^-\} \\
&= \mathcal{O}_k x(t) + \Psi_k^s \hat{E}_{\|\mathcal{U}_t^+}\{e_t^+ \mid \mathcal{P}_t^-\} \\
&= \mathcal{O}_k x(t) - \Psi_k^s \Sigma_{e_t^+ u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} \Sigma_{u_t^+ p_t^-} p_t^-
\end{aligned} \qquad (2.146)$$

and likewise

$$\hat{E}_{\|\mathcal{P}_t^-}\{y_t^+ \mid \mathcal{U}_t^+\} = \Psi_k^d u_t^+ + \Psi_k^s \Sigma_{e_t^+ u_t^+} u_t^+ \;. \qquad (2.147)$$

The last equation of (2.146) follows from

$$\hat{E}_{\|\mathcal{U}_t^+}\{e_t^+ \mid \mathcal{P}_t^-\} = \Sigma_{e_t^+ p_t^-} - \Sigma_{e_t^+ u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} \Sigma_{u_t^+ p_t^-} p_t^- \;,$$

where by definition $\Sigma_{e_t^+ p_t^-} = 0$ (cf. (2.3)). In terms of a linear regression analysis, this result is also given in Ljung and McKelvey (1996a,b).

---

[8]My special thanks to Mr. Victor Solo of the University of New South Wales, Sydney for advising me of his paper.

2.7.2 Methods

There are three general approaches to closed-loop identification (see, e.g., Katayama (2005) and Ljung (2009)) – direct approach, indirect approach, joint input-output approach. In terms of subspace identification, the methods mainly follow either the direct approach or the joint input-output approach. The idea of using the joint input-output approach was introduced in Verhaegen (1993a) as a closed-loop approach for the MOESP methods. It was later used for the closed-loop implementation of the ORT approach in Katayama, Kawauchi, and Picci (2005) and Katayama and Tanaka (2007). The three main methods of the direct approach are the state-space ARX (SSARX) method (Jansson, 2003), predictor-based subspace identification (PBSID) method (Chiuso and Picci, 2005), and a method based on an explicit innovation estimation (PARSIM-E) (Qin and Ljung, 2003a). The advantage of these approaches is that they are also able to work in the open-loop setting. In the following, the focus will be on the PBSID method, as its theoretical derivation is based on the previously introduced coordinate-free framework. The SSARX method is similar to the PBSID method in terms of the basic idea but the calculation of the estimates is slightly different (see, e.g., Chiuso, 2007b).

Based on Proposition 2.4 and as explained in Chiuso and Picci (2005) and Lindquist and Picci (2015, p. 722), the oblique predictor space can be also extracted in a closed-loop setting.

Lemma 2.8. *(Chiuso and Picci, 2005) The oblique predictor space is given by the vector sum*

$$\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{k_f-1} \mathcal{X}_t^h \,, \tag{2.148}$$

where

$$\mathcal{X}_t^h = \hat{E}_{||\mathcal{P}_{[t,t+h)}^+} \{\mathcal{Y}_{t+h} \mid \mathcal{P}_t^-\} \tag{2.149}$$

and for finite dimensional systems $k_f \geq n$.

The relation between the SSARX method and the PBSID method is given by the extraction of the past component. In terms of the SSARX method, the calculation is defined by $\hat{E}\{y_t^+ - \Psi_{k_f}^{s,K} y_t^+ - \Psi_{k_f}^{d,K} u_t^+ \mid \mathcal{P}_t^-\}$ (Chiuso, 2007b). The single oblique projections $\hat{E}_{||\mathcal{P}_{t,t+h}^+} \{\mathcal{Y}_{t+h} \mid \mathcal{P}_t^-\}$ of (2.148) are essentially the same as every future value up to $t+h$ is removed. That is, the single oblique projections are equal to the respective rows of $\hat{E}\{y_t^+ - \Psi_{k_f}^{s,K} y_t^+ - \Psi_{k_f}^{d,K} u_t^+ \mid \mathcal{P}_t^-\}$. Numerical implementations of the PBSID method in terms of a least-squares algorithms are discussed in Chiuso (2006, 2007a). The optimized version of the PBSID algorithm (PBSID$_{\text{opt}}$) derived in Chiuso (2007a) was later also implemented in terms of a vector ARX (VARX) algorithm in Chiuso (2007b). This VARX-based algorithm starts by obtaining the first k_p coefficients of the truncated VARX model (no feed-through)

$$y_{N+k_f}(t) \approx \sum_{i=1}^{k_p} \hat{\Phi}_i p_{N+k_f}(t-i) = C\mathcal{C}_{k_p} P_t^- \,, \tag{2.150}$$

where the coefficients are also given by the product of the output matrix C of the system and the reversed and extended reachability matrix $\mathcal{C}_{k_p}$. In terms of the estimated coefficients, the past component $\hat{Y}_t^{+,P_t^-}$ of Y_t^+ is given by

$$\hat{Y}_t^{+,P_t^-} = \Xi_P P_t^- = \hat{\mathcal{O}}_k^K x_N(t) \,, \tag{2.151}$$

where

$$\Xi_{\mathrm{P}_\mathrm{t}^-} = \begin{bmatrix} \hat{\Phi}_{k_p} & \hat{\Phi}_{k_p-1} & \dots & \hat{\Phi}_{k_p-k_f+1} & \dots & \hat{\Phi}_1 \\ 0 & \hat{\Phi}_{k_p} & \dots & \hat{\Phi}_{k_p-k_f} & \dots & \hat{\Phi}_2 \\ \vdots & \vdots & \ddots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \hat{\Phi}_{k_p} & \dots & \hat{\Phi}_{k_f} \end{bmatrix}.$$

For an explanation of the structure of $\Xi_{\mathrm{P}_\mathrm{t}^-}$, refer to Remark 4.4 in Chiuso (2007b, p. 1043). The observability matrix is extracted from $\hat{Y}_\mathrm{t}^{+,\mathrm{P}_\mathrm{t}^-}$ by the weighted singular value decomposition

$$W_p^{-1}\hat{Y}_\mathrm{t}^{+,\mathrm{P}_\mathrm{t}^-} = U\Sigma V^\mathrm{T} \approx U_n\Sigma_n V_n^\mathrm{T} \tag{2.152}$$

as

$$\hat{\mathcal{O}}_k^K = W_p U_n \Sigma_n^{1/2} . \tag{2.153}$$

Here, the observability matrix is based on the predictor system and is hence constructed from $A_K = A - KC$ and C. The states are given by

$$\mathrm{x_N}(\mathrm{t}) = (\hat{\mathcal{O}}_k^K)^\dagger \hat{Y}_\mathrm{t}^{+,\mathrm{P}_\mathrm{t}^-} \qquad \mathrm{x_N}(\mathrm{t}+1) = (\hat{\mathcal{O}}_{k-1}^K)^\dagger \hat{Y}_\mathrm{t+1}^{+,\mathrm{P}_\mathrm{t}^-} . \tag{2.154}$$

With the states given, the system can be identified equal to the state approach.

Remark 2.1. As outlined in Chiuso and Picci (2005), the PBSID method is, like any other direct approach method, affected by the transient behavior of the innovations if the past horizon is finite. However, the major advantage the PBISD method has over any other method is that it can also deal with the identification of unstable systems. This is a direct consequence of the extraction of the oblique predictor space as done by (2.148) and (2.149); see Chiuso and Picci (2005, pp. 387–388).

2.8 Identification of systems by linear parameter-varying models

In the previous sections the model of the system has been considered to be a LTI model, as it has been assumed that the model will only be used for a limited range of operation, where the system behavior is nearly linear. If the model is however required to cover a much larger range of operation, where the system behavior is clearly non-linear, a different type of model is required. One model for the description of non-linear system behavior is the linear parameter-varying (LPV) representation of a system. The structure of such a LPV model is given by (see, e.g., Verdult and Verhaegen, 2002)

$$\begin{aligned} x(t+1) = & \begin{bmatrix} A^{(0)} & A^{(1)} & \cdots & A^{(s)} \end{bmatrix} \begin{bmatrix} x(t) \\ \mu(t) \otimes x(t) \end{bmatrix} \\ & + \begin{bmatrix} B^{(0)} & B^{(1)} & \cdots & B^{(s)} \end{bmatrix} \begin{bmatrix} u(t) \\ \mu(t) \otimes u(t) \end{bmatrix} \\ & + \begin{bmatrix} K^{(0)} & K^{(1)} & \cdots & K^{(s)} \end{bmatrix} \begin{bmatrix} e(t) \\ \mu(t) \otimes e(t) \end{bmatrix}, \end{aligned} \tag{2.155a}$$

$$y(t) = Cx(t) + Du(t) + e(t) \, , \tag{2.155b}$$

where $\mu(t) \in \mathbb{R}^s$ is the parameter vector. The remaining vectors and matrices correspond to definitions made for the linear time-invariant system description of (2.4). The structure of the above given LPV model reflects the assumption of state affinity. Methods for the LPV identification of systems are proposed in van Wingerden and Verhaegen (2009) and Verdult and Verhaegen (2002). The LPV approach proposed in Verdult and Verhaegen (2002) is an extension of the usual linear open-loop state approach to LPV models. Both the derivation and numerical implementation remain basically the same. The difference concerns only the structure of the data matrices of past and future inputs. In van Wingerden and Verhaegen (2009), a LPV method which is equivalent to the PBSID method is proposed. Thus, this method works in both open-loop and closed-loop settings.

It is important to note that, for the identification of LPV models, the parameter vector $\mu(t)$ needs to be known. In particular, the result of the identification heavily depends on the choice of the interval which $\mu(t)$ covers during the data collection, as it basically defines the range of operating points the resulting model is valid for.

One the major drawback of LPV algorithms is the rapidly growing dimension of the data matrices. For example, in terms of the identification of a MIMO system by the LPV method of Verdult and Verhaegen (2002), where $s = 3$, $k_p = k_f - 1 = 5$, and $m = p = 2$, the row number of the equivalents of the numerical past and future output data matrices P_t^- and Y_t^+ is 1397424. Similarly, in terms of the identification of a MIMO system by the LPV method of van Wingerden and Verhaegen (2009), where $s = 4$, $k_p = 5$, and $m = p = 2$, the row number of the equivalent of the numerical past data matrices P_t^- is 5456. See Table 1 in Verdult and Verhaegen (2002) and Table 1 van Wingerden and Verhaegen (2009) for more examples regarding the row numbers of the respective matrices[9].

2.9 Recursive methods for subspace identification

Similar to a change of the basic model type, the whole approach to the identification can be changed. Instead of recording the data and using all the data at once to calculate the model, it can be determined successively when a new data point becomes available. This type of identification is called recursive identification. In the early stages of recursive subspace identification, the focus was on the reduction of the numerical load (Lovera, Gustafsson, and Verhaegen, 2000). The primary goal was to prevent the direct calculation of the QR decomposition and SVD of large data matrices. Later, as sufficient computational power became available, the focus switched to the update of models of time-varying systems (Kameyama, Ohsumi, et al., 2005). The two basic approaches discussed in the literature will be reviewed in the following. Those are the update of the sample covariance matrices which avoids the QR decomposition and the subspace tracking approach which facilitates the avoidance of the SVD.

---

[9]Without introducing the construction of these numerical data matrices, it is not meaningful to give these tables here.

Update of data matrices

In Oku and Kimura (1999), Takei, Imai, and Wada (2001) or Takei, Nanto, et al. (2006), algorithms for the recursive computation of the observability matrix within the MOESP framework are given. Let the orthogonal projection, which is used for the estimation of the observability matrix, be given by (see (2.127))

$$Z_t^+ = \hat{E}\{Y_t^+ \mid (U_t^+)^\perp\} = \mathcal{O}_k \hat{E}\left\{x_N(t) \,\middle|\, (U_t^+)^\perp\right\} + E_t^+ . \tag{2.156}$$

As pointed out in Oku and Kimura (1999), the eigenvalue decomposition of

$$S_{zz} = Z_t^+ (Z_t^+)^T = \mathcal{O}_k \hat{E}\left\{x_N(t) \,\middle|\, (U_t^+)^\perp\right\} \hat{E}\left\{x_N(t) \,\middle|\, (U_t^+)^\perp\right\}^T \mathcal{O}_k^T \tag{2.157}$$

gives the same estimate of the observability matrix as the singular value decomposition of Z_t^+. Thus, the algorithm presented in Oku and Kimura (1999) concentrates on the recursive estimation of $Z_t^+ (Z_t^+)^T$, which is essentially the sample covariance matrix of Z_t^+. This eventually replaces the QR decomposition. To accomplish this, the auxiliary quantities

$$\alpha^{N+i} = \left(1 + u_t^{+T}(N+i)\left(S_{uu}^{i-1}\right)^{-1} u_t^+(N+i)\right)^{-1} ,$$
$$z_t^+(N+i) = y_t^+(N+i) - S_{yu}^{i-1}\left(S_{uu}^{i-1}\right)^{-1} u_t^+(N+i) \tag{2.158}$$

are introduced, where

$$u_t^+(N+i) = \left[u^T(N+i) \cdots u^T(N+i+k-1)\right]^T ,$$
$$y_t^+(N+i) = \left[y^T(N+i) \cdots y^T(N+i+k-1)\right]^T$$

are the $N+i$th column of $U_t^{+(i)} \in \mathbb{R}^{km \times N+i}$ and[10]

$$S_{uu}^{i-1} = U_t^{+(i-1)}\left(U_t^{+(i-1)}\right)^T .$$

In this context, $U_t^{+(i)}$ denotes the input matrix of the ith recursion step. The initial estimation is thus given by the index 0, where $U_t^{+(0)} = U_t^+$. The matrices S_{yu}^{i-1} and S_{zz}^{i-1} are formed according to the same pattern. After those auxiliary quantities are obtained, the update step is performed by calculating

$$S_{zz}^i = S_{zz}^{i-1} + \alpha^{N+i} z_t^+(N+i)(z_t^+)^T(N+i) \tag{2.159}$$

and

$$(S_{uu}^i)^{-1} = (S_{uu}^{i-1})^{-1} + \alpha^{N+i}(S_{uu}^{i-1})^{-1} u_t^+(N+i)(u_t^+)^T(N+i)(S_{uu}^{i-1})^{-1} ,$$
$$S_{yu}^i = S_{yu}^{i-1} + y_t^+(N+i)(u_t^+)^T(N+i) . \tag{2.160}$$

The updated version of the observability matrix is obtained from the eigenvalue decomposition of S_{zz}^i. Hence, updated estimates of A and C can be determined. Although not pointed out,

---

[10]The matrices S_{zz}^i, S_{yu}^i, and S_{uu}^i are equal to the sample covariance matrices (upto the factor $\frac{1}{N+i}$).

the Toeplitz matrix of the Markov parameters can be estimated as well. Hence, $\hat{B}$ and $\hat{D}$ can be estimated using the algorithm outlined in Verhaegen and Dewilde (1992a). An extension to the PO-MOESP algorithm is given also in Oku and Kimura (1999).

Opposed to the previous version which uses all incoming data, the algorithm proposed in Kameyama, Ohsumi, et al. (2005) uses a sliding window for the recursive calculation. That is, as new data becomes available, old data is discarded. This approach makes the algorithm also suitable for time-varying systems. The idea of the recursive calculation is based on the QR decomposition of two consecutive recursion steps, denoted by the superscripts $i-1$ and i, as

$$\begin{bmatrix} & \mathrm{u_t^+(N+i)} \\ R^{i-1}\left(Q^{i-1}\right)^{\mathrm{T}} & \mathrm{p_t^-(N+i)} \\ & \mathrm{y_t^+(N+i)} \end{bmatrix} = \begin{bmatrix} \mathrm{u_t^+(i-1)} \\ \mathrm{p_t^-(i-1)} \ R^i\left(Q^i\right)^{\mathrm{T}} \\ \mathrm{y_t^+(i-1)} \end{bmatrix} , \tag{2.161}$$

where $\mathrm{p_t^-(N+i)}$ is defined similar to $\mathrm{u_t^+(N+i)}$ or $\mathrm{y_t^+(N+i)}$. Denoting the matrices of the left-hand and right-hand side by the short forms $\begin{bmatrix} A \mid b \end{bmatrix}$ and $\begin{bmatrix} c \mid D \end{bmatrix}$,

$$AA^{\mathrm{T}} + bb^{\mathrm{T}} = \begin{bmatrix} A \mid b \end{bmatrix}\begin{bmatrix} A \mid b \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} c \mid D \end{bmatrix}\begin{bmatrix} c \mid D \end{bmatrix}^{\mathrm{T}} = cc^{\mathrm{T}} + DD^{\mathrm{T}}$$

follows. Thus, the calculation of

$$DD^{\mathrm{T}} = AA^{\mathrm{T}} + bb^{\mathrm{T}} - cc^{\mathrm{T}} \tag{2.162}$$

results in a recursive update of the sample covariance matrices of the past and future data which are given by $\frac{1}{N}AA^{\mathrm{T}}$ and $\frac{1}{N}DD^{\mathrm{T}}$. See Kameyama, Ohsumi, et al. (2005) for a detailed explanation. A direct calculation of the QR decomposition is thus replaced by the update of the sample covariance matrices.

Subspace tracking algorithms

Based on the PAST (projection approximation subspace tracking) and IV-PAST (instrumental variable PAST) algorithms introduced in B. Yang (1995) and Gustafsson (1998), a recursive algorithm for subspace identification is proposed in Lovera, Gustafsson, and Verhaegen (2000). In the context of the PAST algorithm, a measured complex-valued random signal z is defined by

$$z(t) = Ox(t) + e(t) , \tag{2.163}$$

where e is white noise, x is the incoming but not measurable random signal and O describes the measuring device, defining the propagation of x and thereby the signal subspace of z. The signal subspace is also connected through

$$E\{z(t)z^{\mathrm{H}}(t)\} = \Sigma_{zz} = O\Sigma_{xx}O^{\mathrm{T}} + \sigma^2 I \tag{2.164}$$

to the covariance of z. That is, given the eigenvalue decomposition

$$\Sigma_{zz} = U\Lambda U^{\mathrm{H}} = \begin{bmatrix} U_n & U_r \end{bmatrix} \begin{bmatrix} \Lambda_n & 0 \\ 0 & \sigma^2 I \end{bmatrix} \begin{bmatrix} U_n^{\mathrm{H}} \\ U_r^{\mathrm{H}} \end{bmatrix} , \tag{2.165}$$

then $\mathrm{Im}(U_n) = \mathrm{Im}(O)$. The same holds also for a singular value decomposition. In terms of the PAST algorithm, this eigenvalue decomposition is replaced by an unconstrained minimization.

Theorem 2.16. *(B. Yang, 1995) Let the unconstrained minimization*

$$\min_{W(t)} J\left(W(t)\right) = \min_{W(t)} E\left\{\left\|z(t) - W(t)W^H(t)z(t)\right\|^2\right\} , \qquad (2.166)$$

be given. Then, for some arbitrary unitary matrix $Q \in \mathbb{C}^{n \times n}$ and U_n, containing n distinct eigenvectors of (2.165), $W(t) = U_n Q$ if and only if $W(t)$ is the global minimum of (2.166). Thus, $\mathrm{Im}(W(t))$ is an estimate of the signal subspace O.

Using a sample function of z, which is furthermore restricted to real numbers, and introducing a forgetting factor λ, the expectation is replaced by

$$W(t) = \arg\min_{W(t)} \sum_{k=1}^{t} \lambda^{t-k} \|z(k) - W(t)W^{\mathrm{T}}(t)z(k)\|^2 . \qquad (2.167)$$

In terms of this equation, $W(t)$ gives an estimation of the respective eigenvectors of the eigenvalue decomposition of

$$\hat{\Sigma}_{zz} = \sum_{k=1}^{t} \lambda^{t-k} z(k) z^{\mathrm{T}}(k) . \qquad (2.168)$$

In the PAST algorithm, $W^{\mathrm{T}}(t)z(k)$ is replaced by $h(k) = W^{\mathrm{T}}(k-1)z(k)$ to simplify the calculation. Then, $W(t)$ is given by

$$W(t) = \hat{\Sigma}_{zh}\hat{\Sigma}_{hh}^{-1} , \qquad (2.169)$$

where the sample covariance matrices $\hat{\Sigma}_{zh}$ and $\hat{\Sigma}_{hh}$ are formed according to (2.168). The IV-PAST algorithm of Gustafsson (1998) adds the instrumental variable $\xi(t)$ to the calculation to eliminate colored-noise disturbances, which would otherwise influence the estimation of $W(t)$. This lifts the restriction of $e(t)$ to be white noise. Given an instrumental variable $\xi(t)$ $(E\{e(t)\xi^{\mathrm{H}}(t)\} = 0)$ and

$$E\{z(t)\xi^{\mathrm{H}}(t)\} = \Sigma_{z\xi} = \begin{bmatrix} U_n & U_r \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_n^{\mathrm{H}} \\ V_r^{\mathrm{H}} \end{bmatrix} , \qquad (2.170)$$

it again follows that $\mathrm{Im}(U_n) = \mathrm{Im}(O)$.

Theorem 2.17. *(Gustafsson, 1998) Let the unconstrained minimization*

$$\min_{W(t)} J\left(W(t)\right) = \min_{W(t)} E\left\{\left(z(t) - W(t)W^H(t)z(t)\right)\xi^H(t)\right\} , \qquad (2.171)$$

be given. Then, for some arbitrary unitary matrix $Q \in \mathbb{C}^{n \times n}$ and U_n, containing n distinct eigenvectors of (2.170), $W(t) = U_n Q$ if and only if $W(t)$ is the global minimum of (2.171). Thus, $\mathrm{Im}(W(t))$ is an estimation of the signal subspace O.

The numerical minimization then reads

$$W(t) = \arg\min_{W(t)} \sum_{k=1}^{t} \lambda^{t-k} \|z(k)\xi^{\mathrm{T}}(t) - W(t)W^{\mathrm{T}}(t)z(k)\xi^{\mathrm{T}}(t)\|^2 . \qquad (2.172)$$

The resulting $W(t)$ is

$$W(t) = \hat{\Sigma}_{z\xi}\hat{\Sigma}_{h\xi}^{-1} \ , \tag{2.173}$$

where $\hat{\Sigma}_{z\xi}$ and $\hat{\Sigma}_{h\xi}$ are again formed according to (2.168).

In Lovera, Gustafsson, and Verhaegen (2000), the resemblance between (2.163) and

$$\hat{E}\{y_t^+ \mid \mathcal{U}_t^{+\perp}\} = z_t^+ = \mathcal{O}_k\hat{E}\{x(t) \mid \mathcal{U}_t^{+\perp}\} + e(t)$$

of the ordinary MOESP method is used to derive a recursive algorithm for the estimation of the observability matrix $\mathcal{O}_k$ (equal to O in (2.163)). In this context, the value of $W(t)$ is equivalent to U_n of the SVD of (2.123). The most pressing problem is now the avoidance of the QR decomposition, as the SVD is already replaced by the PAST algorithm. Combining the results of the QR decomposition of the previous step with the data of the current step,

$$\begin{bmatrix} (\mathrm{U}_t^+)^{(i-1)} \ \mathrm{u}_t^+(\mathrm{N}+\mathrm{i}) \\ (\mathrm{Y}_t^+)^{(i-1)} \ \mathrm{y}_t^+(\mathrm{N}+\mathrm{i}) \end{bmatrix} = \begin{bmatrix} R_{11}^{i-1} & 0 & \mathrm{u}_t^+(\mathrm{N}+\mathrm{i}) \\ R_{21}^{i-1} & R_{22}^{i-1} & \mathrm{y}_t^+(\mathrm{N}+\mathrm{i}) \end{bmatrix} \begin{bmatrix} \left(Q_1^{i-1}\right)^{\mathrm{T}} & 0 \\ \left(Q_2^{i-1}\right)^{\mathrm{T}} & 0 \\ 0 & 1 \end{bmatrix}$$

is yielded. Using Givens rotations, the extended R matrix is transformed into (Lovera, Gustafsson, and Verhaegen, 2000)

$$\begin{bmatrix} (\mathrm{U}_t^+)^{(i-1)} \ \mathrm{u}_t^+(\mathrm{N}+\mathrm{i}) \\ (\mathrm{Y}_t^+)^{(i-1)} \ \mathrm{y}_t^+(\mathrm{N}+\mathrm{i}) \end{bmatrix} = \begin{bmatrix} R_{11}^i & 0 & 0 \\ R_{21}^i & R_{22}^{i-1} & \mathrm{z}_t^+(\mathrm{N}+\mathrm{i}) \end{bmatrix} (Q^i)^{\mathrm{T}} \ .$$

The calculated $\mathrm{z}_t^+(\mathrm{N}+\mathrm{i})$, where $i = 0, 1, \cdots, t$, are now used in conjunction with the PAST algorithm to determine $W(t)$. Extensions to the PI/PO-MOESP algorithms based on the IV-PAST algorithms are discussed in Lovera, Gustafsson, and Verhaegen (2000) and Wu, Yang, et al. (2008). Once the matrices $\hat{A}$ and $\hat{C}$ are estimated using $W(t)$, a linear regression approach is used to determine the remaining system matrices (see, e.g., Chiuso and Picci, 2001).

A combination of the recursive algorithm outlined in Oku and Kimura (1999) and the PAST algorithm is proposed in Oku and Kimura (2002). Here, the gradient of $J(W(t))$ of (2.166) is formed and used to update the eigenvalue decomposition of (2.157). Consequently, the eigenvalue decomposition of the algorithm of Oku and Kimura (1999) is also replaced and the estimate of observability matrix is recursively updated. Using the basic principle of the PAST approach, an approach, which provides, with respect to the progression of t, an improved convergence of $W(t)$ calculated in (2.167) and (2.172), is outlined in Mercère, Lecoeuche, and Vasseur (2003) and Mercère, Bako, and Lecuche (2008). The basic idea is based on a permutation of the rows of the observability matrix, so that the dynamics of the system are observable in one output. Thus, an auxiliary output is formed from a linear combination of all outputs of the system. Knowing the permutation matrix and the linear combination of the outputs, the observability matrix of the original system can be reconstructed after the recursion step is performed.

2.10 Summary

This chapter introduces the notation as well as the basic principles of subspace identification and the assumptions needed for the discussion and derivation of the later-proposed approach to recursive subspace identification and new algorithms. The briefly explained coordinate-free framework is the theoretical foundation of subspace identification. For both stochastic and joint stochastic-deterministic systems, the definition of the predictor space and its extraction form the data spaces, i.e., the spaces spanned by the output and the input, are discussed. This predictor space is the crucial entity of the whole identification procedure, as it is equivalent to the state of the system. Furthermore, conditions for minimality are discussed. The review of the methods has focuses on the explanation of the basic approaches, i.e., the realization approach and the state approach, and respective methods, which are referred to in the remainder of the thesis. This includes the basic open-loop methods MOESP (realization approach) and CCA (state approach, based on coordinate-free framework), the ORT approach, which decomposes the output into its stochastic and deterministic components, and the closed-loop method PBSID, which follows from a derivation within the coordinate-free framework. To understand the difference between the approach to recursive identification proposed in this thesis and the methods discussed in the literature, these existing methods of recursive subspace identification were discussed as well.

Recapitulation of main results

- The task of subspace identification is to calculate the matrices A, B, C, D, and K, of a linear time-invariant state-space representation of the system as well as the innovation covariance matrix Σ_{ee}. These matrices are equal to the actual system representation (given there is one) up to a global similarity transform. In terms of the methods' descriptions, the recorded data points $\{y(t), u(t), t = 0, 1, ..., T\}$ are considered to be random variables and are divided into a past segment as well as a present and future segment. This segmentation is necessary to decompose the output into a component associated with the state and a component associated with the future input.

- Identification methods are divided into methods of the realization approach and methods of the state-regression approach. Based on

$$y_t^+ = \mathcal{O}_k \mathcal{C}_\infty p_t^- + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ = \mathcal{O}_k x(t) + \Psi_k^d u_t^+ + \Psi_k^s e_t^+ \ ,$$

realization-based methods, like the MOESP method, extract the matrices $\mathcal{O}_k$, Ψ_k^d, and Ψ_k^s and calculate the respective system matrices based on these matrices. State-regression-based method, like the CCA method or the N4SID method, extract the states $x(t)$ and $x(t+1)$ (from y_{t+1}^+, where the "present" time is shifted to $t+1$) and calculate the system matrices based on a regression using $x(t)$, $x(t+1)$, $u(t)$, and $y(t)$. Furthermore, the MOESP method is based on a rationale resulting from the QR decomposition of the numerical data of a system's input and output. The MOESP method is hence not derived on the basis of the coordinate-free framework but on the basis of a coordinate-based framework.

- The ORT approach decomposes the output into its deterministic component, which is governed by the input u, and its stochastic component, which is governed by the innovation process e. It thus facilitates the identification of the deterministic subsystem in the presence of arbitrary disturbances.

- Although derived from the coordinate-free framework, the implementation of the closed-loop method PBSID is based on a VARX estimation (or an involved least-squares solution).

- Existing recursive methods perform the recursion in terms of the data and are not based on the methodological basis of subspace identification, i.e., are not based on the coordinate-free framework.

- The space spanned by a data vector of a stochastic process, e.g., y_t^+, and the row space of the associated numerical data matrix Y_t^+ build from tail matrices of one of the sample functions of y are equivalent if the column count N of the tail matrices goes to infinity, i.e.,

$$
\mathrm{Im}_{\mathrm{Row}}\left(\begin{bmatrix} y_N(t_0) \\ \cdots \\ y_N(t-2) \\ y_N(t-1) \end{bmatrix}\right) \xrightarrow{N \to \infty} \overline{\mathrm{span}}\{y_t^-\} \,,
$$

where

$$
y_N(t) = \begin{bmatrix} y(t) \ y(t+1) \ \cdots \ y(t+N-1) \end{bmatrix} \,.
$$

Hence, the derivations made using the coordinate-free framework or stochastic processes can be transfered to implementations in terms of data matrices without loss of correctness.

- For stochastic systems, a Markovian splitting subspace/predictor space is every subspace of an ambient data space which satisfy

$$
(\mathcal{Y}_t^- \vee \mathcal{X}_t^-) \perp (\mathcal{Y}_t^+ \vee \mathcal{X}_t^+) \mid \mathcal{X}_t \,,
$$

whereas for joint stochastic-deterministic systems an oblique Markovian splitting subspace/oblique predictor space is every subspace which satisfy

$$
(\mathcal{P}_t^- \vee \mathcal{X}_t^-) \perp (\mathcal{Y}_t^+ \vee \mathcal{X}_t^+) \mid \mathcal{X}_t \vee \mathcal{F}_t^+ \,,
$$

where $\mathcal{F}_t^+$ is the space spanned by the future input and wandering subspace (see (2.79)). The minimal predictor spaces are given by (for stochastic systems and joint stochastic-deterministic systems)

$$
\mathcal{X}_t^{+/-} = \hat{E}\{\mathcal{Y}_t^+ \mid \mathcal{Y}_t^-\} \,, \qquad \mathcal{X}_t^{+/-} = \hat{E}_{\|\mathcal{F}_t^+}\{\mathcal{Y}_t^+ \mid \mathcal{P}_t^-\} \,.
$$

The evolution through time of the predictor spaces is governed by (for stochastic systems and joint stochastic-deterministic systems)

$$
\mathcal{X}_{t+1}^{+/-} \subset \mathcal{X}_t^{+/-} \oplus \mathcal{W}_t \,, \qquad \mathcal{X}_{t+1}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{U}_t \vee \mathcal{Y}_t \,.
$$

3 Problem statement and conceptual outline

Based on the review of subspace identification made in the previous chapter, the goal drafted in the introductory chapter can now be stated precisely. This starts by stating the challenges which are to be faced in terms of the identification of industrial processes in Section 3.1. It is then changed over to the issues resulting from existing methods and algorithms when these challenges of real-world identifications are addressed. This serves furthermore as a detailed definition of the problem statement and hence as the motivation of this thesis. From the problem statement, the direction of the work is concluded, and the conceptual outline of the proposed methodological approach to recursive subspace identification and related algorithms of methods is given in Section 3.2.

3.1 Problem statement

Regarding the identification of industrial plants, the following challenges need to be addressed:

1. Multi-variable processes: several inputs and outputs

2. Non-linear processes: plant behavior depending on the operating point

3. Process disturbances: observed plant output subjected to measurement noise and internal disturbances

Hence, the course of action during an identification of such processes depends heavily on the degree these challenges appear. These challenges are however given adequate responses as

1. CCA method facilitates the handling of large data spaces (due to underlying canonical correlation analysis),

2. LPV methods facilitate the identification of nonlinear processes, and

3. ORT approach facilitates the identification of the process even under influences of disturbances.

The general question is hence whether it is possible to combine all three approaches into one. However, with respect to the LPV approach, the following problems can be hardly overcome:

- Dimensionality of the problem and numerical load: usual identification problems of industrial scale might have four and more inputs, four and more outputs, and require rather long past and future horizons. In such cases, the dimensionality and hence the numerical load rapidly increases, see Table 1 in Verdult and Verhaegen (2002) or Table 1 in van Wingerden and Verhaegen (2009).

- Determination of parameter vector: which plant variables are needed for the formation of the parameter vector?

- Plant safety: even if the parameter vector might be found, the remaining question is regarding the feasibility of a thorough identification, which covers the necessary

domain of both the input and the parameter vector. This is limited by possible safety or operational constraints.

These issues lead to the idea of replacing a LPV identification by a recursive re-identification of a LTI model during a change of the operating point. Regarding the existing recursive approaches, the review of the previous chapter points to the following issues:

- Existing recursive approaches are numerical methods, i.e., the problem of recursion is not addressed in terms of a sound theoretical framework like the coordinate-free framework.

- The algorithms are purpose-build, i.e., the basic approach is mainly a modification of the MOESP algorithm and is hence not easily applicable to other methods.

From these challenges and issues, the problem statement of this thesis can be rendered precisely as follows:

> *In order to circumvent the problems of a general identification of linear parameter-varying models by an adaptation of a linear time-invariant model, a methodological approach to recursive subspace identification is needed. This methodological approach is required to facilitate the implementation in terms of the majority of existing methods. This methodological approach needs to be derived within the theoretical framework of subspace identification. Basic identification methods need to be capable to yield sufficiently accurate results even when applied to problems of industrial scale.*

3.2 Conceptual outline

As this problem statement includes the derivation of a general approach to recursive subspace identification as a subtask, this subtask can be separated from the overall task of deriving a suitable recursive algorithm for the identification of industrial systems. Hence, there are the following two separate subtasks:

1. Derivation of new basic algorithms for open-loop and closed-loop identification in order to achieve sufficient results in terms of the identification of industrial systems
   As the goal is the recursive identification of industrial systems, the basic algorithms should not only facilitate the recursive identification but should also give sufficiently good results when applied to industrial systems. Whereas the former point is a problem with respect to the implementations of closed-loop methods, the latter point is a problem in terms of the open-loop identification. Hence, new algorithms for both open-loop and closed-loop identification are needed.

2. Derivation of a new approach for recursive subspace identification
   To be considered a recursive approach, information regarding the model from previous identifications needs to be transfered to following identifications and to be used for the calculation of the updated model. This information is furthermore required to be as close as possible to the actual model in terms of its meaning. By fulfilling this requirement the recursive approach becomes similar to the recursive identification of the classical methods, which update the parameter vector and hence directly update the model.

3.2.1 Conceptual outline for the derivation of basic algorithms

As the first subtask states, not only modifications with respect to the quality of the results are needed but also the modifications which yield algorithms that are in accordance with the state approach as it represents the implementation of the coordinate-free framework. In terms of open-loop identification, the main problem is the capability of the given implementations of the ORT approach, which was concluded to be the core approach of a method for the identification of industrial processes. In terms of closed-loop identification, the main hindrance is the implementation of the direct-approach methods by either VARX-based algorithms or least-squares-based algorithms. These two types of algorithms should be replaced by a state-approach algorithm.

Open-loop identification

Following from the above considerations, the proposed algorithm for the open-loop problem will be combination of the CCA method and the ORT approach. That is, the calculation of the deterministic state

$$x_d(t) = \hat{E}_{\|\mathcal{U}_t^+}\{y_t^{d,+}|\mathcal{U}_t^-\} \ ,$$

where $y_t^{d,+}$ follows from the orthogonal projection

$$y_t^{d,+} = \hat{E}\{y_t^+|\mathcal{U}_t^- \vee \mathcal{U}_t^+\} \ ,$$

should be made in terms of a canonical correlation analysis. On the basis of the CCA method of Katayama and Picci (1999), the archetype algorithm of the state estimation would be based on

$$L_f^{-1}\Sigma_{y_t^{d,+}u_t^-|u_t^+}L_p^{-\mathrm{T}} = U\Sigma V^{\mathrm{T}} \approx U_n\Sigma_nV_n^{\mathrm{T}} \ ,$$

where

$$\Sigma_{y_t^{d,+}y_t^{d,+}|u_t^+} = L_fL_f^{\mathrm{T}} \ ,$$
$$\Sigma_{u_t^-u_t^-|u_t^+} = L_pL_p^{\mathrm{T}} \ .$$

The state is then given by

$$x_d(t) = \mathcal{C}_k\Sigma_{u_t^-u_t^-|u_t^+}^{-1}u_t^- = \Sigma_n^{1/2}V_n^{\mathrm{T}}L_p^{-1}u_t^- \ .$$

This approach naturally leads to a state-approach algorithm and is expected to solve the reliability issues of the existing MOESP-based algorithms proposed in Chiuso and Picci (2004a) and Katayama (2005). These issues are demonstrated by the identification study of Chapter 5. Given this archetype algorithm, the actual goal is the combination of the orthogonal decomposition and the oblique projection, in order to calculate the deterministic state directly from the disturbance-affected y_t^+ instead of $y_t^{d,+}$.

Closed-loop identification

The proposed algorithm for the closed-loop problem will be based on the PBSID method, as this method is derived based on the coordinate-free framework, which also is the basis for the approach to recursive subspace identification. The proposed algorithm will directly

implement the theoretical basis of the PBSID method given by Lemma 2.8, which already describes the state-approach algorithm needed for the recursive identification. That is, the calculation of the oblique predictor space as

$$\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{k_f-1} \mathcal{X}_t^h \,,$$

where

$$\mathcal{X}_t^h = \hat{E}_{\|\mathcal{P}_{[t,t+h)}^+} \left\{ y(t+h) \mid \mathcal{P}_t^- \right\} \,,$$

needs simply to be given a suitable algorithm. By achieving this, the pre-estimation of a VARX model or the least-squares algorithm of the previously disclosed algorithms of the PBSID method, which would otherwise hinder the integration of the recursive scheme, can be avoided.

A long-term objective would be the inclusion of the ORT approach into the resulting algorithm. This is however a problem of its own, as the absence of feedback is a basic element of the derivation of the ORT approach, see Katayama (2005) and Picci and Katayama (1996b). Hence, an in-depth analysis regarding the orthogonal decomposition in a closed-loop setting and an answer to the question whether it is possible at all are needed.

3.2.2 Conceptual outline for the approach to recursive subspace identification

Once the first subtask is cleared, the focus shifts towards the derivation of the approach to recursive subspace identification. This approach is obviously required to reuse all the necessary information from the results of past identifications while keeping the compression of this information as high as possible to avoid the propagation of redundant information (Bathelt, Söffker, and Jelali, 2017). The derivations made within the coordinate-free framework of Chiuso and Picci (2003) and Lindquist and Picci (1996b, 2015) provide an entity, which actually fits all these requirements – the minimal predictor space. This predictor space is the core entity of every identification as its determination is the actual identification, whereas the remainder, i.e., the calculation of the system matrices, is rather a realization of the model than an identification (Akaike, 1974, p. 669, Lindquist and Picci, 1996b). That is, this subspace can be considered to be equal to the model. Considering an isolated identification, the predictor space contains furthermore all the necessary information from the past data needed for the prediction of the future, while being the smallest subspace to do so, i.e., there are no redundant information contained in this space. So does any basis of this space. Therefore, the approach to recursive subspace identification might the given as follows. Given the (oblique) predictor space $\mathcal{X}_t^{+/-}$ of time t, a substitution as

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \mid \mathcal{P}_{t+k}^- \right\} = \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \mid \mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k)}^- \right\}$$

should be eventually possible. That is, the projection of the future outputs of a system onto the joint space spanned by any previous predictor space and the space of the intermediate inputs and outputs should yield the same results as the projection of the future outputs onto the whole past. Such a substitution results in a natural recursive framework for subspace identification, as any future predictor space can then be defined in terms of any past predictor space. This claim needs to be thoroughly proven.

Note that such an approach will not lead to purpose-build algorithms, as the recursive scheme can be directly implemented in terms of any state-approach method, such as CCA or N4SID, and do not need any kind of modification of the respective algorithms. The only difference within the resulting algorithms concerns the past data, i.e., instead of a huge amount of past data, only the past state and the intermediate data would be needed. In terms of realization methods/algorithms, the only additional step is a state estimation. This step does however not alter the general procedure of any realization method/algorithm.

3.3 Summary

In this chapter, the scope of the following work is precisely defined. Considering that the identification of (large) industrial processes by LPV models leads to a number of more or less unsolvable issues, like the dimensionality of the problem or the correct definition of the parameter vector, the focus of the identification of such processes is switched to a recursive re-identification of a LTI model during the change of operating points. Therefore, a suitable approach to recursive subspace identification is needed, which can furthermore be easily implemented within algorithms of existing methods – a requirement existing recursive methods can hardly met. Furthermore, suitable open-loop and closed-loop algorithms for the identification of industrial processes are required.

As the approach will use also future data ($\mathcal{U}_t^+$ and $\mathcal{Y}_t^+$), it might be debatable whether it is a genuine recursive approach or just a recycling of already known information. However, the use of future data is a basic characteristic of all subspace methods. Furthermore, judging the approach proposed here in terms of the existing methods, which also perform the recursion using future data, it can be concluded that the approach proposed here will lead to a recursive methodology.

Recapitulation of main results

- In terms of open-loop identification, an improved algorithm of the ORT approach is needed, i.e., not a MOESP-based algorithm. This algorithm should be able to deal with the nonlinear nature of industrial processes. The idea is to derive a projection which yields the deterministic predictor space by combining the orthogonal decomposition of the ORT approach with the oblique projection used for the determination of the joint stochastic-deterministic predictor space. This projection facilitates in turn a combination of the CCA method with the ORT approach.

- In terms of closed-loop identification, an algorithm of the PBSID method needs to be derived which circumvents the VARX estimation. Based on the theoretical derivation of the PBSID method, such an algorithm is given by numerically realizing Lemma 2.8.

- In terms of recursive identification, an approach, which is based on the predictor space, needs to be derived.

4 Revision of methods for basic identifications

In order to achieve the goal of deriving a recursive subspace identification algorithm, which is able to deal with industrial problems, the task is subdivided into the subtask of deriving suitable basic identification algorithms and the subtask of deriving the recursive scheme. In the following the first subtask will be covered. This subtask is illustrated in Figure 4.1. For the moment, the question regarding an identification during the change of the plant's operating points is left aside. Only the basic identification problem, i.e., deriving a model for the present operating point, is considered. In this context, two new identification algorithms for both open-loop conditions and closed-loop conditions are proposed. These two algorithms are based on existing methods and approaches and are motivated by

- the need for an open-loop algorithm, which combines the capabilities of reliably approximating a suitable LTI model for an operating point of an otherwise nonlinear system and suppressing the influence of process disturbances on the identification result, and

- the need for a closed-loop algorithm, which avoids either an involved least-squares procedure or the pre-estimation of a VARX model, which is done in terms of all closed-loop identification methods of the direct approach.

These two algorithm are then the basis for an integration of the recursive scheme derived in Chapter 6.

In terms of the open-loop algorithm, the fundamental idea of the ORT approach – decomposition of the output into the deterministic and stochastic components – is appealing for the practical use, as usually only the deterministic plant behavior is needed while effects of disturbances need to be suppressed. The algorithm needs however to be furthermore able to deal with the great variety of realistic systems, at least in terms of the identification of a LTI model for an operating point. As however documented by the first trail identifications using Tennessee Eastman Process model, which feature the typical nonlinear behavior of real processes or plants, in Bathelt and Jelali (2014), the results of the existing realization-based algorithm of the ORT approach have been unexpectedly poor. Hence, an analysis of the approach and its previous implementation as well as a subsequent revision of the algorithm are needed. This proposed algorithm will further take changes into consideration, which ease the integration of the recursive scheme. The results and the new algorithm are presented in Section 4.1.

In terms of the closed-loop algorithm, the implication made by the explanations of Section 3.2.2 need to be taken into account. That is, for an easy implementation of the recursive scheme, the identification algorithm should facilitate the partial replacement of the past data by a state estimate. Considering the closed-loop methods, this results in a problem, as the algorithms of all methods include steps which hinder the integration of the recursive scheme. With respect to the favored PBSID method and also in terms of the SSARX method, this is the pre-estimation of a VARX model. Hence, an algorithm will be proposed which avoids this pre-estimation and thus facilitates the incorporation of the past state. This is presented in Section 4.2.
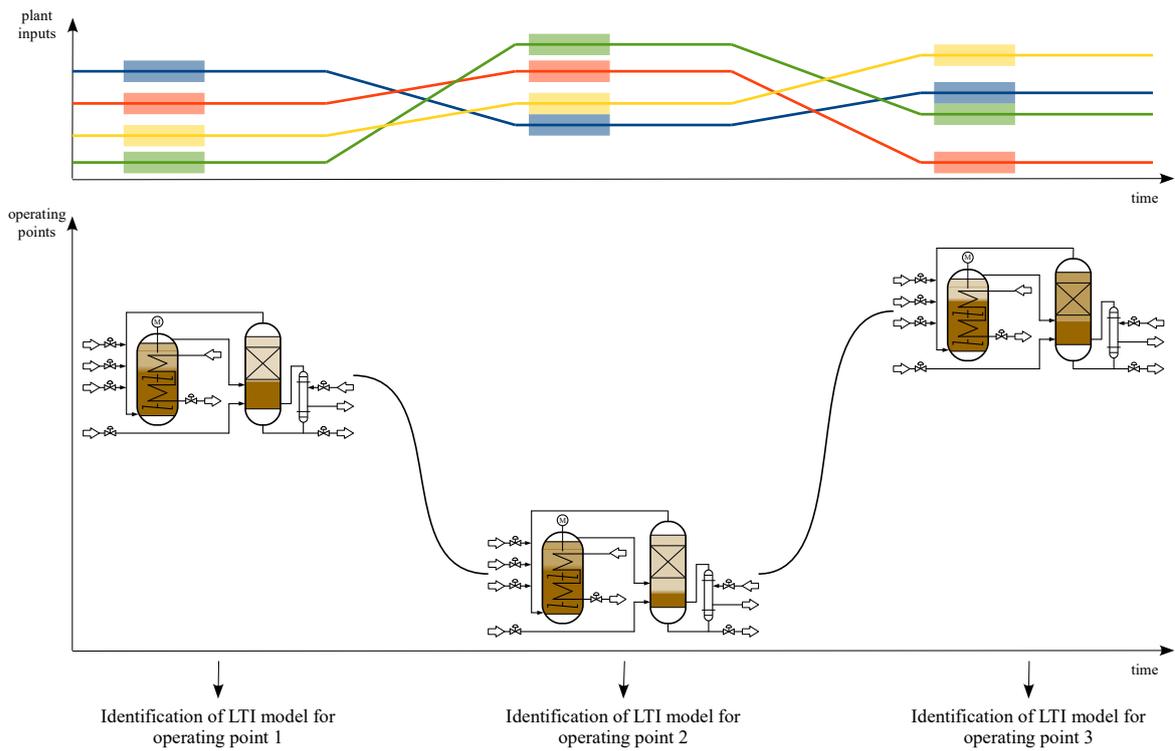
Figure 4.1: Principle of identification of new models for each operating point; the colored areas around the inputs symbolize the excitation for identification

4.1 Basic algorithm for open-loop identification

First trial runs of the ORT approach and its numerical implementation as published in Katayama (2005), which have been made using the Tennessee Eastman Process model and have been published in Bathelt and Jelali (2014), have revealed major problems when more realistic processes are considered. The standard ORT implementation is not able to give satisfying results (cf. results of the study of Chaper 5). This problem is however not limited to the ORT implementation but rather a general problem of realization-based algorithms, as the MOESP algorithms yield similar unsatisfactory results. This points to a more systematic and deep-seated problem centered around the realization approach. A solution for this problem and the also necessary change needed for the recursive approach have led to the concept of combining the fundamental idea of the ORT approach of Katayama (2005) and Picci and Katayama (1996b) with the CCA method of Katayama and Picci (1999).

The first subsection addresses the cause of the unsatisfactory results and thus motivates the switch to a state algorithm, which furthermore offers a basis for the incorporation of the recursive approach. In the second subsection, the method combining CCA and ORT is outlined. The third subsection describes the numerical implementation, and the fourth subsection gives the results of two theoretical examples. In Chapter 5, the results of an identification study are presented, showing that the method derived in the following yields the overall best results in both disturbance-free and disturbance-affected environments. The problem and the algorithm have first been disclosed in Bathelt, Söffker, and Jelali (2015). This section expands the content of this paper by several in-depth explanations.
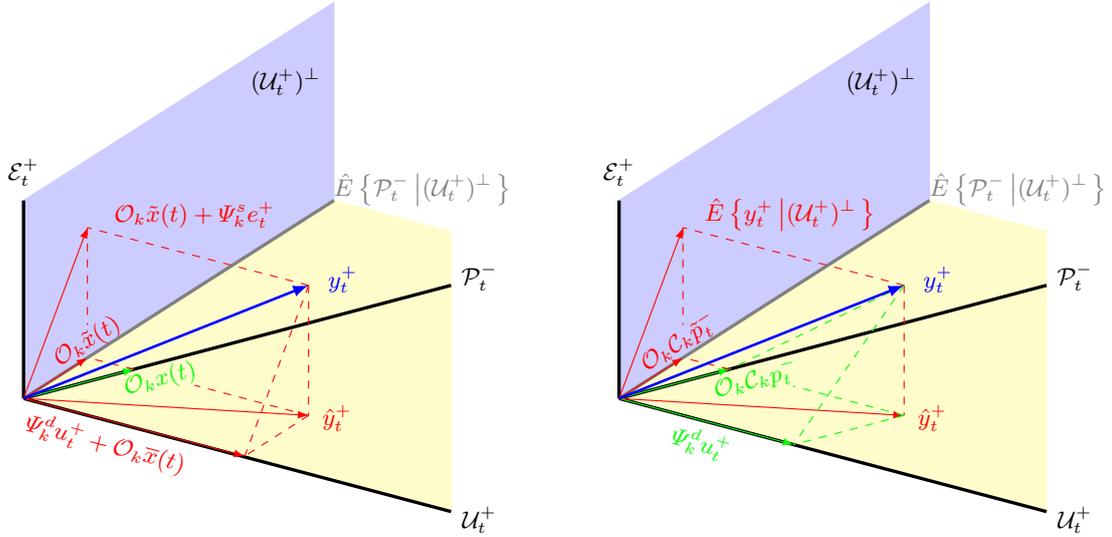
Figure 4.2: Orthogonal decomposition of y_t^+ according to PO-MOESP method (left) and theoretical decomposition of y_t^+ via oblique projections (right)

4.1.1 Problem of the existing algorithm of the orthogonal decomposition

The algorithm for the ORT approach as given in Katayama (2005) is based on the MOESP method. As discussed in Section 2.6.2, the MOESP method is based on an orthogonal decomposition of the future outputs with respect to the future inputs and the respective complement. This is not similar to the orthogonal decomposition of the ORT approach, where the future outputs are decomposed with respect to future *and* past inputs. Taking furthermore the optimal predictor discussed by Theorem 2.13 into consideration, the problem concerns the differences between the orthogonal decomposition of the MOESP method and the oblique decomposition as given by the optimal predictor.

In terms of the PO-MOESP algorithm, this orthogonal decomposition is depicted in the left part of Figure 4.2. The estimation of the matrices A and C is not a problem as the orthogonal projection onto the complement of the past $\hat{E}\left\{\mathcal{P}_t^- \,\big|\, (\mathcal{U}_t^+)^\perp\right\}$ in

$$\hat{E}\left\{y_t^+ \,\Big|\, \hat{E}\left\{\mathcal{P}_t^- \,\big|\, (\mathcal{U}_t^+)^\perp\right\}\right\} = \mathcal{O}_k \tilde{x}(t)$$

only affects the state as

$$\tilde{x}(t) = \hat{E}\left\{x(t) \,\Big|\, \hat{E}\left\{\mathcal{P}_t^- \,\big|\, (\mathcal{U}_t^+)^\perp\right\}\right\}$$

and hence yields, theoretically, the correct observability matrix. The actual problem is created by the associated orthogonal projection

$$\tilde{\Psi}_k u_t^+ = \hat{E}\{y_t^+ | \mathcal{U}_t^+\} \, . \tag{4.1}$$

Following Corollary 2.1, the operator $\tilde{\Psi}$ is non-causal. As the input u is not necessarily white noise, i.e., a processes with mutually orthogonal random variables, the state $x(t)$ correlates with u_t^+ through its deterministic part $x_d(t)$. This becomes evident if $\tilde{\Psi}_k$ is written as

$$\tilde{\Psi}_k = \Psi_k + \mathcal{O}_k \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} \, . \tag{4.2}$$

The lower-triangular structure mandatory for causality is lost, as $\mathcal{O}_k \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1}$ has in general not a lower-triangular structure[1]. As outlined in the MOESP paragraph of Section 2.6.2, the elimination of this term is achieved by the left-multiplication with $\left(\mathcal{O}_k^\perp\right)^{\mathrm{T}}$ (cf. (2.129)). Thus,

$$\left(\mathcal{O}_k^\perp\right)^{\mathrm{T}} \tilde{\Psi}_k = \left(\mathcal{O}_k^\perp\right)^{\mathrm{T}} \Psi_k + \left(\mathcal{O}_k^\perp\right)^{\mathrm{T}} \mathcal{O}_k \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} = \left(\mathcal{O}_k^\perp\right)^{\mathrm{T}} \Psi_k \tag{4.3}$$

follows. However, this relation holds only true if the congruence of the range space of $\mathcal{O}_k$ and the cokernel of $\mathcal{O}_k^\perp$ is exact, or, in terms of the numerical implementation, at least sufficiently exact.

If the specified order used for the estimation of $\mathcal{O}_k$ (see Section 2.6.2) is the same as the order of the true system, (4.3) is fulfilled despite the use of an estimate of $\mathcal{O}_k^\perp$. In cases of considerable order reduction, as for example occurring during the estimation of linear approximate models for nonlinear systems, which are in this context in general of high order (see Remark 5.1), the deviations between the cokernel of $\hat{\mathcal{O}}_k^\perp$ and the range space of $\mathcal{O}_k$ (in this context only of symbolic character) are considerable. Hence, (4.3) using $\hat{\mathcal{O}}_k$ yields

$$\left(\hat{\mathcal{O}}_k^\perp\right)^{\mathrm{T}} \tilde{\Psi}_k = \left(\hat{\mathcal{O}}_k^\perp\right)^{\mathrm{T}} \Psi_k + R, \tag{4.4}$$

where

$$R = \left(\hat{\mathcal{O}}_k^\perp\right)^{\mathrm{T}} \mathcal{O}_k \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1}. \tag{4.5}$$

As such a residual term is not taken into account in the theoretical derivation of the calculation of B and D, it will affect those matrices and will hence worsen the model. The effect R has on the matrices depends on the choice of n. The following examination shows that a reduction of the order n results in an enlargement of R.

Assuming that the identification is carried out as described in Section 2.6.2, the estimate of the observability matrix follows then as a result of the SVD (2.123) with $\hat{\mathcal{O}}_k = U_n \Sigma_n^{1/2}$. An estimate of $\mathcal{O}_k^\perp$ is consequently given by

$$\hat{\mathcal{O}}_k^\perp = U_r.$$

It follows that R is given by

$$
\begin{aligned}
R &= \left(\hat{\mathcal{O}}_k^\perp\right)^{\mathrm{T}} \mathcal{O}_k \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} \\
&= U_r^{\mathrm{T}} U S^{1/2} \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} \\
&= U_r^{\mathrm{T}} \begin{bmatrix} U_n & U_r \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_r \end{bmatrix}^{1/2} \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_r \end{bmatrix}^{1/2} \Sigma_{x_d(t)u_t^+} \Sigma_{u_t^+ u_t^+}^{-1}.
\end{aligned}
$$

---

[1] The entity $\overline{x}(t)$ in Figure 4.2 is hence given by $\overline{x}(t) = \hat{E}\{x(t)|\mathcal{U}_t^+\} = \hat{E}\{x_d(t)|\mathcal{U}_t^+\}$.

Thus, the smaller the order is chosen, the more singular values of considerable modulus are contained in Σ_r and R grows larger. Another effect that influences the residual is the covariance of the past and future of u, as it enters the equation through $\Sigma_{x_d(t)u_t^+}$.

In contrast to the orthogonal decomposition of the output, methods decomposing the data as described in Theorem 2.13, do not suffer from this problem. The oblique decomposition of the output according to

$$\hat{E}_{\|\mathcal{U}_t^+}\{y_t^+|\mathcal{P}_t^-\} = \mathcal{O}_k\mathcal{C}_\infty p_t^- = \mathcal{O}_k x(t) \ ,$$
$$\hat{E}_{\|\mathcal{P}_t^-}\{y_t^+|\mathcal{U}_t^+\} = \Psi_k^d u_t^+$$

supersedes any secondary calculations for the extraction of Ψ_k^d. In fact, as shown in the right part of Figure 4.2, the oblique decomposition and orthogonal decomposition differ only in terms of the recovery of the data vectors within the space of the past inputs and past outputs or the space of the future inputs. Following Lemma 2.3, the oblique projection yielding $\mathcal{O}_k x(t)$ is given by

$$
\begin{aligned}
\hat{E}_{\|\mathcal{U}_t^+}\{y_t^+ \mid \mathcal{P}_t^-\} =& E\left\{\hat{E}\left\{y_t^+ \left|(\mathcal{U}_t^+)^\perp\right.\right\}\hat{E}\left\{p_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}^\mathrm{T}\right\} \\
& \times E\left\{\hat{E}\left\{p_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}\hat{E}\left\{p_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}^\mathrm{T}\right\}^{-1} p_t-
\end{aligned}
\tag{4.6}
$$

whereas the orthogonal projection on the complement of the past is given by

$$
\begin{aligned}
\hat{E}\left\{y_t^+ \left|\hat{E}\left\{\mathcal{P}_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}\right.\right\} =& E\left\{\hat{E}\left\{y_t^+ \left|(\mathcal{U}_t^+)^\perp\right.\right\}\hat{E}\left\{p_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}^\mathrm{T}\right\} \\
& \times E\left\{\hat{E}\left\{p_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}\hat{E}\left\{p_t^- \left|(\mathcal{U}_t^+)^\perp\right.\right\}^\mathrm{T}\right\}^{-1} \tilde{p}_t^- \ ,
\end{aligned}
\tag{4.7}
$$

where $\tilde{p}_t^- = p_t^- - \hat{E}\left\{p_t^- \left|\mathcal{U}_t^+\right.\right\}$. As both equations show, the difference is only in terms of the data vector p_t^-. It is the missing $\hat{E}\left\{p_t^- \left|\mathcal{U}_t^+\right.\right\}$ in (4.7) which reappears in (4.1) in terms of $\mathcal{O}_k\Sigma_{x_d(t)u_t^+}\Sigma_{u_t^+u_t^+}^{-1}u_t^+$.

Another advantage of the oblique decomposition is the possibility of directly estimating the state. Once the states $x(t)$ and $x(t+1)$ are estimated, the system matrices are calculated through (2.114). Possible estimation errors, which may affect $\hat{x}(t)$ and $\hat{x}(t+1)$ due to the linearization (see Remark 4.4) of the data of a nonlinear system, are mitigated by the joint least-squares calculation of the system matrices based on the original data ($y(t)$ and $u(t)$). Thus, the influence due to the violation regarding the assumption of linearity, which the methods are derived under, will be reduced, so that the identification of linear low-order models of an operating point of nonlinear systems will not be an issue.

Remark 4.1. It should be emphasized that the problem outlined above does not reside within the ORT approach. It is rather an implementation-based issue of the MOESP approach. In fact, the derivation of the ORT approach as given in Katayama (2005) uses a two-step strategy, which includes the optimal predictor of Theorem 2.13. Subsequent to the orthogonal projection (2.140),

$$\mathcal{Y}_t^{d,+} = \hat{E}\{\mathcal{Y}_t^+|\mathcal{U}_t^- \vee \mathcal{U}_t^+\} \ ,$$

$$\mathcal{Y}_t^{s,-} \vee \mathcal{Y}_t^{s,+} = \mathcal{Y}^s$$

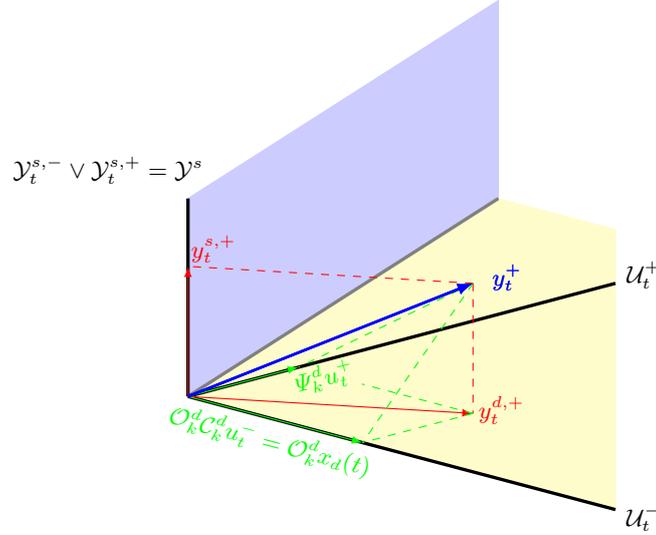$$\mathcal{O}_k^d \mathcal{C}_k^d u_t^- = \mathcal{O}_k^d x_d(t)$$

Figure 4.3: Extraction of a deterministic state estimate

which removes the disturbance effects, the predictor space of the deterministic component of the system is calculated as

$$\mathcal{X}_t^{d,+/-} = \hat{E}_{||\mathcal{U}_t^+}\{\mathcal{Y}_t^{d,+}|\mathcal{U}_t^-\} \ .$$

The claim that the problems are caused by the numerical implementation based on a MOESP-like algorithm of the ORT method, is backed up by the results of Chapter 5. Here, the CCA-based ORT algorithm proposed in the following subsection shows the best overall performance.

4.1.2 New algorithm for the orthogonal decomposition approach based on canonical correlation analysis

The consequence of the discussion of the previous subsection is the idea of combining the orthogonal decomposition of the ORT approach with the calculations of the optimal predictor. This points to a state estimation which includes the orthogonal decomposition and in turn yields the deterministic state of the system. As outlined in Section 2.6.2 by Theorem 2.15, the identification of the deterministic subsystem by the ORT approach is based on the fact that the orthogonal projection

$$y_d(t) = \hat{E}\{y(t) \,|\, \mathcal{U}\}$$

yields a causal estimation of the deterministic component of $y(t)$, i.e., $\hat{E}\{y(t) \,|\, \mathcal{U}\} = \hat{E}\{y(t) \,|\, \mathcal{U}_{t+1}^-\}$. Hence, the oblique projection

$$\mathcal{X}_t^{d,+/-} = \hat{E}_{||\mathcal{U}_t^+}\{\mathcal{Y}_t^{d,+}|\mathcal{U}_t^-\}$$

gives the deterministic predictor space. This relation is illustrated in Figure 4.3. Further inspection shows also

$$\mathcal{Y}_t^{s,+} \perp \mathcal{U}_t^+ \vee \mathcal{U}_t^- \ . \tag{4.8}$$

Together with the fact that

$$\hat{E}_{||\mathcal{C}}\{\mathcal{A}_1 \oplus \mathcal{A}_2|\mathcal{B}\} = \hat{E}_{||\mathcal{C}}\{\mathcal{A}_1|\mathcal{B}\} \tag{4.9}$$

if $\mathcal{A}_2 \perp \mathcal{B} \vee \mathcal{C}$, the predictor space $\mathcal{X}_t^{d,+/-}$ can be directly calculated.

Proposition 4.1. *(Bathelt, Söffker, and Jelali, 2015) The combined orthogonal decomposition of the future output and calculation of the deterministic predictor space is given by*

$$\mathcal{X}_t^{d,+/-} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+ | \mathcal{U}_t^-\} \, . \tag{4.10}$$

Proof. An element $y(t+h)$, $h = 0, 1, \cdots, k$, of the future outputs y_t^+ can be represented by

$$y(t+h) = \sum_{i=-\infty}^{t+h} G_{t+h-i} u(i) + \sum_{i=-\infty}^{t+h} H_{t+h-i} e(i)$$
$$= y_d(t+h) + y_s(t+h) \, ,$$

where

$$G_t = \begin{cases} D & t = 0 \\ C_d A_d^{t-1} B & t > 0 \end{cases} , \qquad H_t = \begin{cases} I & t = 0 \\ C_s A_s^{t-1} K & t > 0 \end{cases} .$$

Due to the orthogonality of u and e, $y_d \perp y_s$ follows. Based on the decomposition $\mathcal{P}_t^- = \mathcal{U}_t^- \oplus \mathcal{Y}_t^{s,-}$, the oblique projection $\hat{E}_{\|\mathcal{U}_t^+}\{y(t+h) \mid \mathcal{P}_t^-\}$ can be split into a deterministic and a stochastic part as

$$
\begin{aligned}
\hat{E}_{\|\mathcal{U}_t^+}\{y(t+h) \mid \mathcal{P}_t^-\} =& \hat{E}_{\|\mathcal{U}_t^+}\{y_d(t+h) + y_s(t+h) \mid \mathcal{U}_t^- \oplus \mathcal{Y}_t^{s,-}\} \\
=& \hat{E}_{\|\mathcal{U}_t^+}\{y_d(t+h) + y_s(t+h) \mid \mathcal{U}_t^-\} \\
& + \hat{E}_{\|\mathcal{U}_t^+}\{y_d(t+h) + y_s(t+h) \mid \mathcal{Y}_t^{s,-}\} \\
=& \hat{E}_{\|\mathcal{U}_t^+}\{y_d(t+h) \mid \mathcal{U}_t^-\} + \hat{E}\{y_s(t+h) \mid \mathcal{Y}_t^{s,-}\} \, .
\end{aligned}
\tag{4.11}
$$

As $\mathcal{U}_t^+ \perp \mathcal{Y}_t^{s,-}$, the second oblique projection becomes an orthogonal projection. The oblique projection $\hat{E}_{\|\mathcal{U}_t^+}\{y_d(t+h)|\mathcal{U}_t^-\}$ yields

$$
\begin{aligned}
\hat{E}_{\|\mathcal{U}_t^+}\{y_d(t+h)|\mathcal{U}_t^-\} &= C_d A_d^h \sum_{i=-\infty}^{t-1} A_d^{t-1-i} B u(i) \\
&= C_d A_d^h x_d(t) \\
&= y_d^-(t+h) \, .
\end{aligned}
$$

The elements $y_d^-(t+h)$, $h = 0, 1, \cdots, k$, span the deterministic predictor space $\mathcal{X}_t^{d,+/-}$. The orthogonal projection of the last equation in (4.11) defines the predictor space of the stochastic subsystem, i.e., $\mathcal{X}_t^{s,+/-}$. $\qquad \square$

Remark 4.2. Based on the interpretation of projections as estimations of high-order ARX models, as made for example in Jansson and Wahlberg (1996), Ljung and McKelvey (1996a), and Qin, Lin, and Ljung (2005), projection (4.10) resembles the estimation of a FIR (finite impulse response) model. In the standard algorithms, the inclusion of $\mathcal{Y}_t^-$ in the past data is only needed for the identification of the stochastic subsystem. This in turn gives rise to the autoregressive part of the high-order ARX model. In terms of the deterministic part of $\mathcal{Y}_t^-$, i.e., for $\mathcal{Y}_t^{d,-} = \hat{E}\{\mathcal{Y}_t^- | \mathcal{U}_t^-\}$, the inclusion $\mathcal{Y}_t^{d,-} \subset \mathcal{U}_t^-$ holds. That is, $\mathcal{Y}_t^{d,-}$ contains no additional information regarding the past if $\mathcal{U}_t^-$ is given. Thus, $\mathcal{Y}_t^{d,-}$ is not needed for the

prediction of the future of the deterministic outputs $y_t^{d,+}$, and hence the autoregressive part is redundant and the moving average part over $\mathcal{U}_t^-$ remains.

The calculation of the state $x_d(t)$ is based on the CCA method of Katayama and Picci (1999). In the context of the here proposed ORT algorithm, the data used for the underlying canonical correlation analysis needs to be changed according to the differences between the standard projection $\hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+|\mathcal{P}_t^-\}$ and the projection $\hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+|\mathcal{U}_t^-\}$ of Proposition 4.1. Thus, the conditional covariance matrices are now based only on the past input (instead of the past input *and* past output) and are given by

$$
\begin{aligned}
\Sigma_{y_t^+ y_t^+|u_t^+} &= E\left\{\left(y_t^+\left|\mathcal{U}_t^{+\perp}\right.\right)\left(y_t^+\left|\mathcal{U}_t^{+\perp}\right.\right)^{\mathrm{T}}\right\}, \\
\Sigma_{y_t^+ u_t^-|u_t^+} &= E\left\{\left(y_t^+\left|\mathcal{U}_t^{+\perp}\right.\right)\left(u_t^-\left|\mathcal{U}_t^{+\perp}\right.\right)^{\mathrm{T}}\right\}, \\
\Sigma_{u_t^- u_t^-|u_t^+} &= E\left\{\left(u_t^-\left|\mathcal{U}_t^{+\perp}\right.\right)\left(u_t^-\left|\mathcal{U}_t^{+\perp}\right.\right)^{\mathrm{T}}\right\}.
\end{aligned}
\tag{4.12}
$$

With the Cholesky factorizations

$$
\Sigma_{y_t^+ y_t^+|u_t^+} = LL^{\mathrm{T}}, \qquad \Sigma_{u_t^- u_t^-|u_t^+} = MM^{\mathrm{T}},
\tag{4.13}
$$

the SVD of the canonical correlation analysis is given by

$$
L^{-1}\Sigma_{y_t^+ u_t^-|u_t^+}M^{-\mathrm{T}} = U\Sigma V^{\mathrm{T}} = \begin{bmatrix} U_n & U_r \end{bmatrix}\begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_r \end{bmatrix}\begin{bmatrix} V_n^{\mathrm{T}} \\ V_r^{\mathrm{T}} \end{bmatrix},
\tag{4.14}
$$

and the estimate of a deterministic state of order n subsequently follows with

$$
\hat{x}_d(t) = \Sigma_n^{1/2}V_n^{\mathrm{T}}M^{-1}u_t^-.
\tag{4.15}
$$

Although the necessity of explicitly determining the state estimate $\hat{x}_d(t+1)$ is avoided by the numerical implementation of the CCA method (see following subsection), a possible approach to the coherent estimation of $\hat{x}_d(t)$ and $\hat{x}_d(t+1)$ can be outlined as follows (see, e.g., Chiuso, 2007b; Chiuso and Picci, 2004d). From (4.14), an estimate of the observability matrix is found as

$$
\hat{\mathcal{O}}_k = LU_n\Sigma_n^{1/2}.
$$

By construction $\hat{\mathcal{O}}_k$ is injective, i.e., the pair (A, C) of $\hat{\mathcal{O}}_k$ is observable. Thus, it has a left pseudo-inverse fulfilling $\hat{\mathcal{O}}_k^\dagger\hat{\mathcal{O}}_k = I_n$. Then, the state estimates are calculated by

$$
\begin{aligned}
\hat{x}_d(t) &= \hat{\mathcal{O}}_k^\dagger\Sigma_{y_t^+ u_t^-|u_t^+}\Sigma_{u_t^- u_t^-|u_t^+}u_t^-, \\
\hat{x}_d(t+1) &= \hat{\mathcal{O}}_k^\dagger\Sigma_{y_{t+1}^+ u_{t+1}^-|u_{t+1}^+}\Sigma_{u_{t+1}^- u_{t+1}^-|u_{t+1}^+}u_{t+1}^-,
\end{aligned}
$$

where y_{t+1}^+ contains $y(t+h), h = 1, \cdots, k+1$. The model is then determined by the regression

$$
\min_{A_d, B_d, C_d, D_d}\left\|\begin{bmatrix} \hat{x}_d(t+1) \\ y_d(t) \end{bmatrix} - \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix}\begin{bmatrix} \hat{x}_d(t) \\ u(t) \end{bmatrix}\right\|_{\mathcal{H}}^2,
\tag{4.16}
$$

where $y_d(t)$ is the deterministic part of the output given by[2]

$$y_d(t) = \hat{E}\{y(t)|\mathcal{U}_{t+1}^-\} . \tag{4.17}$$

Remark 4.3. A formulation of the regression in terms of the usual output $y(t)$ would yield the same result as the stochastic part $y_s(t)$ of $y(t)$ is orthogonal to both $u(t)$ and $\hat{x}_d(t)$. That is, the regression

$$\min_{A_d,B_d,C_d,D_d} \left\| \begin{bmatrix} \hat{x}_d(t+1) \\ y(t) \end{bmatrix} - \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix} \begin{bmatrix} \hat{x}_d(t) \\ u(t) \end{bmatrix} \right\|_{\mathcal{H}}^2$$

can also be used. However, due to the lack of sufficiently accurate orthogonality in the case of numerical entities, the use of $y_d(t)$ is preferred for the numerical implementation.

Due to the use of the CCA approach and the direct calculation of the deterministic predictor space by an oblique projection, this algorithm of the ORT approach differs not only from the MOESP-based algorithm of Katayama (2005) but also from the algorithm of Picci and Katayama (1996b, pp. 157–158). The latter one uses an approach via an intersection between $\mathcal{U}_t^-$ and the extended future subspace spanned by $\hat{E}\{y_t^+|\mathcal{U}\}$ and $x(t)$ (see Lemma 4.2 in Picci and Katayama, 1996b). It is also noteworthy that the approach presented in Chiuso and Picci (2004d) is different, as it is based on the determination of the complementary state, i.e., the state resulting from an orthogonal projection of y_t^+ onto $u_t^- - \hat{E}\{u_t^-|\mathcal{U}_t^+\}$. This creates a state, which is orthogonal to u_t^+. However, by Corollary 2.1 and Katayama (2005, p. 282), this state is non-causal. Thus, a correct estimation of a model of the system is not possible as B_d and D_d can not be estimated.

4.1.3 Numerical implementation

In terms of the implementation, the data vectors u_t^- and y_t^+, u_t^+ are limited to the intervals $[t-k, t-1]$ and $[t, t+k-1]$. The interval length k might be chosen arbitrarily, but have to fulfill the conditions for informative experiments as outlined by Proposition 2.1 and in Chui and Maciejowski (2005). If u is a regular stochastic process, these conditions and the condition of $\mathcal{U}_t^+ \cap \mathcal{U} = \{0\}$, which is required for the oblique projection, are naturally met. The order n is assumed to be known or estimated by the methods discussed in Section 2.6.3. The algorithm hereafter referred to as CCA-ORT is based on the following steps:

1. Calculation of the QR decomposition

$$\begin{bmatrix} U_t^+ \\ U_t^- \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{21} & R_{22} & 0 \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{bmatrix}$$

---

[2]It should be kept in mind that $\mathcal{U}_t^-$ includes only the past of u, whereas $y(t)$ also depends on the current $u(t)$. Thus, the projection needs to use $\mathcal{U}_{t+1}^- = \mathcal{U}_t^- \vee \mathcal{U}_t$.

2. Calculation of the numerical equivalents of the conditional covariance matrices of equation (4.12) as

$$N\Sigma_{y_t^+ y_t^+ | u_t^+} \approx S_{y_t^+ y_t^+ | u_t^+} = R_{32} R_{32}^T + R_{33} R_{33}^T ,$$

$$N\Sigma_{y_t^+ u_t^- | u_t^+} \approx S_{y_t^+ u_t^- | u_t^+} = R_{32} R_{22}^T ,$$

$$N\Sigma_{u_t^- u_t^- | u_t^+} \approx S_{u_t^- u_t^- | u_t^+} = R_{22} R_{22}^T .$$

As the result of the QR decomposition is not only needed to determine the conditional sample covariance matrices but also $\hat{y}_N^d(t)$, a scaling with $\frac{1}{\sqrt{N}}$ as done in Katayama (2005) and Katayama and Picci (1999) to adjust $S_{y_t^+ y_t^+ | u_t^+}$, $S_{y_t^+ u_t^- | u_t^+}$, $S_{u_t^- u_t^- | u_t^+}$ to $\Sigma_{y_t^+ y_t^+ | u_t^+}$, $\Sigma_{y_t^+ u_t^- | u_t^+}$, $\Sigma_{u_t^- u_t^- | u_t^+}$ is not possible.

3. Calculation of the factorizations

$$S_{y_t^+ y_t^+ | u_t^+} = LL^T , \qquad S_{u_t^- u_t^- | u_t^+} = MM^T$$

using the SVDs

$$S_{y_t^+ y_t^+ | u_t^+} = U_L \Sigma_L V_L^T , \qquad S_{u_t^- u_t^- | u_t^+} = U_M \Sigma_M V_M^T$$

as

$$L = V_L \Sigma_L^{-1/2} U_L^T , \qquad M = V_M \Sigma_M^{-1/2} U_M^T$$

4. Calculation of the SVD

$$L^{-1} S_{y_t^+ u_t^- | u_t^+} M^{-T} = U\Sigma V^T = \begin{bmatrix} U_n & U_r \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_r \end{bmatrix} \begin{bmatrix} V_n^T \\ V_r^T \end{bmatrix}$$

5. Determination of the state estimate

$$\hat{x}_N^d(t) = \Sigma_n^{1/2} V_n^T M^{-1} U_t^- .$$

It should be noted that this state estimate is $N^{-1/2}$ times smaller than the one given in (4.15). This is due to the use of $S_{u_t^- u_t^- | u_t^+}$ instead of $\Sigma_{u_t^- u_t^- | u_t^+}$. This will however only influence the estimation in terms of a similarity transformation or a scaling factor for B_d and C_d, which is hardly an issue as the calculated model represents just one of the countless possible realizations of the system. The calculation of the state estimate as

$$\hat{x}_N^{d,*}(t) = N^{1/2} \Sigma_n^{1/2} V_n^T M^{-1} U_t^-$$

is consistent with (4.15). This follows if the above numerical quantities are replaced by the conditional sample covariance matrices

$$\Sigma_{y_t^+ y_t^+ | u_t^+} \approx \tilde{S}_{y_t^+ y_t^+ | u_t^+} = \frac{1}{N} S_{y_t^+ y_t^+ | u_t^+}$$

$$\Sigma_{u_t^- u_t^- | u_t^+} \approx \tilde{S}_{u_t^- u_t^- | u_t^+} = \frac{1}{N} S_{u_t^- u_t^- | u_t^+}$$

and the factorizations (with relation to L and M)

$$\tilde{S}_{u_t^- u_t^- | u_t^+} = \tilde{L}\tilde{L}^T = \frac{1}{N^{1/2}} L \frac{1}{N^{1/2}} L^T \,, \qquad \tilde{S}_{u_t^- u_t^- | u_t^+} = \tilde{M}\tilde{M}^T = \frac{1}{N^{1/2}} M \frac{1}{N^{1/2}} M^T \,.$$

With these relations, it follows from $\tilde{M} \approx M$ and (4.15) that

$$\hat{x}_N^{d,*}(t) = \Sigma_n^{1/2} V_n^T \tilde{M}^{-1} U_t^-$$

is the actual numerical equivalent of $\hat{x}_d(t)$. Replacing $\tilde{M}$ by the above relation with respect to M gives the equation of $\hat{x}_N^{d,*}(t)$ in terms of M. Note that there is no scaling involved within the SVD of Step 4, as the respective factors are mutually canceled.

6. Determination of the numerical estimates of $x_d(t)$ and $x_d(t+1)$ by choosing $\hat{x}_{N-1}^d(t) = \hat{x}_N^d(t)(:,1:N-1)$ and $\hat{x}_{N-1}^d(t+1) = \hat{x}_N^d(t)(:,2:N)^3$. Such a partition of $x_N^d(t)$ yields coherent numerical estimates of $x_d(t)$ and $x_d(t+1)$. Observing that

$$\mathcal{C}_k = \Sigma_n^{1/2} V_n^T \tilde{M}^{-1}$$

is the deterministic (reversed) reachability matrix, the partition is equal to (assuming the order of the row vectors in U_t^- to be in accordance with $\mathcal{C}_k$)

$$N^{-1/2}\hat{x}_{N-1}^d(t) = \mathcal{C}_k U_t^-(:,1:N-1) \,,$$
$$N^{-1/2}\hat{x}_{N-1}^d(t+1) = \mathcal{C}_k U_t^-(:,2:N) = \mathcal{C}_k U_{t+1}^-(:,1:N-1) \,.$$

This reasoning can be as well extended to the joint stochastic-deterministic case.

7. Calculation of the orthogonal projection $\hat{E}\{y(t)|\mathcal{U}_{t+1}^-\}$ as

$$\hat{y}_{N-1}^d(t) = L_{31}(1:p,:)Q_1^T(:,1:N-1) + L_{32}(1:p,:)Q_2^T(:,1:N-1)$$

8. Solving the set of overdetermined equations

$$\begin{bmatrix} \hat{x}_{N-1}^d(t+1) \\ \hat{y}_{N-1}^d(t) \end{bmatrix} = \begin{bmatrix} A_d & B_d \\ C_d & D_d \end{bmatrix} \begin{bmatrix} \hat{x}_{N-1}^d(t) \\ u_{N-1}(t) \end{bmatrix}$$

in terms of a least-squares solution.

The respective MATLAB® implementation of the CCA-ORT algorithm can be found in Section B.1. This algorithm contains the estimation of both the deterministic model (as shown here) and the stochastic model. The order estimation is not included within the there shown implementation.

Remark 4.4. With regard to identifications of nonlinear systems like the Tennessee Eastman Process, the QR decomposition can be also seen as a linearization procedure. Considering the QR decomposition of Step 1, $L_{33}Q_3^T$ contains both the disturbances effects as well as linearization errors. As the rows of Q_3^T are orthogonal to the rows Q_1^T and Q_2^T, the linearization is given in terms of a least-squares fit. Thus, the QR decomposition produces essentially a high-order FIR model, or ARX model if P_t^- is used instead of U_t^-. In a second step, the

---

[3] The notation is used in the style of MATLAB®'s matrix notation, where $A(i:j,k:l)$ describes the submatrix of A consisting of rows i through j and columns k through l. The colon : itself indicates that either all rows or all columns are used.

model is extracted. This model extraction is basically a version of Ho-Kalman's method for the realization of deterministic systems (see Katayama, 2005). This becomes evident if the connection of the two conditional covariance matrices of the oblique projection (4.6) is considered. Assuming that the random vectors in p_t^- are lined up in a descending order, it follows that they are connected as

$$
E\left\{\hat{E}\left\{y_t^+\left|(\mathcal{U}_t^+)^\perp\right.\right\}\hat{E}\left\{p_t^-\left|(\mathcal{U}_t^+)^\perp\right.\right\}^\mathrm{T}\right\} = HE\left\{\hat{E}\left\{p_t^-\left|(\mathcal{U}_t^+)^\perp\right.\right\}\hat{E}\left\{p_t^-\left|(\mathcal{U}_t^+)^\perp\right.\right\}^\mathrm{T}\right\}\ ,\quad (4.18)
$$

where

$$
H = \begin{bmatrix} C\begin{bmatrix}B_K & K\end{bmatrix} & CA_K\begin{bmatrix}B_K & K\end{bmatrix} & CA_K^2\begin{bmatrix}B_K & K\end{bmatrix} & \cdots \\ CA\begin{bmatrix}B_K & K\end{bmatrix} & CAA_K\begin{bmatrix}B_K & K\end{bmatrix} & CAA_K^2\begin{bmatrix}B_K & K\end{bmatrix} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ CA^{k-1}\begin{bmatrix}B_K & K\end{bmatrix} & CA^{k-1}A_K^2\begin{bmatrix}B_K & K\end{bmatrix} & CA^{k-1}A_K^2\begin{bmatrix}B_K & K\end{bmatrix} & \cdots \end{bmatrix} = \mathcal{O}_k\mathcal{C}_\infty^K\ . \quad (4.19)
$$

The structure of H follows from the equations given in Section 2.2.1. Whereas the projections recover H, the singular value decomposition step is essentially used to decompose this matrix into $\mathcal{O}_k$ and $\mathcal{C}_\infty^K$, or rather $\mathcal{C}_k^K$. In terms of the proposed algorithm of the ORT approach, H is given by

$$
H = \begin{bmatrix} CB & CAB & CA^2B & \cdots \\ CAB & CA^2B & CA^3B & \cdots \\ CA^2B & CA^3B & CA^4B & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ CA^{k-1}B & CA^kB & CA^{k+1}B & \cdots \end{bmatrix} = \mathcal{O}_k\mathcal{C}_\infty^d\ .
$$

If the order of the random vectors in p_t^- or u_t^- is as in (2.6), $\mathcal{C}_\infty^K$ and $\mathcal{C}_\infty^d$ would be reversed reachability matrices and H would not exhibit the classical Hankel structure.

4.1.4 Examples

The two following examples are taken out of Katayama (2005) and should be considered as illustrative examples. The results are supposed to show that the proposed algorithm performs as stated and is not inferior to existing methods. The first example is an ARMAX system, i.e., the dynamics of the stochastic and deterministic subsystems are the same. The second example is a Box-Jenkins system, i.e., the dynamics of the stochastic and deterministic subsystems are disjoint.

This second example somewhat illustrates the worst case of a system identification. That is, the case of an identification with unknown and non-white disturbances acting on the output of the system with non-white input signals. Although one might argue that this is actually not a problem as the correct order of the joint stochastic-deterministic system can be estimates by methods explained in Section 2.6.3, those order estimation methods might fail as later shown by the identification study of Chapter 5. In such a case, it is necessary that the identification methods still perform its duty flawlessly.

ARMAX system

The ARMAX example, whose system structure is shown in Figure 4.4, is based on Case 2 of the examples in Katayama (2005, pp. 292–296). The system to be identified is given by

$$P(z) = \frac{B(z)}{A(z)} = \frac{0.0275z^{-4} + 0.0551z^{-5}}{1 - 2.3443z^{-1} + 3.081z^{-2} - 2.5274z^{-3} + 1.2415z^{-4} - 0.3686z^{-5}} \, .$$

The only difference with respect to the example in Katayama (2005) concerns the choice of the number of columns N and horizon lengths k, which are here set to $N = 2000$, $k = 30$. All other parameters remain the same. The input u is a white-noise sequence with unity variance. The disturbances v_1 and v_2 are white-noise sequences with $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.09$. The number of data sets is 100. Over the 100 simulations, the input is kept the same whereas the disturbances are different for each simulation. The methods/algorithms used for the identifications are the state algorithm of the CCA method, N4SID Algorithm 1, PO-MOESP, PI-MOESP, ORT (MOESP-based implementation), and CCA-ORT. The implementations of the algorithms are either based on the associated files of van Overschee and De Moor (1996) or as outlined in the appendices of Katayama (2005). The pre-defined order of the models is $n = 5$. As explained in Katayama (2005, p. 294), the disturbance of the system shown above are defined by

$$v(t) = H(z)e(t) = \begin{bmatrix} 1 & P(z) \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix} \, ,$$

and hence the noise filter $H(z)$ is a minimum-phase transfer function with the same poles/eigenvalues as $P(z)$.

The evaluations of identification results are made in terms of the models' eigenvalues, bode plots, and coefficients in Figure 4.5 through Figure 4.7. As expected from the structure of the system, both the eigenvalues in Figure 4.5 and the bode plots in Figure 4.6 show that the all methods can handle the given problem. This is also reflected by the boxplots of the coefficients of $P(z)$ in Figure 4.7. The often reported problem of insufficient identification of zeros by subspace methods presents itself by the relatively large dispersion of the numerator coefficients. Regarding the order estimation, which has been done in parallel for test purposes, the methods NIC and SVC yielded incorrect results if the penalty term $C(T)$ is not chosen correctly. By setting $C(T) = k_f k_p \log(T)$, the order has been correctly estimated in all identifications. In terms of the N4SID algorithms, it is however required to use the CVA weighting (see van Overschee and De Moor, 1996, p. 114). The method of Fujikoshi and Veitch (1979) has been able to correctly estimate the order in 60% of the identifications whereas the remainder of estimates has resulted in overestimations.
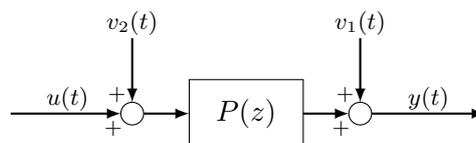

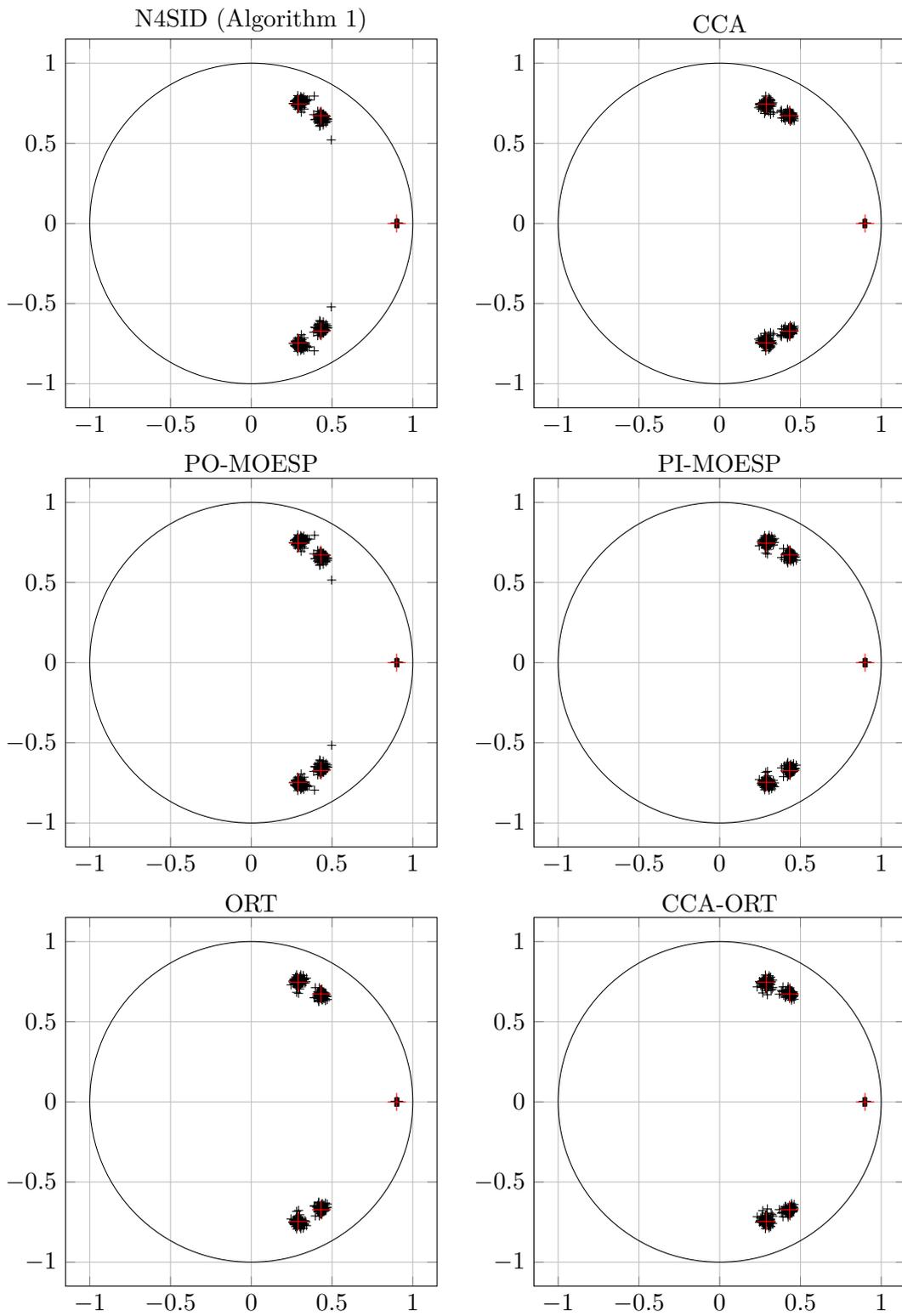
Figure 4.4: System structure of the ARMAX example

Figure 4.5: Eigenvalues of the identified models for the ARMAX system (red cross: nominal value of $P(z)$, black cross: value of identified models)
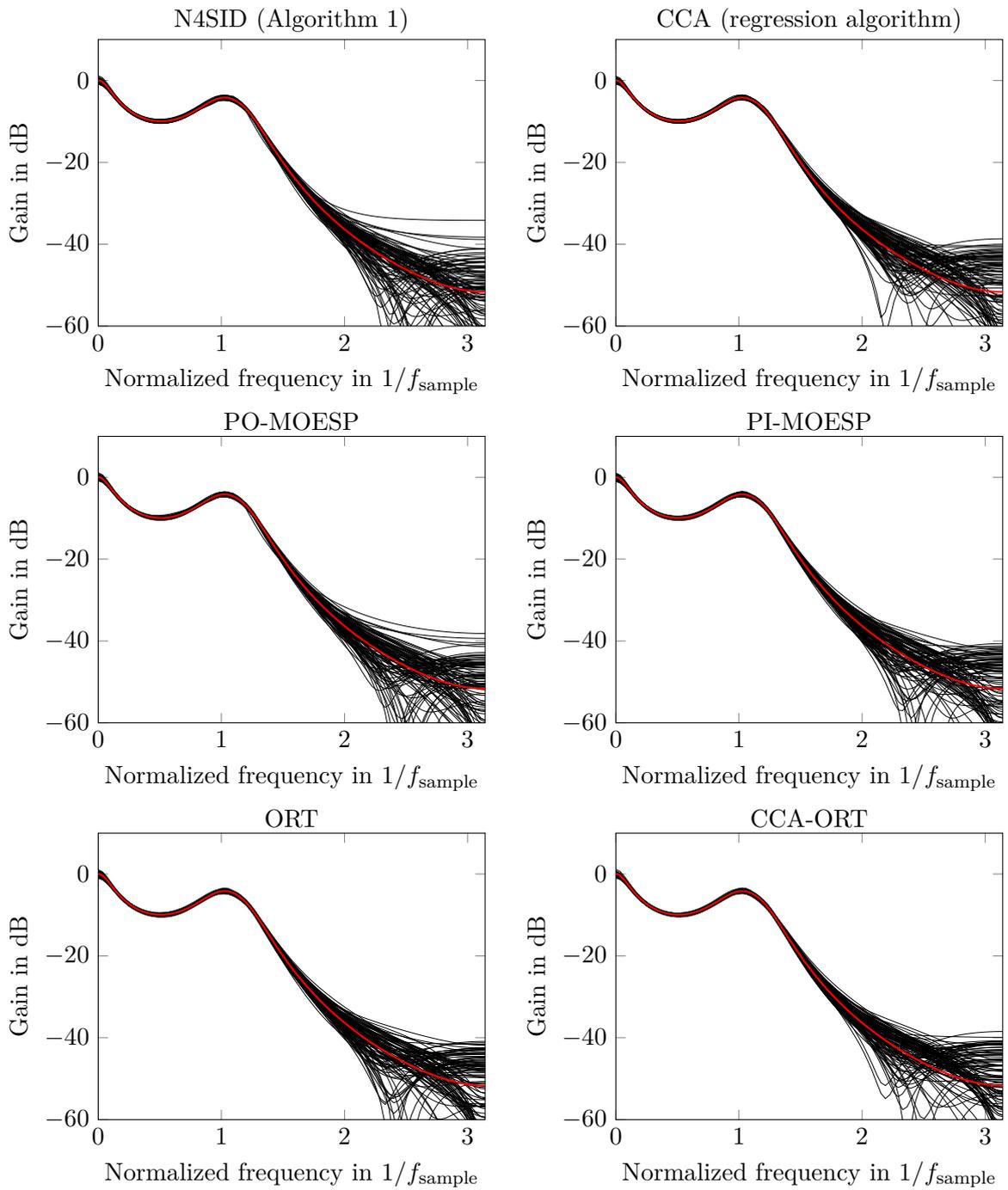
Figure 4.6: Bode gain plots of the identified models for the ARMAX system (red: nominal value of $P(z)$, black: value of the identified models)
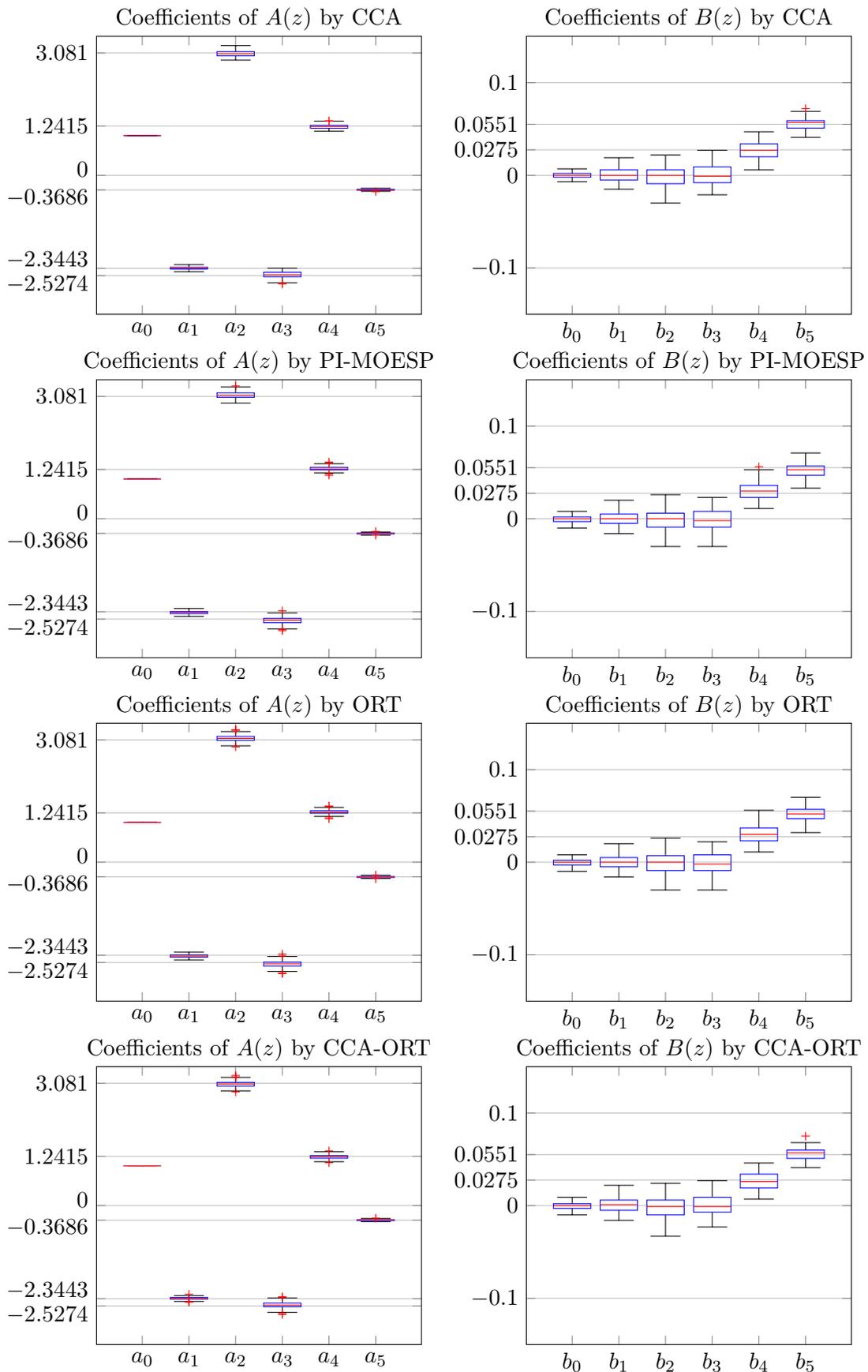
Figure 4.7: Coefficients of the transfer functions of the identified models for the ARMAX system (whisker length: 1.5 interquartile range)

Box-Jenkins system

The Box-Jenkins example, whose system structure is shown in Figure 4.8, is based on Case 2 of the examples in Katayama (2005, pp. 261–264). The system to be identified is again given by

$$P(z) = \frac{0.0275z^{-4} + 0.0551z^{-5}}{1 - 2.3443z^{-1} + 3.081z^{-2} - 2.5274z^{-3} + 1.2415z^{-4} - 0.3686z^{-5}} ,$$

whereas the input model $C(z)$ and the disturbance noise model $H(z)$ are described by

$$C(z) = \frac{\sqrt{1 - 0.9}}{1 - 0.9z^{-1}} , \qquad H(z) = \frac{1 - 0.2z^{-1} - 0.48z^{-2}}{1 + 0.4z^{-1} + 0.4z^{-2}} .$$

Again, the number of columns and horizon lengths are $N = 2000$ and $k = 30$. The white noise e_u has unity variance whereas e is adjusted to yield $\sigma^2_{y_s} = 0.01$. The input sequence is the same for all 100 simulations whereas the disturbance is varied for each simulation. The aim is the identification of $P(z)$ and hence the order of the model is set to $n = 5$. The methods/algorithms are again the state algorithm of the CCA method, N4SID Algorithm 1, PO-MOESP, PI-MOESP, ORT (MOESP-based implementation), and CCA-ORT.

The evaluation of the poles/eigenvalues of the resulting models in Figure 4.9 shows that not all methods are able to deal with the increased complexity of the example. By introducing a colored-noise disturbance with different dynamics, the specification of the model order by the order of the deterministic subsystem is not sufficient for methods identifying a joint stochastic-deterministic model (N4SID, CCA, PO-MOESP). In order to determine a model of order five, the actual joint model with an order of seven is reduced. Hence, considering also the poles of the stochastic system at $-0.2 \pm 0.6i$, the resulting poles of the model of order five are a blend between the seven poles of the joint system. As shown by the other diagrams, the ORT approach does not exhibit these problems, as the stochastic part y_s is suppressed during the identification. PI-MOESP takes a special position. Although yielding the same results as the orthogonal decomposition approach, the underlying IV approach is different; see the comment following the figures. Comparing the MOESP-based ORT algorithm and the CCA-ORT algorithm, CCA-ORT has a slight advantage in terms of the dispersion of the identified eigenvalues. A similar pattern presents itself in terms of the Bode gain plots of Figure 4.10. CCA-ORT is the best among the methods, whereas the methods for the joint identification show a clear bias in the high frequency range. The large scattering of the frequency response observed for ORT, CCA-ORT and PI-MOESP stem from the difference of the power spectra of y_d and y_s. In the high-frequency range, y_s holds considerably more power than u. For numerical reasons and due to finite horizons, the orthogonal decomposition is not completely exact and hence the scattering appears. The effects seen here are also addressed in Chiuso and Picci (1999, 2004a,d). For the joint identification, the boxplots of the coefficients in Figure 4.11 and Figure 4.12 finally depict the deviation of the coefficient of the models from
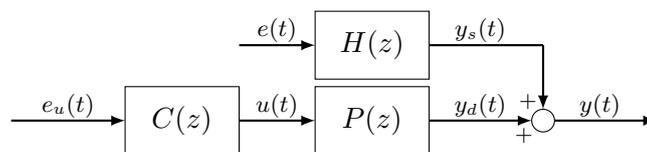

Figure 4.8: System structure of the Box-Jenkins example

the actual coefficients. Again, the CCA-ORT algorithm achieves a smaller coefficient variance than MOESP-based ORT algorithm.
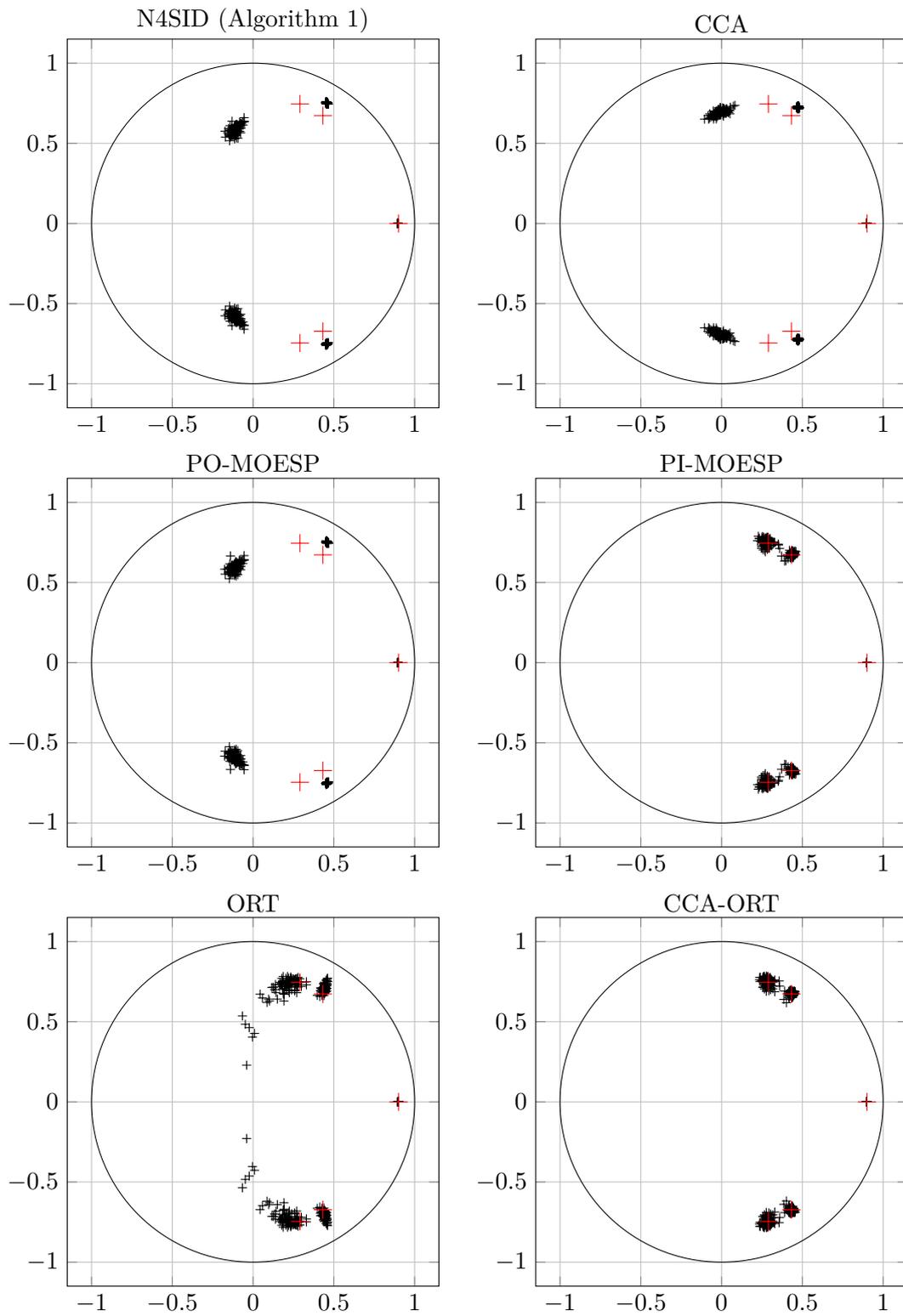
Figure 4.9: Eigenvalues of the identified models for deterministic subsystem of the Box-Jenkins system (red cross: nominal value of $P(z)$, black cross: value of identified models)
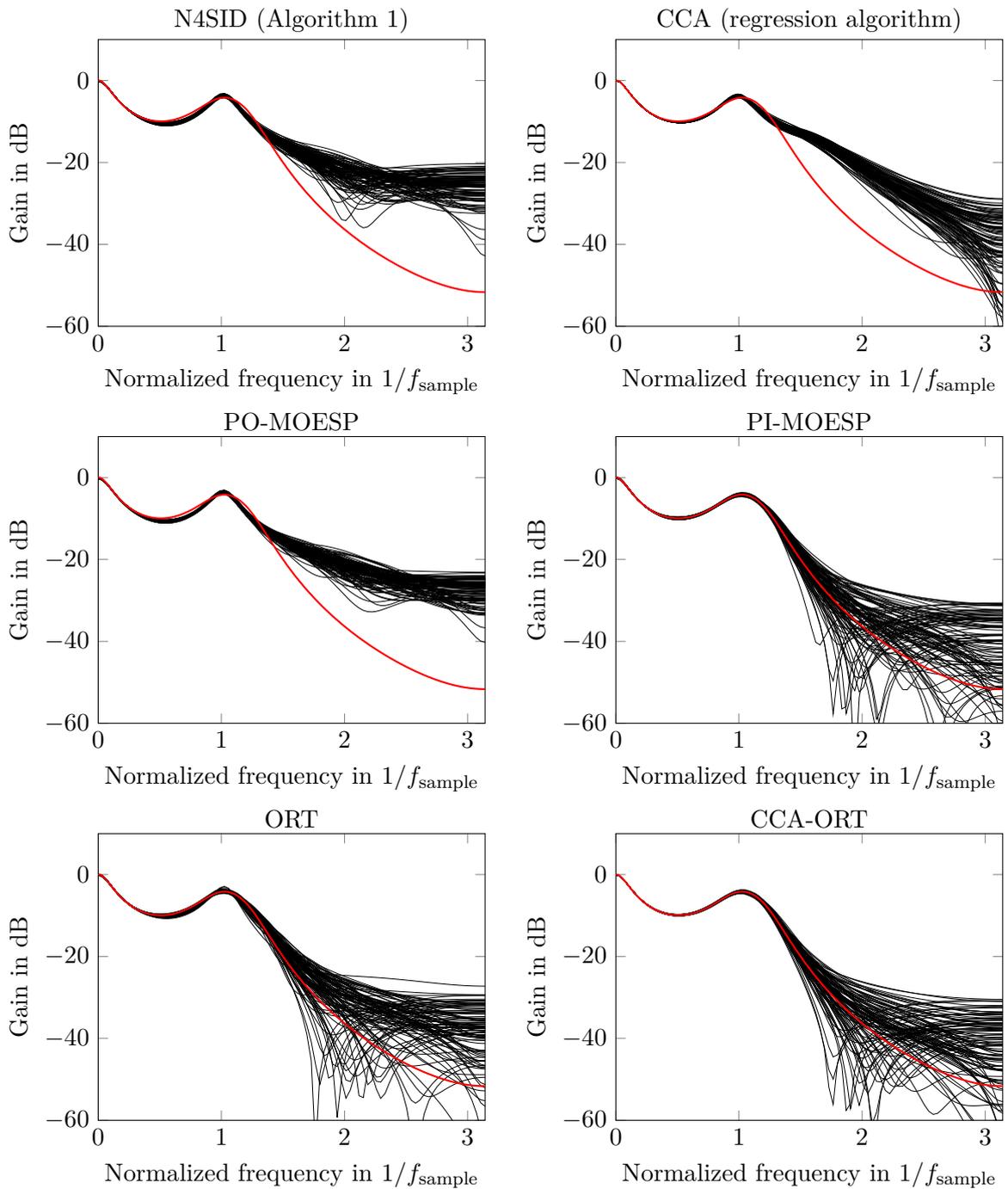
Figure 4.10: Bode gain plots of the identified models for deterministic subsystem of the Box-Jenkins system (red: nominal value of $P(z)$, black: value of identified models)
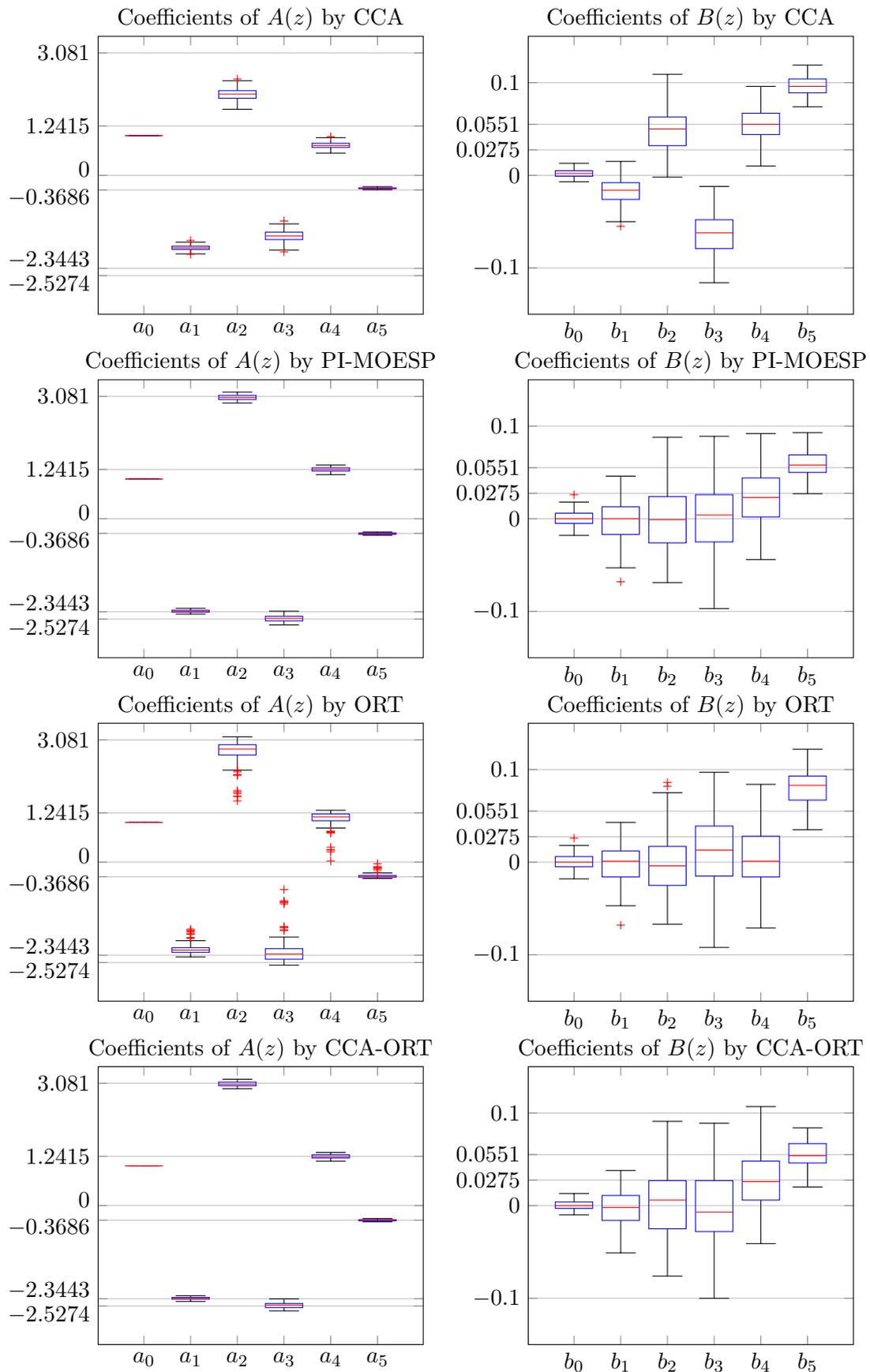
Figure 4.11: Coefficients of the transfer functions of the identified models for the deterministic subsystem of the Box-Jenkins system (whisker length: 1.5 interquartile range)
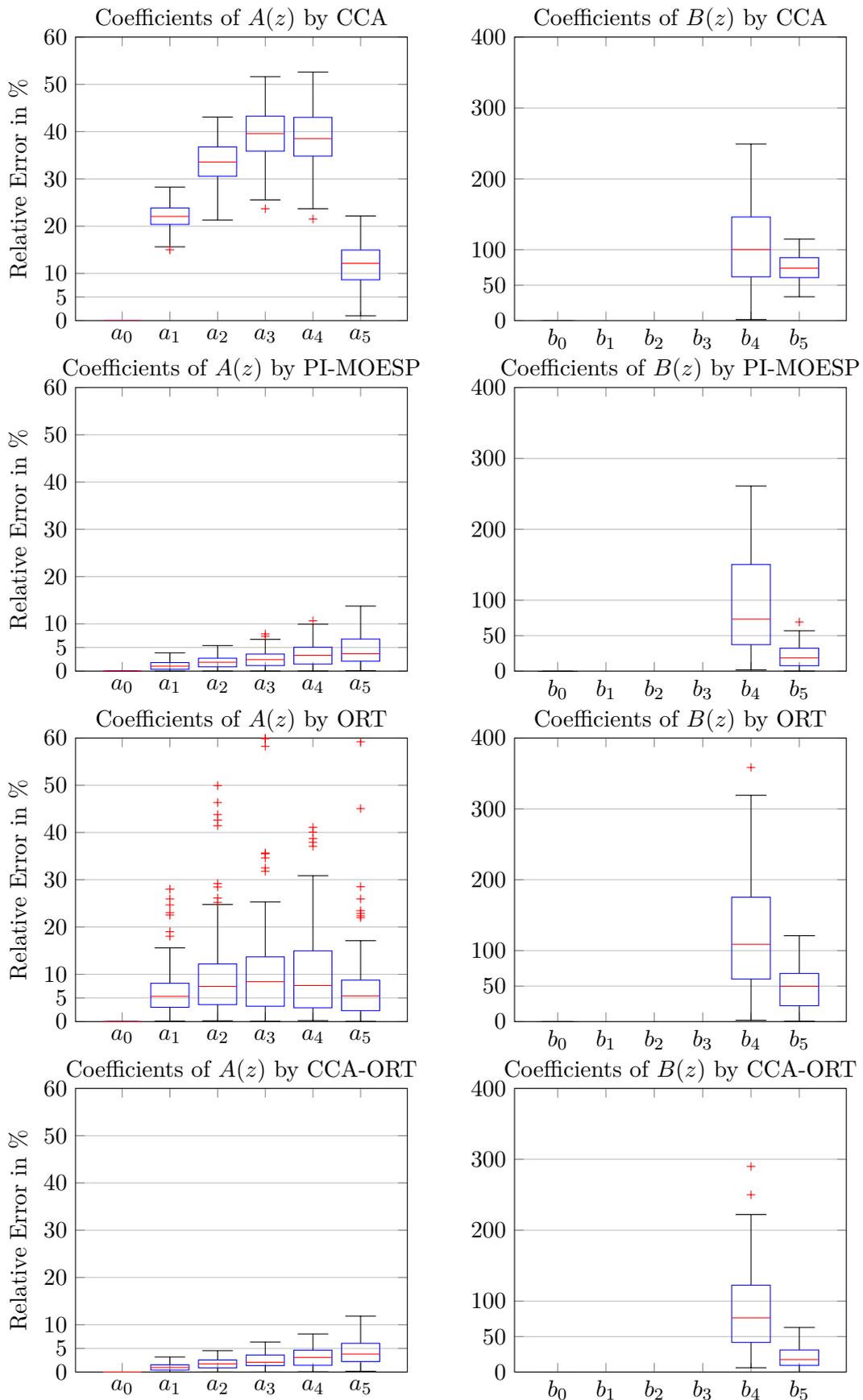
Figure 4.12: Relative error of the transfer function coefficients of the identified models for the deterministic subsystem of the Box-Jenkins system (whisker length: 1.5 interquartile range)

The good results of the PI-MOESP algorithm stem from the fact that it is essentially based on the SVD of $E\left\{\hat{E}\{y_t^+ \mid (\mathcal{U}_t^+)^\perp\}\hat{E}\{u_t^- \mid (\mathcal{U}_t^+)^\perp\}^\mathrm{T}\right\} \approx \Sigma_{y_f u_p \mid u_f}$ or, in terms of the QR decomposition of the CCA-ORT algorithm, is based on the SVD of $\mathrm{R}_{22}\mathrm{R}_{32}^\mathrm{T}$. Hence, the basis of the identifications by the CCA-ORT and PI-MOESP algorithms is the same. The difference is only in terms of the subsequent steps, which only influence the results if there is a system reduction. This should however not hide the fact that the PI-MOESP algorithm is an IV algorithm, which is bound to fail when the linearity of the system is dropped. Based on the results of the identification study on the Tennessee Eastman Process, the PI-MOESP algorithm is not suited for such systems, whereas the CCA-ORT algorithm is able to deal with such systems.

Regarding the order estimation, the result has been the same as for the ARMAX example. The correct order of the joint stochastic-deterministic system ($n = 7$) has been estimated for the joint models by both NIC/SVC and the method of Fujikoshi and Veitch (1979) (with the aforementioned 40% of over-estimations). In terms of the CCA-ORT method, the order $n_d = 5$ has been likewise recovered by the method of Fujikoshi and Veitch (1979).

4.1.5 Summary

In this section, an algorithm, called CCA-ORT, for the open-loop identification which combined the ORT approach with the CCA method is presented. This derivation of an new algorithm is motivated by the observation that the existing algorithm of the ORT approach is not able to cope with identifications of real systems like the one which is emulated by the Tennessee Eastman Process. The problem is traced back to the implementation based on the MOESP method (realization-based method), which does not decompose the future outputs based on two oblique projections onto the past data and the future inputs, but based on two orthogonal projections onto the future inputs and the respective complement of the past. This in turn is shown to introduce a bias within the estimation of the input and feed-through matrices if the assumed order of the model is not equal to the order of the system – a problem which is pronounced by nonlinearities of real systems. Modifying the oblique decompositions of the optimal predictor with respect to the idea of the ORT approach, the deterministic component of the state is estimated. The core calculation of the proposed CCA-ORT algorithm is given by the canonical correlation analysis, which orders the the dimensions of the data spaces with respect to the strongest correlation between the past and the future. An estimation of a state of order n based on this calculation contains hence the n most important data combinations from the past data. As the proposed algorithm is based on the state-regression approach, the integration of the recursive scheme is also eased. The examples based on SISO LTI systems already illustrate that the proposed algorithm achieves better results than its MOESP-based counterpart.

4.2 Basic algorithm for closed-loop identification

Similar to the open-loop case, a new identification algorithm for the closed-loop case is needed. The problem is however not related to the quality of the results or to a problem of a respective method itself as it was the case for the open-loop identification. Now, the problem is rather related to the algorithmic implementation of the methods. Among the closed-loop methods, the PBSID method is favored, as it is already based on the coordinate-free framework, which is

also the basis for the derivation of the recursive approach. Although its theoretical description allows for an easy integration of the recursive scheme, the algorithms highly deviates from the methodical description. Before the focus is turned to the derivation of a new algorithm for closed-loop identification, it should be noted that an inclusion of the ORT approach into the closed-loop method as done above for the CCA method is however not possible, as the basis of the ORT approach is the freedom of feedback.

The recursive methodology that will be proposed in Chapter 6 requires the methods to be able to use the past predictor space. Whereas this is not a major problem for open-loop methods, it is a problem if the algorithm of the closed-loop method PBSID (the same holds for the SSARX method) is considered. As discussed in Section 2.7.2, the method is based on a pre-identification of a high-order VARX model, whose parameters are approximately the Markov parameters of the predictor system (2.5). This algorithmic approach is based on the fact that the determination of the predictor space as in (2.148) and (2.149) with

$$
\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{k_f - 1} \mathcal{X}_t^h \, , \tag{4.20}
$$

where

$$
\mathcal{X}_t^h = \hat{E}_{\| \mathcal{P}_{[t,t+h)}^+} \{ y(t+h) \mid \mathcal{P}_t^- \} \, , \tag{4.21}
$$

is not possible if the usual QR decomposition of the data according to

$$
\begin{bmatrix} U_t^+ \\ P_t^- \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} \\ R_{21} \ R_{22} \\ R_{31} \ R_{32} \ R_{33} \end{bmatrix} Q^{\mathrm{T}} \tag{4.22}
$$

is used, see Chiuso (2007a). As the VARX-based algorithm follows the rationale of estimating the system's Markov parameters over a suitable horizon, it requires the associated past input-output data. Hence, a replacement of the majority of this past by some past predictor space is problematic. If however the VARX estimation can be avoided and the estimation of $\mathcal{X}_t^{+/-}$ can be made according to the theoretical expressions of (4.20) and (4.21), the incorporation of the recursive scheme into the resulting algorithm is possible.

As this problem is related to the numerical implementation of the oblique projections (4.21), another way of implementing these projections is needed. Such an idea was presented in van Overschee and De Moor (1996, p. 167), which is subsequently borrowed and applied to the problem at hand. The idea outlined in van Overschee and De Moor (1996) is based on the fact that an oblique projection is merely a part of a superordinate orthogonal projection. In fact, the coefficients, which govern the orthogonal projection[4]

$$
\hat{E}\{Y_t^+ \mid \mathcal{P}_t^- \vee \mathcal{U}_t^+\} = \begin{bmatrix} L_{\mathcal{P}_t^-} & L_{\mathcal{U}_t^+} \end{bmatrix} \begin{bmatrix} P_t^- \\ U_t^+ \end{bmatrix} \, ,
$$

---

[4]The matrices $L_{\mathcal{P}_t^-}$ and $L_{\mathcal{U}_t^+}$ remain italic as they are numerically also the same in terms of the calculation based on stochastic values. The upright notation is used only for numerical values which replace the stochastic values.

can be used to define the respective oblique projections by

$$\hat{E}_{\|\mathcal{U}_t^+}\{Y_t^+ \mid \mathcal{P}_t^-\} = L_{\mathcal{P}_t^-}P_t^-, \qquad \hat{E}_{\|\mathcal{P}_t^-}\{Y_t^+ \mid \mathcal{U}_t^+\} = L_{\mathcal{U}_t^+}U_t^+.$$

Hence, defining the QR decomposition as

$$\begin{bmatrix} P_t^- \\ U_t^+ \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} & & \\ R_{21} & R_{22} & \\ R_{31} & R_{32} & R_{33} \end{bmatrix} Q^T,$$

the coefficients $L_{\mathcal{P}_t^-}$ and $L_{\mathcal{U}_t^+}$ are given by

$$\begin{bmatrix} L_{\mathcal{P}_t^-} & L_{\mathcal{U}_t^+} \end{bmatrix} = \begin{bmatrix} R_{31} & R_{32} \end{bmatrix} \begin{bmatrix} R_{11} & \\ R_{21} & R_{22} \end{bmatrix}^{-1}.$$

The inversion is replaced by a pseudo inversion if the matrix is rank-deficient. It is however assumed in the following that this is not case, which is a sound assumption as u is assumed to be a regular stochastic process.

Using this approach and changing the data arrangement of the matrix of the left-hand side of (4.22), it will be shown that an implementation of (4.20) and (4.21) is possible without compromising the causality of the estimator. The resulting advantage is the omission of the VARX estimation and hence the feasibility of incorporating the recursive scheme into the resulting algorithm.

4.2.1 Restructuring of the QR decomposition for the predictor-based subspace identification method

The following discussion is based on the assumption that the system has no feed-through, i.e., $D = 0$. To this end, consider the numerical data matrix and its associated QR decomposition to be structured as

$$\begin{bmatrix} P_t^- \\ p_N(t) \\ p_N(t+1) \\ \vdots \\ p_N(t+k_f-1) \end{bmatrix} = \begin{bmatrix} R_{11} & & & & \\ R_{21} & R_{22} & & & \\ R_{31} & R_{32} & R_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ R_{(1+k_f)1} & R_{(1+k_f)2} & R_{(1+k_f)3} & \cdots & R_{(1+k_f)(1+k_f)} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ \vdots \\ Q_{1+k_f}^T \end{bmatrix}, \qquad (4.23)$$

where

$$p_N(t) = \begin{bmatrix} y_N(t) \\ u_N(t) \end{bmatrix}.$$

The dimensions of the matrices are $R_{11} \in \mathbb{R}^{k_p(m+p) \times k_p(m+p)}$, $R_{i1} \in \mathbb{R}^{(m+p) \times k_p(m+p)}$, $i \geq 2$, and $R_{ij} \in \mathbb{R}^{(m+p) \times (m+p)}$, $i, j \geq 2$. The restructuring of the data matrix as in (4.23) follows from the consideration to facilitate the calculation of the orthogonal projections

$$\hat{E}\{y(t+h) \mid \mathcal{P}_t^- \vee \mathcal{P}_{[t,t+h)}^+\} = \hat{E}_{\|\mathcal{P}_{[t,t+h)}^+}\{y(t+h) \mid \mathcal{P}_t^-\} + \hat{E}_{\|\mathcal{P}_t^-}\{y(t+h) \mid \mathcal{P}_{[t,t+h)}^+\}$$
$$= L_{\mathcal{P}_t^-}^{t+h} p_t^- + L_{\mathcal{P}_{[t,t+h)}^+}^{t+h} p_{[t,t+h)}^+ , \tag{4.24}$$

which the oblique projections (4.21) are part of. The coefficients of these orthogonal projections are defined in terms of (4.23) as

$$L_{\mathcal{P}_t^-}^t = R_{21}^y R_{11}^{-1} ,$$

$$\begin{bmatrix} L_{\mathcal{P}_t^-}^{t+1} & L_{\mathcal{P}_t}^{t+1} \end{bmatrix} = \begin{bmatrix} R_{31}^y & R_{32}^y \end{bmatrix} \begin{bmatrix} R_{11} & \\ R_{21} & R_{22} \end{bmatrix}^{-1} ,$$

$$\begin{bmatrix} L_{\mathcal{P}_t^-}^{t+2} & L_{\mathcal{P}_{[t,t+2)}^+}^{t+2} \end{bmatrix} = \begin{bmatrix} R_{41}^y & R_{42}^y & R_{43}^y \end{bmatrix} \begin{bmatrix} R_{11} & & \\ R_{21} & R_{22} & \\ R_{31} & R_{32} & R_{33} \end{bmatrix}^{-1} , \tag{4.25}$$

$$\vdots$$

where the R_{ij}^y stem from the partition of the decompositions of the $p_N(t+h)$, $h = 0, 1, \cdots, k_f - 1$ in (4.23) with

$$p_N(t+h) = \begin{bmatrix} R_{(2+h)1} \cdots R_{(2+h)(1+h)} & R_{(2+h)(2+h)} & \overbrace{0_{(m+p)} \cdots 0_{(m+p)}}^{k_f - 1 - h \text{ times}} \end{bmatrix} Q^T$$

$$= \begin{bmatrix} \begin{bmatrix} R_{(2+h)1}^y \\ R_{(2+h)1}^u \end{bmatrix} \cdots \begin{bmatrix} R_{(2+h)(1+h)}^y \\ R_{(2+h)(1+h)}^u \end{bmatrix} & \begin{bmatrix} R_{(2+h)(2+h)}^y & 0_{p \times m} \\ & R_{(2+h)(2+h)}^u \end{bmatrix} & 0_{(m+p)} \cdots 0_{(m+p)} \end{bmatrix} Q^T$$

$$\tag{4.26}$$

where $R_{(2+h)1}^y \in \mathbb{R}^{p \times k_p(m+p)}$, $R_{(2+h)i}^y \in \mathbb{R}^{p \times (m+p)}$ for $i = 2, \cdots, 1+h$, $R_{(2+h)(2+h)}^y \in \mathbb{R}^{p \times p}$, and $R_{(2+h)i}^u \in \mathbb{R}^{m \times (m+p)}$ for $i = 2, \cdots, 2+h$. By concatenating these calculations, the state estimate can be retrieved.

Lemma 4.1. *Let the QR decomposition of data be given according to (4.23). Then, the estimation of $\mathcal{O}_k^K x_N(t)$ is given by*

$$\hat{\mathcal{O}}_k^K \hat{x}_N(t) = \begin{bmatrix} \hat{E}\{y_N(t) \mid \mathcal{P}_t^-\} \\ \hat{E}_{\|\mathcal{P}_t}\{y_N(t+1) \mid \mathcal{P}_t^-\} \\ \vdots \\ \hat{E}_{\|\mathcal{P}_{[t,t+k_f-1)}^+}\{y_N(t+k_f-1) \mid \mathcal{P}_t^-\} \end{bmatrix} = L_{\mathcal{P}_t^-} P_t^- , \tag{4.27}$$

where $k_f \geq n$ and $L_{\mathcal{P}_t^-}$ follows from

$$
\begin{bmatrix} L_{\mathcal{P}_t^-} & L_{\mathcal{P}_t^+} \end{bmatrix} = \begin{bmatrix} R_{21}^y & & & 0 \\ R_{31}^y & R_{32}^y & & \\ \vdots & \vdots & \ddots & \\ R_{(1+k_f)1}^y & R_{(1+k_f)2}^y & \cdots & R_{(1+k_f)(1+k_f-1)}^y \end{bmatrix}
$$

$$
\times \begin{bmatrix} R_{11} & & & 0 \\ R_{21} & R_{22} & & \\ \vdots & \vdots & \ddots & \\ R_{(k_f)1} & R_{(k_f)2} & \cdots & R_{(k_f)(k_f)} \end{bmatrix}^{-1} .
$$

(4.28)

The matrix $L_{\mathcal{P}_t^+}$ has zeros above and on the main diagonal, and hence the estimate $\mathcal{O}_K \mathbf{x}_N(t)$ is causal.

Proof. As seen from the applied sub-matrices of R of the QR decomposition of (4.23) in (4.25), one part of the causality of the calculation of the orthogonal projections is already secured by the use of the respective past data. Hence, to prove causality of the estimator, it just remains to show that all elements on and above the main diagonal of

$$
L_{\mathcal{P}_t^+} = \begin{bmatrix} L_{\mathcal{P}_t}^t & L_{\mathcal{P}_{t+1}}^t & \cdots & L_{\mathcal{P}_{t+k_f-2}}^t \\ L_{\mathcal{P}_t}^{t+1} & L_{\mathcal{P}_{t+1}}^{t+1} & \cdots & L_{\mathcal{P}_{t+k_f-2}}^{t+1} \\ \vdots & \vdots & \ddots & \vdots \\ L_{\mathcal{P}_t}^{t+k_f-2} & L_{\mathcal{P}_{t+1}}^{t+k_f-2} & \cdots & L_{\mathcal{P}_{t+k_f-2}}^{t+k_f-2} \\ L_{\mathcal{P}_t}^{t+k_f-1} & L_{\mathcal{P}_{t+1}}^{t+k_f-1} & \cdots & L_{\mathcal{P}_{t+k_f-2}}^{t+k_f-1} \end{bmatrix}
$$

are zero, i.e., $L_{\mathcal{P}_{t+k}}^{t+h} = 0,\ h = 0, 1, \cdots, k_f - 1,\ k \geq h$. This condition follows, as $L_{\mathcal{P}_t^+}$ has a similar meaning as $\Psi_k^{d,K}$ and $\Psi_k^{s,K}$ in (2.113). Note that $L_{\mathcal{P}_t^+}$ has k_f block rows but only $k_f - 1$ columns. This results, as the orthogonal projection (4.24) uses only data of the past, i.e., in terms of $y(t + k_f - 1)$ the last data set is $p(t + k_f - 2)$.

From the inversion of block matrices, the inversion of a lower triangular block matrix follows with

$$
\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix} .
$$

Hence, (4.28) can be rewritten as

$$
\begin{bmatrix} L_{\mathcal{P}_t^-} & L_{\mathcal{P}_t^+} \end{bmatrix} =
\begin{bmatrix}
\begin{pmatrix}
R_{21}^y & & & 0 \\
R_{31}^y & R_{32}^y & & \\
\vdots & \vdots & \ddots & \\
R_{(k_f)1}^y & R_{(k_f)2}^y & \cdots & R_{(k_f)(k_f-1)}^y
\end{pmatrix} & 0_{(p(k_f-1)\times p)} \\
\begin{pmatrix} R_{(1+k_f)1}^y & \cdots & R_{(1+k_f)(1+k_f-2)}^y \end{pmatrix} & R_{(1+k_f)(1+k_f-1)}^y
\end{bmatrix}
$$

$$
\times
\begin{bmatrix}
\begin{pmatrix}
R_{11} & & & 0 \\
R_{21} & R_{22} & & \\
\vdots & \vdots & \ddots & \\
R_{(k_f-1)1} & R_{(k_f-1)2} & \cdots & R_{(k_f-1)(k_f-1)}
\end{pmatrix}^{-1} & 0_{(k_p(p+m)+(k_f-1)p\times p+m)} \\
(\cdots)^{-1} & R_{(k_f)(k_f)}^{-1}
\end{bmatrix}^{-1}
$$

$$
=
\begin{bmatrix}
\left(L_{\mathcal{P}_t^-}^{[t,t+k_f-2]} \quad L_{\mathcal{P}_{[t,t+k_f-2)}^+}^{[t,t+k_f-2]} \right) & L_{\mathcal{P}_{t+k_f-2}^+}^{[t,t+k_f-2]} \\
\left(L_{\mathcal{P}_t^-}^{t+k_f-1} \quad L_{\mathcal{P}_{[t,t+k_f-2)}^+}^{t+k_f-1} \right) & L_{\mathcal{P}_{t+k_f-2}^+}^{t+k_f-1}
\end{bmatrix}.
$$

If follows that $L_{\mathcal{P}_{t+k_f-2}}^{[t,t+k_f-2]} = 0$. This matrix is constructed from coefficients $L_{\mathcal{P}_{t+k_f-2}}^{t+h}$, $h = 0, 1, \cdots k_f - 2$. Repeating this rationale for the sub-matrices in the parentheses, the required condition $L_{\mathcal{P}_{t+k}}^{t+h} = 0$, $h = 0, 1, \cdots, k_f - 1$, $k \geq h$ follows. Hence, the estimation of $\mathcal{O}_k^K \mathrm{x}_N(t)$ is causal. $\qquad\square$

Although the lower triangular structure in $L_{\mathcal{P}_t^+}$ is secured, the drawback of this algorithm is the loss of the Toeplitz structure of $L_{\mathcal{P}_t^+}$, which would however theoretically be given (cf. Chiuso, 2007a). The numerical calculation of the elements on the diagonals of $L_{\mathcal{P}_t^-}$ will not lead to the theoretically prescribed equality of these elements. As the numerical examples will show, this problem does however not compromise the results.

A deeper analysis of the meaning of the operation show that the fundamental idea of this numerical algorithm of the PBSID method is comparable to the idea of the PARSIM (open-loop method) or PARSIM-E (closed-loop method) methods as presented in Qin, Lin, and Ljung (2005) and Qin and Ljung (2003b). In terms of the PARSIM methods, causality is enforced by successive QR decompositions and calculations, which are basically implementations of (4.25). Regarding the closed-loop estimation by PARSIM-E, estimates of the innovations are also calculated, which are than reused in the next iteration. However, the approach proposed here only needs one global calculation, which is a consequence of the reordering of the data. The proposed approach can hence be seen as a continuation of the ideas of PARSIM and PARSIM-E. The innovation estimation is not needed, as the innovation process is implicitly estimated by the PBSID method by using the predictor system.

Remark 4.5. As the assumption of no feed-through is a build-in feature of the above proposed algorithm, systems having a feed-through can not be correctly identified. However, by

changing the data structure of the future part of the data matrix in the left-hand side of (4.23) to

$$p_N(t) = \left[u_N^T(t) \ y_N^T(t) \right]^T$$

and adding the row $R^u_{(1+k_f)i}$, $i = 1, \cdots 1 + k_f$ in (4.28), the feed-through can also be included. Note that the structure of the matrix consisting of R^y_{ij} also changes as the estimations of $\mathcal{X}^h_t$, $h = 0, \cdots, k_f - 1$ in

$$\mathcal{X}^{+/-}_t = \bigvee_{h=0}^{k_f-1} \mathcal{X}^h_t$$

are now given by

$$\mathcal{X}^h_t = \hat{E}_{\|(\mathcal{P}^+_{[t,t+h)} \vee \mathcal{U}_{t+h})}\{y(t+h) \mid \mathcal{P}^-_t\} \ ,$$

which in turn are based on the orthogonal projections $\hat{E}\{y(t+h) \mid \mathcal{P}^-_t \vee \mathcal{P}_{[t,t+h)} \vee \mathcal{U}_{t+h}\}$. That is, a term attributed to the additional subspace $\mathcal{U}_{t+h}$ needs to appear in each of the inverted matrices of (4.25) after the last block row of these matrices.

4.2.2 Numerical implementation

As stated for the implementation of the CCA-ORT algorithm, the data vectors u^-_t, y^-_t, and y^+_t, u^+_t are limited to the intervals $[t-k, t-1]$ and $[t, t+k-1]$. The requirements for the length are again given by the conditions for informative experiments as outlined in Chui and Maciejowski (2005). From (4.20), the length of the horizons has at least to fulfill $k > n$, where the order n is either known or estimated. The QR decomposition-based algorithm of the PBSID method, subsequently referred to as PBSID$_{QR}$, is then given by the following steps (the calculation of the Kalman gain is borrowed from the CCA algorithm in Katayama, 2005, pp. 291–292):

1. Construction of the numerical data matrix as in the left-hand side of (4.23) and calculation of the respective QR decomposition

2. Construction of the right-hand side matrices of (4.28), where the sub-matrices R^y_{ij} are extracted as defined in (4.26).

3. Calculation and extraction of $L_{\mathcal{P}^-_t}$, and calculation of the estimator of $\hat{\mathcal{O}}^K_k \hat{x}_N(t)$ as

$$L_{\mathcal{P}^-_t} R_{11} = \hat{\mathcal{O}}^K_k \hat{x}_N(t) Q_1 \ .$$

Note that Q_1^T is a matrix only containing the basis vectors of P^-_t. That is, R_{11} contains the necessary information regarding the actual structure of P^-_t and hence $L_{\mathcal{P}^-_t} R_{11}$ contains the information regarding $\hat{\mathcal{O}}^K_k \hat{x}_N(t)$. As Q_1^T is furthermore an orthogonal matrix, its inverse is given by its transpose Q_1. Hence,

$$L_{\mathcal{P}^-_t} P^-_t = L_{\mathcal{P}^-_t} R_{11} Q_1^T = \hat{\mathcal{O}}^K_k \hat{x}_N(t) \Rightarrow L_{\mathcal{P}^-_t} R_{11} = \hat{\mathcal{O}}^K_k \hat{x}_N(t) Q_1$$

4. Calculation of the SVD

$$L_{\mathcal{P}_t^-} R_{11} = U\Sigma V^T = \begin{bmatrix} U_n & U_r \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_r \end{bmatrix} \begin{bmatrix} V_n^T \\ V_r^T \end{bmatrix}$$

5. Determination of the state estimate

$$x_N(t) = \Sigma_n^{1/2} V_n^T Q_1^T$$

6. Determination of the numerical estimates of $x(t)$ and $x(t+1)$ by choosing $\hat{x}_N(t) = x_N(t)(:, 1 : N-1)$ and $\hat{x}_N(t+1) = x_N(t)(:, 2 : N)$[5]. The underlying reasoning of this determination is explained in Section 4.1.3.

7. Solving of the set of overdetermined equations

$$\begin{bmatrix} \hat{x}_{N-1}(t+1) \\ \hat{y}_{N-1}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{x}_{N-1}(t) \\ u_{N-1}(t) \end{bmatrix}$$

in terms of a least-squares solution under the constraint $D = 0$. By virtue of the Frobenius norm, this constraint can be directly included into the calculation as

$$\min_{A,B,C} \left\| \begin{bmatrix} x_{N-1}(t+1) \\ y_{N-1}(t) \end{bmatrix} - \begin{bmatrix} A & B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_{N-1}(t) \\ u_{N-1}(t) \end{bmatrix} \right\|_F^2$$

$$= \min_{A,B} \left\| x_{N-1}(t+1) - \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x_{N-1}(t) \\ u_{N-1}(t) \end{bmatrix} \right\|_F^2 + \min_C \left\| y_{N-1}(t) - C x_{N-1}(t) \right\|_F^2$$

8. Calculation of the residuals

$$\begin{bmatrix} w_{N-1}(t) \\ e_{N-1}(t) \end{bmatrix} = \begin{bmatrix} x_{N-1}(t+1) \\ y_{N-1}(t) \end{bmatrix} - \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & 0 \end{bmatrix} \begin{bmatrix} x_{N-1}(t) \\ u_{N-1}(t) \end{bmatrix}$$

and its error covariance matrices

$$\begin{bmatrix} \hat{\Sigma}_{ww} & \hat{\Sigma}_{we} \\ \hat{\Sigma}_{ew} & \hat{\Sigma}_{ee} \end{bmatrix} = \frac{1}{N-1} \begin{bmatrix} w_{N-1}(t)w_{N-1}^T(t) & w_{N-1}(t)e_{N-1}^T(t) \\ e_{N-1}(t)w_{N-1}^T(t) & e_{N-1}(t)e_{N-1}^T(t) \end{bmatrix}$$

9. Solving of the algebraic Riccati equation

$$P = \hat{A}P\hat{A}^T - (\hat{A}P\hat{C}^T + \hat{\Sigma}_{we})(\hat{C}P\hat{C}^T + \hat{\Sigma}_{ee})^{-1}(\hat{A}P\hat{C}^T + \hat{\Sigma}_{we})^T + \hat{\Sigma}_{ww}$$

---

[5]The notation is used in the style of MATLAB®'s matrix notation, where $A(i : j, k : l)$ describes the submatrix of A consisting of rows i through j and the columns k through l. The colon $:$ itself indicates that either all rows or all columns are used.

and calculation of the Kalman gain as

$$K = (\hat{A}P\hat{C}^{\mathrm{T}} + \hat{\Sigma}_{we})(\hat{C}P\hat{C}^{\mathrm{T}} + \hat{\Sigma}_{ee})^{-1}$$

The respective MATLAB® implementation of the QR decomposition-based algorithm of the PBSID method can be found in Section B.2.

Remark 4.6. If the identification of the feed-through matrix D is to be included in the algorithm, changes are in Steps 1 and 2 owing to the definition of the data as $\mathrm{p}_N(t) = \left[\mathrm{u}_N^{\mathrm{T}}(t)\ \mathrm{y}_N^{\mathrm{T}}(t)\right]^{\mathrm{T}}$. The other obvious modification is with the regression of Step 7, which then reads

$$\begin{bmatrix} \hat{A}\ \hat{B} \\ \hat{C}\ \hat{D} \end{bmatrix} = \arg \min_{A,B,C,D} \left\| \begin{bmatrix} \mathrm{x}_{N-1}(t+1) \\ \mathrm{y}_{N-1}(t) \end{bmatrix} - \begin{bmatrix} A\ B \\ C\ D \end{bmatrix} \begin{bmatrix} \mathrm{x}_{N-1}(t) \\ \mathrm{u}_{N-1}(t) \end{bmatrix} \right\|_F^2 .$$

4.2.3 Examples

The two numerical examples shall illustrate that the proposed algorithm is equal in comparison with the existing closed-loop methods. In particular, the comparison to the original PBSID$_{\mathrm{opt}}$ algorithm is of importance, as the proposed PBSID$_{\mathrm{QR}}$ algorithm is expected to yield the same results. The first example is an open-loop system based on the examples given in Katayama (2005). The second example is taken from Jansson (2003) and Ljung and McKelvey (1996b) and is a closed-loop example.

Open-loop example

As seen from the system structure in Figure 4.13, the open-loop example is based on an ARMAX system, whose transfer function is given by

$$P(z) = \frac{0.0275z^{-4} + 0.0551z^{-5}}{1 - 2.3443z^{-1} + 3.081z^{-2} - 2.5274z^{-3} + 1.2415z^{-4} - 0.3686z^{-5}} .$$

The example is again based on Case 2 of the examples in Katayama (2005, pp. 292–296). It is also equal to the first example of Section 4.1.4. Similar to that example, the number of columns N and horizon lengths k are set to $N = 2000$, $k = 30$. The disturbances v_1 and v_2 are white-noise sequences with $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.09$. In deviation from the examples in Katayama (2005) and in Section 4.1.4, the input u is a colored-noise sequence defined by

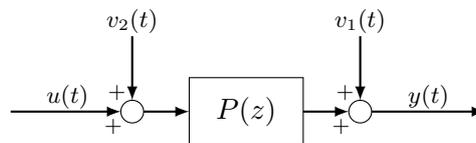$$C(z) = \frac{\sqrt{1 - 0.9}}{1 - 0.9z^{-1}}$$

Figure 4.13: System structure of the ARMAX example

and a white-noise process with unity variance. The number of data sets is 100. Over the 100 simulations, the input is kept the same, whereas the disturbances are different. The methods/algorithms used for the identification are the open-loop methods CCA (state algorithm), PO-MOESP, the closed-loop methods based on the explicit VARX estimation $PBSID_{opt}$ and SSARX, PARSIM-E and $PBSID_{QR}$. The implementations of the algorithms of the closed-loop methods are based on the explanations in Chiuso (2007b), Jansson (2003), and Qin and Ljung (2003a). For the VARX-based algorithms, the horizon length $k = 30$ defines also the number of past coefficients of the VARX model. The implementation of the $PBSID_{QR}$ algorithm which includes the identification of the feed-through matrix performed virtually the same in comparison to the primary implementation without estimation of D. Hence, the results of this implementation of the $PBSID_{QR}$ algorithm are omitted here.

As shown by the evaluation of the results in terms of the models' eigenvalues and Bode gain plots given in Figure 4.14 and Figure 4.15, all methods are able to identify the system. The deterioration of the results in comparison to the ARMAX example given in Section 4.1.4 stems from the input, which is here a colored-noise sequence. As, in comparison to the noise, less power is placed in the high-frequency range by the input, the identification becomes more difficult. This is illustrated by the frequency responses in Figure 4.15, showing a high scattering of the individual frequency responses (cf. Figure 4.6). Although this increased complexity of the identification problem resulted also in a higher spreading of the eigenvalues, the PBSID algorithms and SSARX show the smallest scattering (with $PBSID_{opt}$ dominating among those three). The apparent similarity of the results of SSARX and $PBSID_{QR}$ can be traced back to the fact that the calculations of both methods are – technically speaking – the same. Whereas SSARX first identifies an high-order VARX model and then removes the effects of the future form y_t^+, the same is done by $PBSID_{QR}$ in terms of projections. In contrast to those two algorithms, the algorithm of $PBSID_{opt}$ uses the VARX coefficients to construct $\mathcal{O}_{k_f}^K \mathcal{C}_{k_p}^K$.
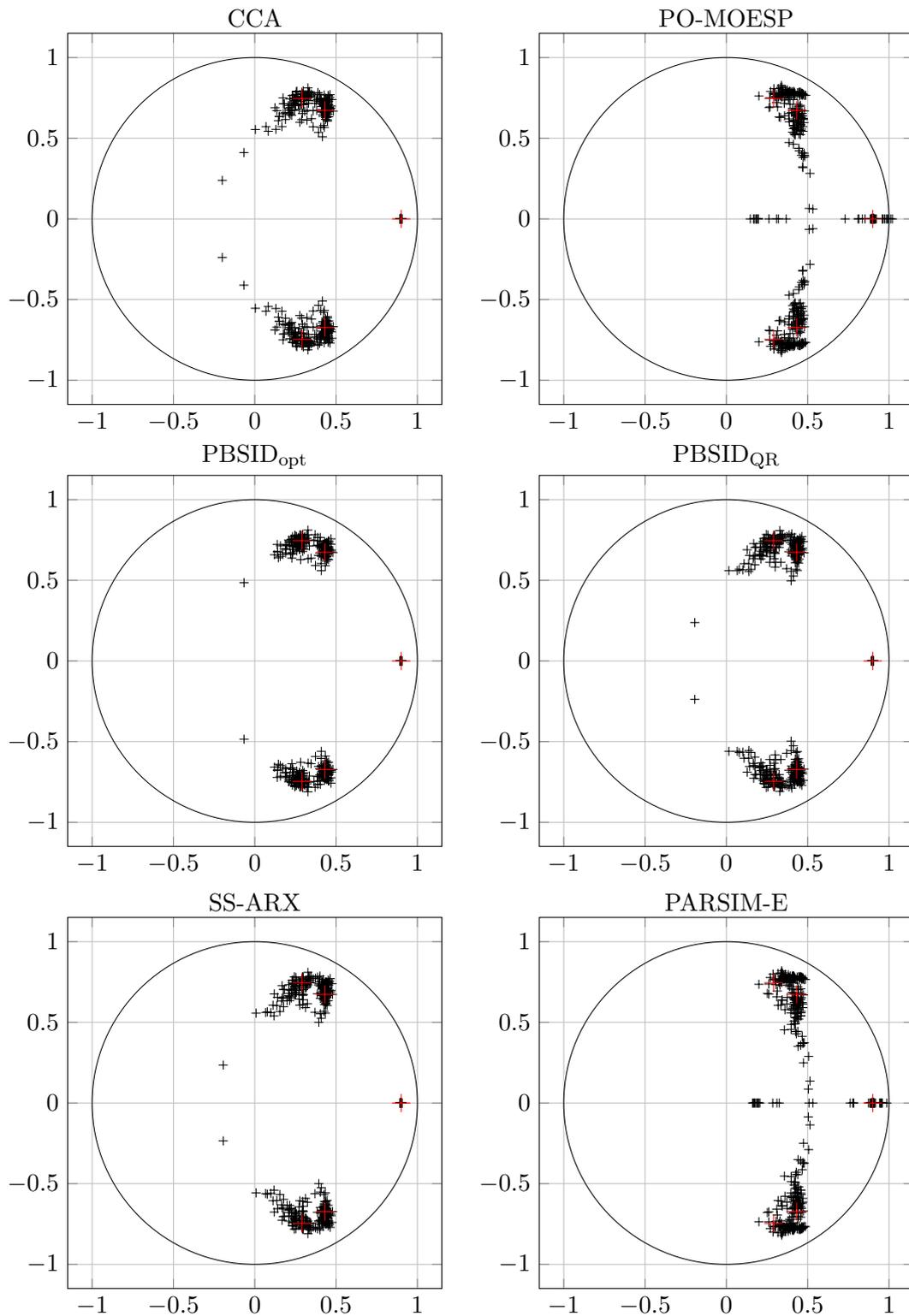
Figure 4.14: Eigenvalues of the identified models for the ARMAX system (red cross: nominal value of $P(z)$, black cross: value of identified models)
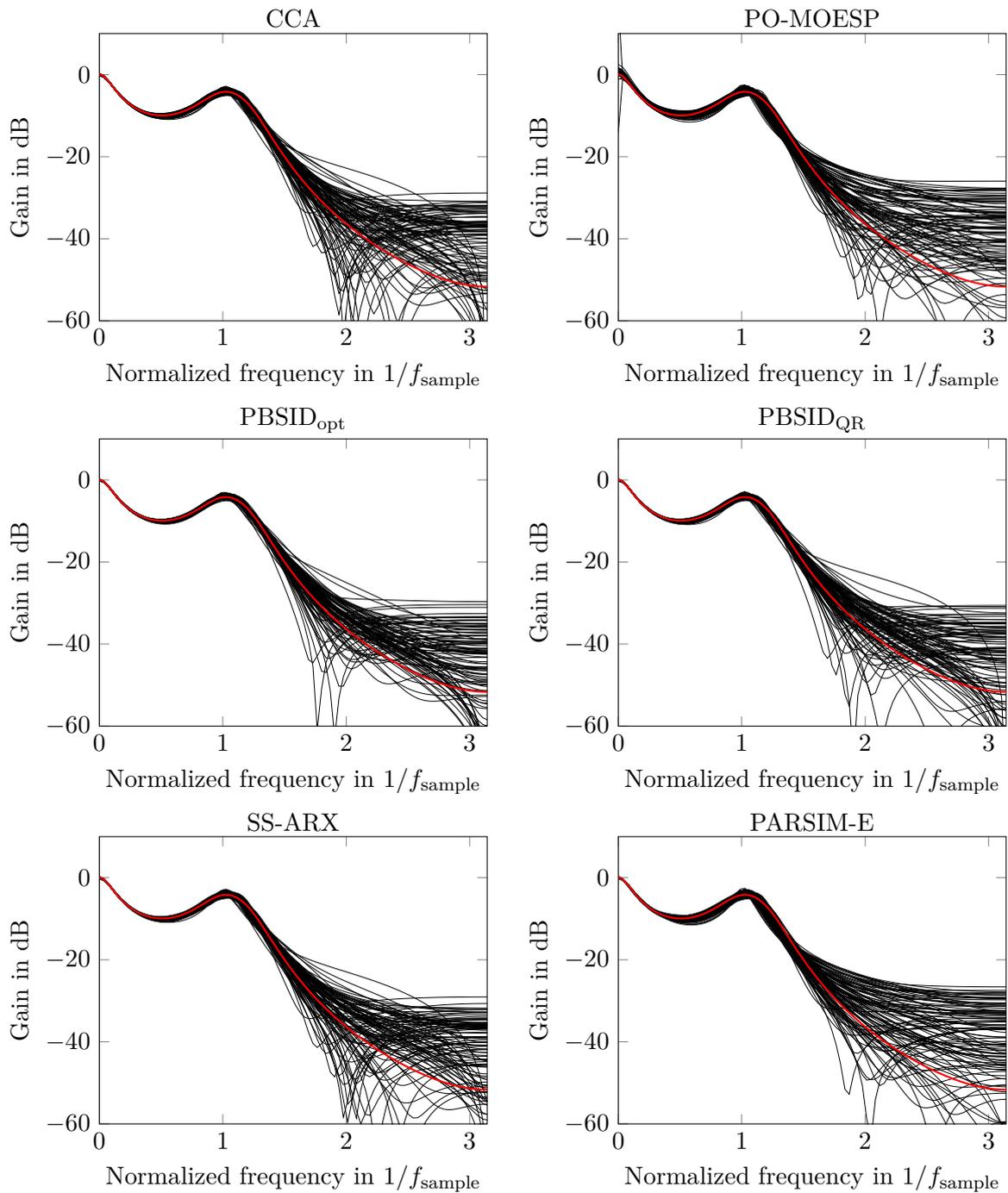
Figure 4.15: Bode gain plots of the identified models for the ARMAX system (red: nominal value of $P(z)$, black: value of identified models)

Closed-loop example

For the evaluation of the closed-loop capabilities of the proposed algorithm, several test with examples presented in the literature were performed. The here discussed results are based
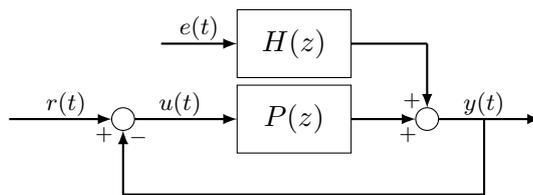
Figure 4.16: System structure of the closed-loop example (Box-Jenkins system)

on the example given in Jansson (2003) and Ljung and McKelvey (1996b). The structure of the example is shown in Figure 4.16. The transfer functions $P(z)$ and $H(z)$ are defined by

$$P(z) = \frac{0.21z^{-1} + 0.07z{-2}}{1 - 0.6z^{-1} + 0.8z^{-2}}$$

and

$$H(z) = \frac{1}{1 - 0.98z^{-1}} \,.$$

The closed-loop system is excited be means of the reference signal r, which is here a white-noise process with unity variance. The variance of the noise process e is $\sigma_e^2 = 4$. The number of columns N and horizon lengths k are set to $N = 2000$, $k = 30$. The number of data sets is 100. The input is kept the same and only the disturbances differ. The methods/algorithms used for the identification are the previously chosen CCA (state algorithm), PO-MOESP, PBSID$_{\text{opt}}$, PBSID$_{\text{QR}}$, SS-ARX, and PARSIM-E.

The results illustrated in terms of the eigenvalues and Bode gain plots in Figure 4.17 and Figure 4.18 showcase the results given for all identifications whose system included a genuine stochastic subsystem, i.e., not only a white-noise disturbance acting on the output of the system. Although the deterministic eigenvalues, which are in this example located at $0.3 \pm 0.843i$, are identified correctly (see Figure 4.17), the defective identification of the stochastic subsystem (eigenvalue at 0.98) results in a overall faulty identification. For cases in which $H(z) = 1$ was chosen, e.g., Chiuso (2007b), the results are as expected. However, for identifications including the modeling of the dynamics of a stochastic subsystem, be it a Box-Jenkins system as in this example or an ARMAX system as in the example of Qin and Ljung (2003a), neither of the methods have been capable to deliver the expected (and published) results. This is in so far intriguing as the examples of the respective papers could not be repeated in a satisfactory manner.

This points to a much more severe underlying problem in the implementation of the methods used for this example. A simple implementation error would also have shown itself during identification of open-loop systems. An implementation error regarding the VARX estimation seems hence highly likely, as this step is not elaborately explained in neither of the papers Chiuso (2007b), Jansson (2003), and Ljung and McKelvey (1996b). The problem might also concern the order of the VARX model, i.e., the length of the past horizon k_p. Although PARSIM-E does not include an explicit estimation of a VARX model, the recurrent orthogonal projections/regressions are similar to a VARX estimation.

This problem, which appears irrespective of the method, calls for a comparison of the methods and experiment implementations used for the disclosed results in the literature and here. This time consuming work is however postponed to a later date, as the main focus of this work is on the derivation of a framework for recursive subspace identification.
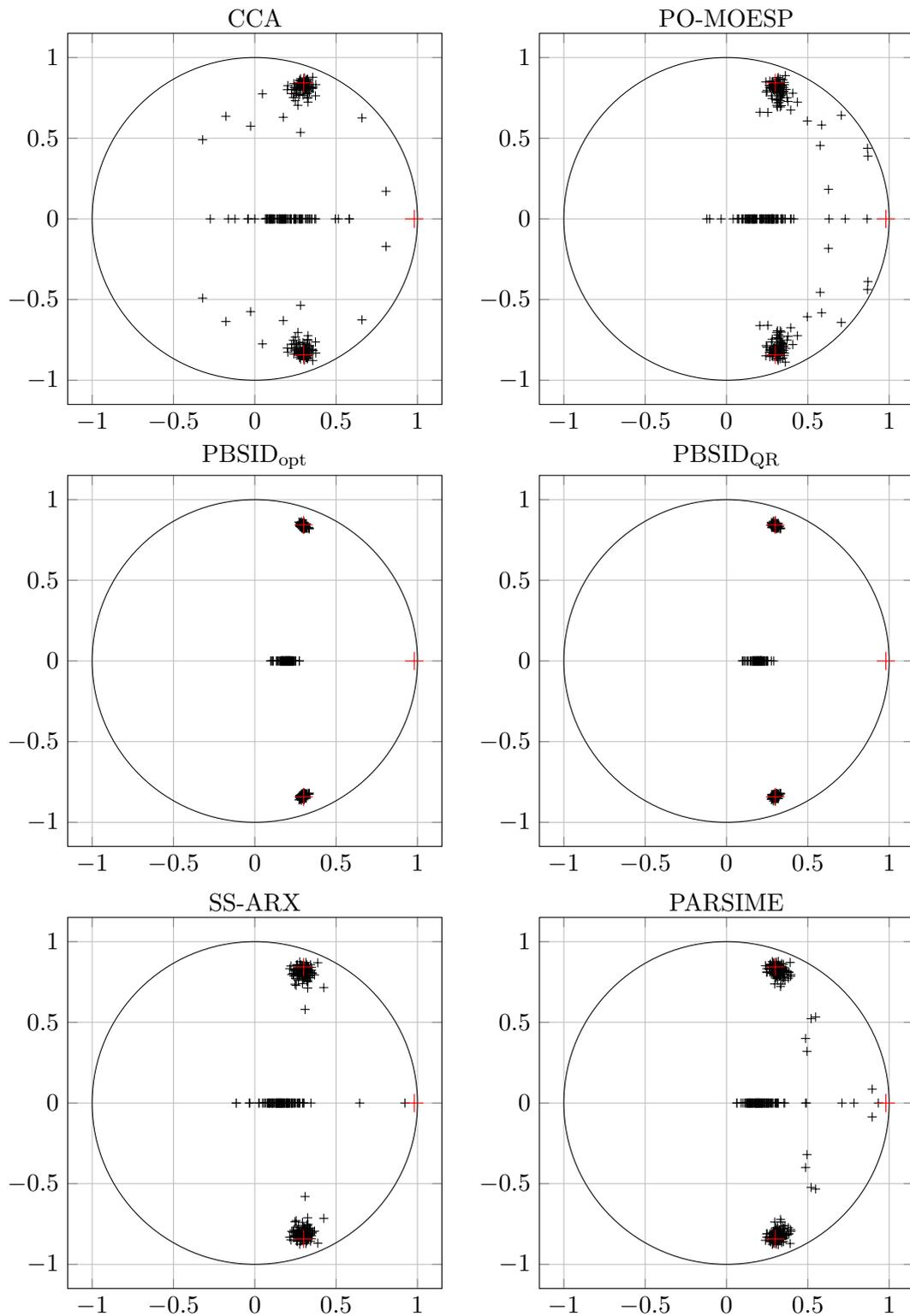
Figure 4.17: Eigenvalues of identified models for the Box-Jenkins system (red cross: nominal value of $P(z)$ and $H(z)$, black cross: value of identified models)
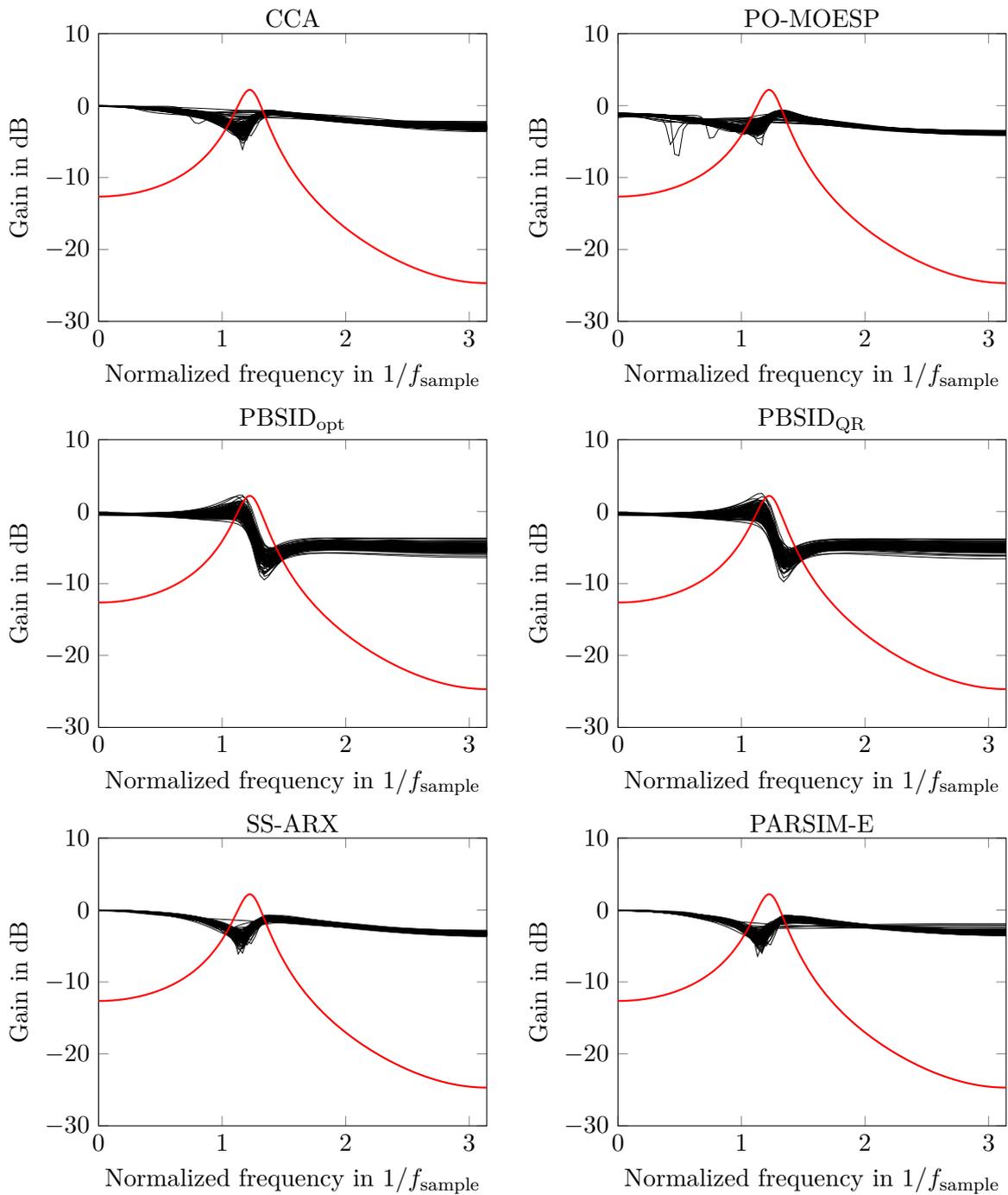
Figure 4.18: Bode gain plots of the identified models for deterministic subsystem of the Box-Jenkins system (red: nominal value of $P(z)$, black: value of identified models)

4.2.4 Summary

In this section an algorithm, called PBSID_{QR}, for the identification under closed-loop condition is derived. This algorithm is based on the PBSID method. The goal is to directly implement the theoretical description of the method and hence to avoid the pre-estimation of

a VARX model or teh involved least-squares calculation of previous algorithms of the PBSID method. Although not rendering impossible, these two algorithms would however hinder the integration of the recursive scheme derived in Chapter 6. The direct implementation is based on a rearrangement of the data of the future inputs and outputs within the data matrix, which the usual QR decomposition is applied to. The future data is not arranged in blocks with respect to the inputs or outputs but rather with respect to time. That is, the first entry of the joint-input-output block of the future is the input-output-pair of time t, followed by the pair of time $t + 1$ and so on. By means of this interleaving data structure, the oblique projections of the state estimation

$$\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{k_f - 1} \mathcal{X}_t^h \,,$$

where

$$\mathcal{X}_t^h = \hat{E}_{\|\mathcal{P}_{[t,t+h)}^+} \left\{ y(t + h) \mid \mathcal{P}_t^- \right\} \,,$$

can be directly constructed from the results of the QR decomposition. Hence, the PBSID method is directly implemented and the recursive scheme can be integrated, as these oblique projection do not prohibit to do so. Based on open-loop and closed-loop examples, it is shown that the PBSID$_{\mathrm{QR}}$ algorithm achieves results equal to the given PBSID$_{\mathrm{opt}}$ algorithm, where the PBSID$_{\mathrm{opt}}$ algorithm is slightly better if the spread of the eigenvalues is considered. A problem which needs further investigation is however concerned with the identification under closed-loop conditions. Although the algorithms used for the examples are implemented as given by the respective references, the disclosed results could not be recovered. This raises the question regarding the sensitivity with respect to the necessary parametrization of the algorithms or the implementation itself.

4.3 Summary

This chapter discusses modifications of algorithms of the ORT approach and the closed-loop method PBSID. The intention is the derivation of basic algorithms, which can be used for the integration of the recursive scheme derived in Chapter 6. As it has became apparent that the ORT algorithm based on the MOESP method does not perform satisfactory when dealing with realistic identification problems (e.g., the identification of the Tennessee Eastman Process), the goal has also been to change the actual identification algorithm in order to facilitate the identification of these kind of processes by the ORT approach. In terms of the ORT approach, both tasks are solved by changing the algorithm from a realization-based algorithm to a state-regression algorithm based on the CCA method, called CCA-ORT. Therefore, the orthogonal decomposition of the output into the deterministic and stochastic components is embedded into the oblique projection used to determine the predictor space. By virtue of this combination, the deterministic predictor space is directly determined, instead of being the result of a two-step procedure. Regarding the modifications of the algorithm of the PBSID method, a circumvention of the explicit VARX estimation has been necessary. This is achieved by directly implementing the underlying projections based on the results of a QR decomposition of a data Hankel matrix featuring a rearranged structure. Although the proposed PBSID$_{\mathrm{QR}}$ algorithm is able to perform similar to the PBSID$_{\mathrm{opt}}$ algorithm, closed-loop examples reveal a problem in terms of the correct parametrization of the methods, as both algorithm give (the same) incorrect results.

Recapitulation of main results

- The orthogonal decomposition of the output space into the deterministic and stochastic subspaces, referred to as ORT approach, and the subsequent determination of the deterministic predictor space can be combined as

$$\mathcal{X}_t^{d,+/-} = \hat{E}_{\|\mathcal{U}_t^+}\{\mathcal{Y}_t^+ | \mathcal{U}_t^-\} \ .$$

- An algorithm embedding this approach of the orthogonal decomposition into the CCA method, called CCA-ORT, is presented. Based on the SVD

$$\Sigma_{y_t^+ y_t^+ | u_t^+}^{-1/2} \Sigma_{y_t^+ u_t^- | u_t^+} \Sigma_{u_t^- u_t^- | u_t^+}^{-T/2} = U \Sigma V^T \approx U_n \Sigma_n V_n^T \ ,$$

where U_n, Σ_n, V_n^T are reduced according to the first n singular values, the observability matrix of the deterministic subsystem is estimated by

$$\hat{\mathcal{O}}_k = \Sigma_{y_t^+ y_t^+ | u_t^+}^{1/2} U_n \Sigma_n^{1/2}$$

and subsequently the deterministic states as

$$\hat{x}_d(t) = \hat{\mathcal{O}}_k^\dagger \Sigma_{y_t^+ u_t^- | u_t^+} \Sigma_{u_t^- u_t^- | u_t^+} u_t^- \ ,$$

$$\hat{x}_d(t+1) = \hat{\mathcal{O}}_k^\dagger \Sigma_{y_{t+1}^+ u_{t+1}^- | u_{t+1}^+} \Sigma_{u_{t+1}^- u_{t+1}^- | u_{t+1}^+} u_{t+1}^- \ .$$

With these values, the system matrices of the deterministic subsystem can be determined by the state-regression approach. CCA-ORT slightly improves the results of the MOESP-based ORT approach in terms of the identification of theoretical examples.

- An algorithm of PBSID method, directly implementing the determination of the predictor space by

$$\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{k_f - 1} \mathcal{X}_t^h$$

where

$$\mathcal{X}_t^h = \hat{E}_{\|\mathcal{P}_{[t,t+h)}^+}\{y(t+h) \mid \mathcal{P}_t^-\}$$

is presented. This is facilitated by changing the arrangement of the data in data Hankel matrix $\left[(U_t^+)^T \ (P_t^-)^T \ (Y_t^+)^T \right]^T$ to $\left[(P_t^-)^T \ p_N^T(t) \ p_N^T(t+1) \ \cdots \ p_N^T(t+k_f-1) \right]^T$, where $p_N^T(t) = \left[y_N^T(t) \ u_N^T(t) \right]$. As a consequence of this rearrangement, the oblique projections yielding $\mathcal{X}_t^h$, $h = 0, \cdots k_f - 1$ can be implemented, and a causal estimation of $\mathcal{O}_k^K x(t)$ is achieved. This algorithm is called PBSID$_{QR}$.

5 Tennessee Eastman Process identification study for non-recursive identification methods

In order to illustrate how the various methods stack up against each other if the identification of a real system is considered, the results of an identification study using the Tennessee Eastman Process described in Downs and Vogel (1993) are presented in this chapter. This study illustrates the ability of the proposed algorithms to deal with realistic processes. In particular, it shows that the CCA-ORT algorithm is able to deliver the expected results, i.e., determination of the deterministic model even if the process is increasingly affected by disturbances, which is an environment other methods tend to fail in. As the Tennessee Eastman Process is a first-principles model of a real process plant, the presented results are exemplary for the results which are to be expected for identifications of real systems.

5.1 Experiment description

This identification study aims at a comparison of different subspace methods against the realistic process model of the Tennessee Eastman Process. A short overview of this model is given in Appendix A. As the process is inherently unstable, some of the inner (stage 1) control loops of the control strategy explained in McAvoy and Ye (1994) and McAvoy, Ye, and Gang (1995) are implemented. In terms of this identification study, the task of these loops is to stabilize the process and first and foremost to keep the respective process units off its shutdown limits (see Table A.3). Thus, the process runs "as open-loop as possible". The resulting process structure is shown in the piping and instrument diagram (P&ID) of Figure 5.1. Due to this configuration of the process, this study differs from the study disclosed in Juricek, Seborg, and Larimore (2001), where a fully controlled process has been identified. Such a configuration however tends to obscures the characteristics of the process, i.e., the nonlinear behavior of the process. A summary of the key points of the experiments, e.g., inputs, outputs, number of identifications, etc. can be found in Table 5.1.

In order to compare the methods also under realistic conditions, two experimental arrangements are considered. In terms of the first arrangement, identifications without disturbances or measurement noise are considered. These first identifications shed light on the capability of the methods to deal with non-academic examples. The results are used for a basic comparison of the methods' behavior and represent also the individual benchmarks of the methods for the second experimental arrangement. Here, the outputs of the process are additionally subjected to arbitrarily colored-noise disturbances. By affecting the process outputs with disturbances, the resulting system structure of the second identifications bears resemblance to the Box-Jenkins example of Section 4.1.4. The internal process disturbances are not used as they do not comply with the assumption of being (colored) noise sequences uncorrelated with the input and hence exhibit some kind of variance with the input signals.
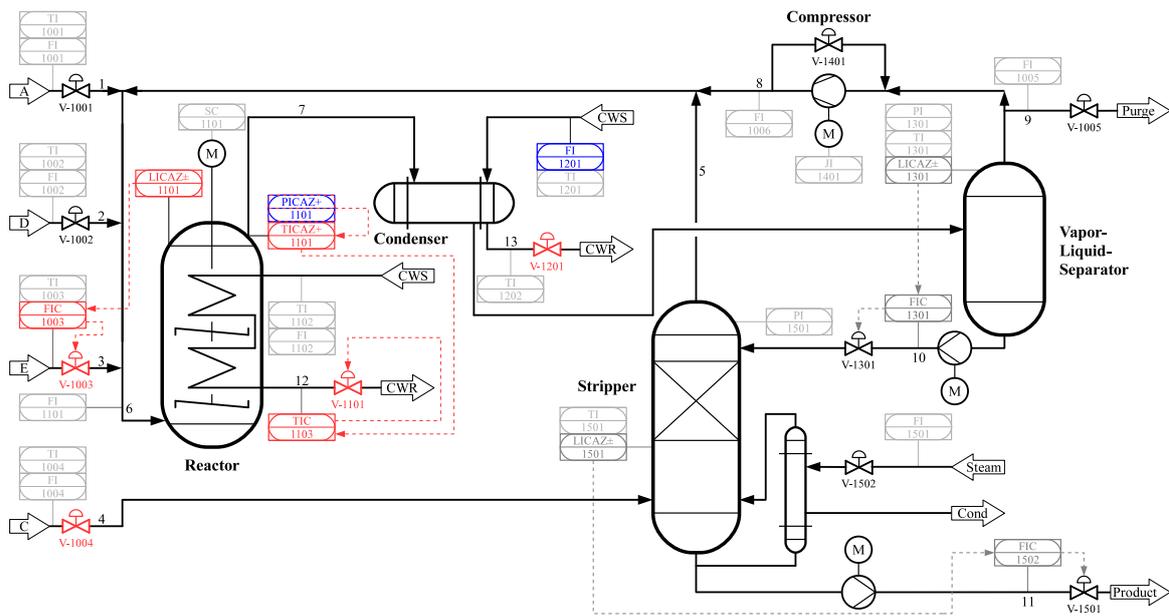
Figure 5.1: P&ID of the stabilized Tennessee Eastman Process (Bathelt, Söffker, and Jelali, 2015)

The identifications are based on a small subprocess of the Tennessee Eastman Process model. The outputs of this subprocess (in Figure 5.1 in blue) are the condenser cooling water flow (measuring point FI1201) and the reactor pressure (measuring point PICAZ+1101). The inputs (in Figure 5.1 in red) are the position of the C feed valve (V-1004), the position of the condenser cooling water valve (V-1201), the set-point of the reactor level control cascade (from LICAZ±1101 via FIC1003 to V-1003), and the set-point of the reactor pressure control cascade (from PICAZ+1101 via TIC1103 to V-1101). This was mainly done to keep the numerical load of the identifications to a reasonable level and due to the lack of additional information on the methods' behavior gained from a larger model. Another reason is also given by a potential ill-conditioning of the process, i.e., the possibility of the process being dominated by the dynamics of one process unit. However, the subprocess containing the mentioned inputs and outputs displays two interesting features. First, it contains the rather sensible path from V-1004 to the reactor pressure. Infusing too much components A, B, and C through the C feed leads to a dis-balancing of the reaction and usually to a subsequent shutdown of the process as one of the components starts to accumulate in the gas phase of the reactor and subsequently raises the reactor pressure beyond the shutdown limit. Secondly, the condenser cooling water flow depends only on the position of the valve V-1201. That is, this output highlights the methods' capabilities to correctly assign the observed effects to the respective inputs. Furthermore, due to the internal feedback (see streams 5 and 8 in the P&ID of Figure 5.1), the two outputs feature the reaction of the entire process to changes of the respective inputs, i.e., changing the position of V-1201 will affect not only the cooling water flow but also the reactor (through the separator and stripper).

Regarding the signals of the four inputs, multi-variable colored-noise sequences are used during the data generating simulations. Those are defined by

$$C(z) = \sigma_i^2 \frac{\sqrt{1 - 0.995}}{1 - 0.995z^{-1}} \ ,$$

Table 5.1: Overview of the key parameters of the comparative study

Inputs	C feed valve (V-1004), condenser cooling water valve (V-1201), set-point reactor pressure (PICAZ+1101), set-point reactor level (LICAZ±1101)
Outputs	Reactor pressure (PICAZ+1101), condenser cooling water flow (FI1201)
Number of simulations	100
Types of identifications	Distrubance-free & output subjected to colored-noise sequences
Methods/algorithms	PO-MOESP, PI-MOESP, ORT, N4SID Algorithm 1, N4SID Robust Algorithm, $PBSID_{opt}$, $PBSID_{QR}$, CCA, CCA-ORT
Sample time (of data)	2 min
N	10000
$k_f = k_p = k$	425
Order (see Section 5.2)	6
Validation methods (see Section 5.3)	Coefficient of determination/variance accounted for & final prediction error
Validation data sets	3 stochastic processes with 5 realizations each

where $\sigma_i^2 = 4$, $i = 1, 2, 3$, for the valve positions and the reactor level set-point and $\sigma_4^2 = 400$ for the reactor pressure set-point. A total of 100 input sequences are generated, which are equally used for the identifications with and without disturbances. The disturbances acting on the reactor pressure and the condenser cooling water flow are governed by the filter

$$H(z) = \sigma_i^2 \frac{1 - 1.3z^{-2} + 0.42z^{-2}}{1 - 1.8z^{-2} + 0.97z^{-2}} \, ,$$

where σ_i^2, $i = 1, 2$, are adjusted to comply with predefined signal-to-noise ratios (SNR, set to 50 and 10) as

$$\sigma_i^2 = \frac{\sigma_{d,i}^2}{\text{SNR} \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega} \, ,$$

with $\sigma_{d,i}^2$ being the variance of the respective outputs. The methods/algorithms used are PO-MOESP, PI-MOESP, ORT (MOESP-based implementation), N4SID Algorithm 1, N4SID Robust Algorithm, $PBSID_{opt}$, $PBSID_{QR}$, CCA (state algorithm), and CCA-ORT. For each method, the past and future horizons k_p and k_f are set to $k = 425$, whereas the number of columns is $N = 10000$.

5.2 Order estimation

The original intended use of the order estimation methods, which are described in Section 2.6.3, has turned out to be pointless as neither of the methods has been able to yield a

Table 5.2: Estimated orders by different order estimation methods (exemplarily for one data set; number of total singular values: 850)

Method	PO-MOESP	N4SID 1	N4SID Robust	CCA	CCA-ORT
SVC	246	1	1	1	1
NIC	850	1	1	1	2
C_n	–	–	–	436	236

reasonable estimate for the order of the model. The resulting estimates are exemplarily shown in Table 5.2 for one of the 100 data sets. Regarding the use of the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) as done in Juricek, Seborg, and Larimore (2001), the following needs to be taken into consideration. Both the AIC and BIC are criteria based on maximum likelihood estimates of the innovation variance (Hannan and Deistler, 2012, p. 161; see also description in Akaike, 1976). It is hence highly questionable whether the use of these criteria in conjunction with subspace methods is theoretically sound, as the resulting estimates of the innovations are the results of a minimal variance estimation (cf. definitions of projections in Section 2.2.3). Even in Larimore and Baillieul (1990, p. 599), the solution given by the CVA method is referred to as an "approximate maximum likehood solution", i.e., the CVA method used in Juricek, Seborg, and Larimore (2001) does not give a maximum likelihood estimates of the innovation variance.

Hence, the order determination has been done by resorting to the heuristic procedure of evaluating the singular values of the SVD step of the methods. In terms of this procedure, a distinct "gap" (see van Overschee and De Moor, 1994, p. 90) between the dominant singular values and residual singular values, which, from the theoretical point of view, are considered to be zero, marks the point which determines the order, i.e., the ordinal number of the last singular value before the gap determines the order.

In this context, the order has been determined using the singular values, which are given by the CCA and N4SID (Algorithm 1) methods for the first data set. These are shown in Figure 5.2. The singular values of the N4SID method show its gap between the fourth and the fifth singular value, whereas, for the CCA method, this gap is to be found in the
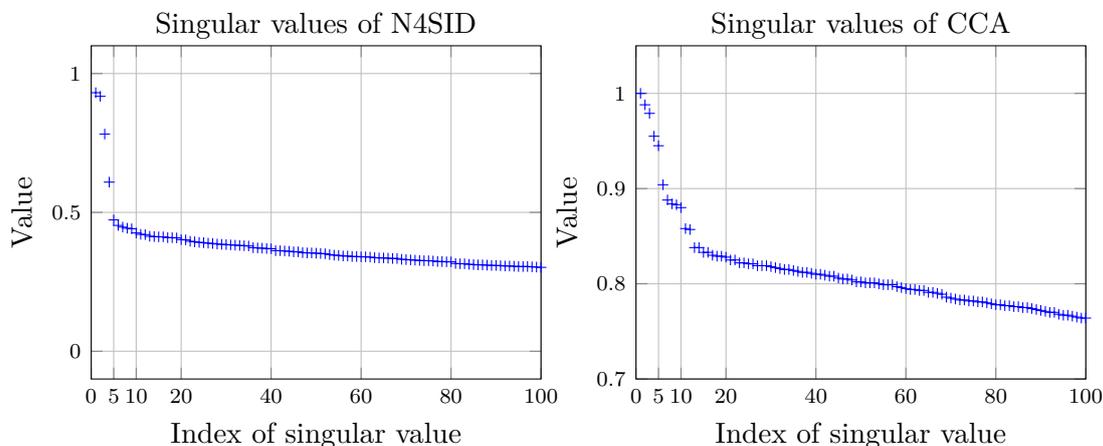

Figure 5.2: Singular values of N4SID (left) and CCA (right); first data set

range between the fifth and thirteenth singular value. In addition to the evaluation of the singular values, test identifications with orders between 4 and 12 has been made to evaluate the influence of the specified order. It turned out that there are six or seven eigenvalues, which are located in approximately certain regions of the complex plane. Taking this and the more significant meaning of the singular values of the CCA method as (conditional) canonical correlations between the past and future into consideration, the order has been chosen with 6. The difference of the order determined by the "gap" in contrast to the order determined in Juricek, Seborg, and Larimore (2001)[1], which was specified with 23, stems from the differences of the considered subprocess. Here, a rather small subprocess is considered, hence the order is smaller.

Remark 5.1. In terms of time-discrete systems, the order estimation problem can be given a rather simple interpretation. In terms of the infinite Hankel matrix of impulse response coefficient

$$H = \begin{bmatrix} G_1 & G_2 & G_3 & G_4 & \cdots \\ G_2 & G_3 & G_4 & G_5 & \cdots \\ G_3 & G_4 & G_5 & G_6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} , \tag{5.1}$$

where the impulse response coefficients are given by the state-space realization of the system or model as (for deterministic realizations and joint stochastic-deterministic realizations)

$$G_i = CA^{i-1}B , \qquad \text{or} \qquad G_i = CA^{i-1} \begin{bmatrix} B & K \end{bmatrix} , \tag{5.2}$$

the order of the system or model is equal to $\mathrm{rank}(H)$. For linear systems, this is governed by the Cayley-Hamilton theorem, i.e., by

$$A^n + a_1 A^{n-1} + \cdots + a_{n-1}A + a_n I = 0 , \tag{5.3}$$

where the coefficients a_i are given by the characteristic polynomial of A. This relation is comprehensively explained in Katayama (2005, pp. 65–67). It is noteworthy that this holds independently of the chosen sample time as a change of the sample time results only in a change of the coefficients a_i but not in a change of the order of the characteristic polynomial. Regarding nonlinear systems, this clear-cut relation is no longer given. Although the order of the system is in a broader sense still determined by the Hankel matrix of impulse response coefficients, namely by the SVD step of the methods, the resulting order varies with respect to the sample time. Although the rank estimation of a matrix is determined by the number of linear independent rows or columns, it is numerically determined by the number of singular values above a certain threshold. This means for the rank estimation of a Hankel matrix, which consists of impulse-response coefficients of a nonlinear system, that the rank is in the worst case equal to the number of non-zero impulse response coefficients as linear dependency can be not achieved until the coefficients become zero or close to zero. This number is however equal to the number of coefficients of the settling time. Hence, the sample time influences the estimated order. In Juricek, Seborg, and Larimore (2001), the sample time has been set to 30 min, whereas here the sample time is 2 min. Thus, the estimated orders deviate; in Juricek, Seborg, and Larimore (2001) the AIC estimated the order with 23.

---

[1]The disclosed plots in Juricek, Seborg, and Larimore (2001) (Figure 7) do unfortunately not allow for an exact location of the "gap".

5.3 Validation and evaluation

The validation and evaluation of the identified models is based on the coefficient of determination R^2 and Akaike's final prediction error (FPE) J_{FPE} (see Ljung, 2009, p. 500 or Kroll and Schulte, 2014). The former one, also known as "variance accounted for" (VAF), as it specifies the ratio of the output variance explained by the model (see Ljung, 2009, pp. 549–550), has the advantage to give an objective good-bad-evaluation of a model and not only a relative comparison between two models. The latter one has the advantage that it also takes the model complexity, i.e., the number of parameters, into consideration and hence gives a better evaluation of the model's actual accuracy[2]. If the error covariance of a model of, e.g., order 6 is similar to the error covariance of a model of higher order, the low-order model is superior, as it gives the same accuracy while being less complex. The coefficient of determination/variance accounted for measure is defined by

$$R^2 = J_{VAF} = \left(1 - \frac{\sum_{k=1}^{T}\left(y(k) - \hat{y}(k)\right)^2}{\sum_{k=1}^{T} y^2(k)}\right) \cdot 100\%$$

and is limited from below to 0, i.e., any value smaller than 0 is set to 0. The final prediction error criterion is given by

$$J_{FPE} = \frac{1 + \dim(\mathcal{M})/T}{1 - \dim(\mathcal{M})/T} \frac{1}{T} \sum_{k=1}^{T}\left(y(k) - \hat{y}(k)\right)^2 \ ,$$

where $\dim(\mathcal{M})$ denotes the number of (independent) parameters of the model. In terms of the state-space model as given by (2.4)[3] (Hannan and Deistler, 2012, p. 69),

$$\dim(\mathcal{M}) = n(m + p) + np + mp \ .$$

For the numerical evaluations and in the wake of the fixed order of the models, the value of $\dim(\mathcal{M})$ is set to $np + nm + mp$ – the number of independent parameters of the deterministic subsystem.

In terms of the possible validation signals, the limitation to stochastic signals is obvious. The final prediction error is furthermore restricted to signals having the same second-order properties as the signals used during the identification/estimation (Ljung, 2009, p. 503). Hence, the input-output data of the validation data sets are generated by inputs, which are again defined by

$$C(z) = \sigma_i^2 \frac{\sqrt{1 - a}}{1 - az^{-1}} \ ,$$

where $\sigma_i^2 = 4$, $i = 1, 2, 3$ for the valve positions and the reactor level set-point and $\sigma_4^2 = 400$ for the reactor pressure set-point. Three validation sets with five realizations of stochastic processes defined by the value of a were generated. The parameter a was set to $a = 0.995$ for the first validation set, to $a = 0.9$ for the second validation set, and to $a = 0.9995$ for the third validation set. The rationale of having three validation sets is to test the models not only with signals which are a statistically the same as the estimation data sets but also with signals, which are, in terms of its dynamics, faster (validation set 2) or slower (validation set 3). The

---

[2]The measures have been chosen before the problem of the order estimation became apparent. This measure was intended to be used for an objective comparison of the resulting models despite having different orders. In the light of a fixed order, the value of the FPE is equal to the error covariance.

[3]See also discussion regarding the number of invariant parameters in Section 7.2.

number of realizations within each validation set is governed by the rationale of mitigating the possibility of having a process realization that is particularly unsuitable for the given model, i.e., yielding an unjustified poor result for one model of a method with respect to a model of another method despite being actually the better model (e.g., in terms of the step response). For the calculation of the two measures, the length of the generated realizations reaches up to a length of 504 h. The first 24 h are however discarded to accommodate the transient behavior of the process (response of the Tennessee Eastman Process on the activation of the noise input). The two outputs of the models – reactor pressure and condenser cooling water flow – are separability evaluated. See Table 5.3, Table 5.4, Table 5.5, and Table 5.6 for the resulting values given by the coefficient of determination and final prediction error with respect to the respective validation set. The tables show the median as well as the upper and lower quartile values for the resulting values of the measures.

In terms of the results of identifications with undisturbed data, which are illustrated for the reactor pressure in Figure 5.3, it is evident that, although the median of the models is acceptable, the realization-based methods are not suitable for the identification of systems that violate the linearity assumption. Furthermore, the methods' results become better, the less the methods' derivations are based on a coordinate-based framework and the more the general underlying data relations are used (increasing accuracy from N4SID to CCA). The plot of the final prediction error mirrors the results given by the coefficient of determination as the results improve from the realization-based methods to the state-regression-based methods and from N4SID to CCA. The outliers of PI-MOESP are as high as 10^{34}. In terms of the condenser cooling water flow, CCA and CCA-ORT give the best results as both methods identified the transfer functions from V-1004, PICAZ+1101, and LICAZ±1101 with nearly zero while correctly representing the response on a change of the valve position. This is reflected the best by the small values of final prediction error in Table 5.6. The values of the CCA-based algorithms are roughly one order of magnitude better than the results of N4SID and slightly better than the results of the PBSID algorithms. In terms of the identification with disturbances the expected behavior occurs. The accuracy of the models starts to decrease; compare upper diagrams of Figure 5.3 and Figure 5.4 for an illustration of the coefficient of determination of the reactor pressure models. Figure 5.4 also shows that the models determined by CCA-ORT algorithm exhibit a better accuracy in the high-frequency range as the drop of the coefficient of determination is not as pronounced as it is the case with the other methods; compare upper and middle diagrams of Figure 5.4. Regarding the transition from identifications without disturbances to identifications with disturbances, an intriguing effect was shown by the CCA method. In terms of the step responses, the accuracy increases if the disturbances are small (SNR < 50, or white-noise disturbances), cf. Figure 5.5 and Figure 5.6. This is but an effect of the numerical implementation. If there are no disturbances, the data matrices of the input and output tend to become numerically parallel. This in turn adversely affects the estimation. If there is a slight disturbance, this parallelism vanishes and the identification algorithm works more closely according to the theoretical derivations. Although the differences between the CCA method and CCA-ORT algorithm for increasing disturbances seems not to be severe in terms of the measures, the step responses shown in Figure 5.6 and Figure 5.7 for disturbance levels of SNR = 50 and SNR = 10 draw a contrary picture regarding the influence of disturbances on the identification results. Hence, considering the step responses, the CCA-ORT algorithm gives clearly the better results even in conditions with severe disturbances (SNR = 10).

Table 5.3: Resulting coefficients of determination R^2 of reactor pressure (all values in %)

		PO-MOESP	PI-MOESP	ORT	N4SID 1	N4SID Robust	PBSID$_{opt}$	PBSID$_{QR}$	CCA	CCA-ORT
un-disturbed	upper quartile	95.4779	0.6347	96.0672	98.9693	98.9343	98.9935	98.8942	99.3319	99.2833
	median	82.3606	0	93.7378	97.8232	98.2171	98.7064	98.6040	99.1208	99.1056
	lower quartile	13.6241	0	0	96.8895	97.0934	98.3620	98.2294	98.7776	98.7678
SNR = 50, val. 1	upper quartile	61.7955	2.4744	64.8333	95.5019	96.5098	98.8100	98.8896	98.6642	99.2836
	median	50.8550	0	44.793	91.7372	95.6264	98.5132	98.5741	98.3592	99.1195
	lower quartile	36.6535	0	0	84.7664	94.3125	98.1395	98.2294	97.9757	98.7943
SNR = 50, val. 2	upper quartile	38.4086	13.2308	40.2511	82.5272	66.2707	91.8786	92.4170	89.2076	94.4607
	median	30.9748	0	17.8873	73.5030	62.2013	90.1205	90.8617	88.0166	94.0296
	lower quartile	0	0	0	66.2347	49.2227	88.2501	89.0874	86.5693	93.5027
SNR = 50, val. 3	upper quartile	74.6245	46.4281	77.5276	97.6396	99.5359	99.8473	99.8580	99.8418	99.9102
	median	65.1515	0	61.9446	94.1120	99.3160	99.7607	99.7733	99.7392	99.8409
	lower quartile	13.7286	0	0	86.7176	99.1016	99.6833	99.6990	99.6799	99.7922
SNR = 10, val. 1	upper quartile	19.1732	0	10.1188	74.6859	88.9535	96.6469	96.7069	96.6652	98.7870
	median	11.1064	0	0	57.5484	84.3283	96.0309	96.1208	96.1028	98.4329
	lower quartile	0	0	0	21.2435	0	95.3214	95.4128	95.4042	97.9595
SNR = 10, val. 2	upper quartile	0	17.9570	0	37.7370	0	65.9920	67.1718	67.2387	90.9923
	median	0	0	0	25.3345	0	62.0521	63.6476	62.7639	85.9176
	lower quartile	0	0	0	10.3337	0	58.9080	60.3099	58.1632	81.2110
SNR = 10, val. 3	upper quartile	22.8354	19.5392	19.9100	74.1662	98.3891	99.5908	99.5986	99.5924	99.8564
	median	13.1179	0	0	52.9442	97.5014	99.3864	99.3973	99.4211	99.7857
	lower quartile	0	0	0	6.8535	0	99.2821	99.2837	99.2591	99.6950

Table 5.4: Resulting final prediction error J_{FPE} of reactor pressure

		PO-MOESP	PI-MOESP	ORT	N4SID 1	N4SID Robust	PBSID$_{opt}$	PBSID$_{QR}$	CCA	CCA-ORT
un-disturbed	upper quartile	328.7183	2750.8368	377.1845	11.0644	10.5682	6.1642	6.5771	4.4854	4.4282
	median	67.6392	1105.1922	22.3791	8.3382	6.8336	5.0094	5.3796	3.4997	3.6451
	lower quartile	17.6999	385.6282	16.0467	6.3183	4.0149	4.0114	4.3326	2.6593	2.8491
SNR = 50, val. 1	upper quartile	214.9215	2485.3603	21420.3110	57.8468	20.5654	6.8777	6.5670	7.4559	4.3606
	median	180.6571	859.8662	15987.9741	32.7925	16.1018	5.8040	5.4975	6.3073	3.4839
	lower quartile	160.1567	416.1966	149.2623	17.3076	14.1382	4.7341	4.4520	5.2266	2.7710
SNR = 10, val. 1	upper quartile	3230.6998	3685.0265	1.6303e10	298.2663	3501.2881	16.8965	16.5714	16.8975	7.7314
	median	377.0323	1072.9459	15987.9742	163.4816	56.8129	15.3928	15.1046	15.2471	6.0314
	lower quartile	269.3639	535.2413	372.6657	97.7823	42.2882	14.0536	13.7219	13.5476	4.6108

Table 5.5: Resulting coefficients of determination R^2 of condenser cooling water flow (all values in %)

Condition	Statistic	PO-MOESP	PI-MOESP	ORT	N4SID 1	N4SID Robust	PBSID_{opt}	PBSID_{QR}	CCA	CCA-ORT
undisturbed	upper quartile	33.8920	0	33.0848	98.9549	99.1072	99.9408	99.9406	99.9449	99.9449
	median	30.1897	0	29.8706	98.8143	99.0328	99.9356	99.9356	99.9406	99.9407
	lower quartile	27.5201	0	23.8237	98.6411	98.9045	99.9256	99.9264	99.9335	99.9335
SNR = 50, val. 1	upper quartile	13.7486	0	14.6413	98.7846	98.9149	99.8281	99.8657	99.3991	99.9420
	median	10.8451	0	10.4207	98.4871	98.7362	99.7825	99.8296	99.2819	99.9367
	lower quartile	0	0	0	97.9196	98.1092	99.7391	99.7926	99.1548	99.9281
SNR = 50, val. 2	upper quartile	8.6842	0	9.0552	82.3743	79.6368	96.9570	97.5952	88.3160	98.9151
	median	4.8689	0	0	81.7287	77.7226	96.3939	97.1241	86.6229	98.8491
	lower quartile	0	0	0	80.8967	65.3055	95.9533	96.6830	84.8272	98.8144
SNR = 50, val. 3	upper quartile	14.3087	0	15.0332	99.6639	99.8981	99.9796	99.9850	99.9451	99.9946
	median	11.7102	0	11.0243	99.3567	99.7972	99.9680	99.9756	99.9279	99.9942
	lower quartile	7.9955	0	0	98.7078	99.6927	99.9406	99.9579	99.8549	99.9874
SNR = 10, val. 1	upper quartile	11.5893	0	9.6807	98.1224	97.8245	99.2020	99.4967	98.1887	99.8615
	median	5.9397	0	0	97.0919	95.7313	98.9981	99.3564	97.8783	99.7941
	lower quartile	0	0	0	94.9520	0	98.7872	99.1952	97.6222	99.6761
SNR = 10, val. 2	upper quartile	0	0	0	78.9933	60.6182	87.4057	91.8478	63.2751	97.4405
	median	0	0	0	77.2595	22.9500	85.0434	89.96330	60.1073	96.1431
	lower quartile	0	0	0	73.4769	0	82.7591	88.2215	56.7876	93.7578
SNR = 10, val. 3	upper quartile	13.9134	0	10.8293	99.1059	99.7437	99.8989	99.9349	99.8388	99.9852
	median	7.6429	0	0	98.0929	99.4057	99.8432	99.8979	99.7991	99.9732
	lower quartile	0	0	0	95.3736	46.3350	99.7310	99.8208	99.5784	99.9577

Table 5.6: Resulting final prediction error J_{FPE} of condenser cooling water flow

|		PO-MOESP	PI-MOESP	ORT	N4SID 1	N4SID Robust	PBSID$_{opt}$	PBSID$_{QR}$	CCA	CCA-ORT
un-disturbed	upper quartile	24.1094	110.4446	27.2454	0.3648	0.2704	0.0181	0.0181	0.0164	0.0164
	median	18.8802	45.7641	19.4753	0.3158	0.2620	0.0176	0.0176	0.0163	0.0163
	lower quartile	16.9427	33.3676	16.8815	0.2938	0.2534	0.0168	0.0168	0.0155	0.0155
SNR = 50, val. 1	upper quartile	33.1057	87.3832	66.2048	0.5790	0.5202	0.0705	0.0551	0.2263	0.0178
	median	24.3738	47.1345	25.3930	0.4195	0.3308	0.0603	0.0481	0.2006	0.0172
	lower quartile	22.0729	34.5393	22.4291	0.3395	0.3017	0.0506	0.0389	0.1755	0.0166
SNR = 10, val. 1	upper quartile	99.2059	145.8713	5.4081e8	1.3558	118.3369	0.3338	0.2214	0.6280	0.0908
	median	26.7640	69.1063	437.2996	0.8228	1.1618	0.2737	0.1798	0.5786	0.0574
	lower quartile	23.2750	37.3245	25.3839	0.5207	0.5998	0.2325	0.1438	0.5379	0.0392
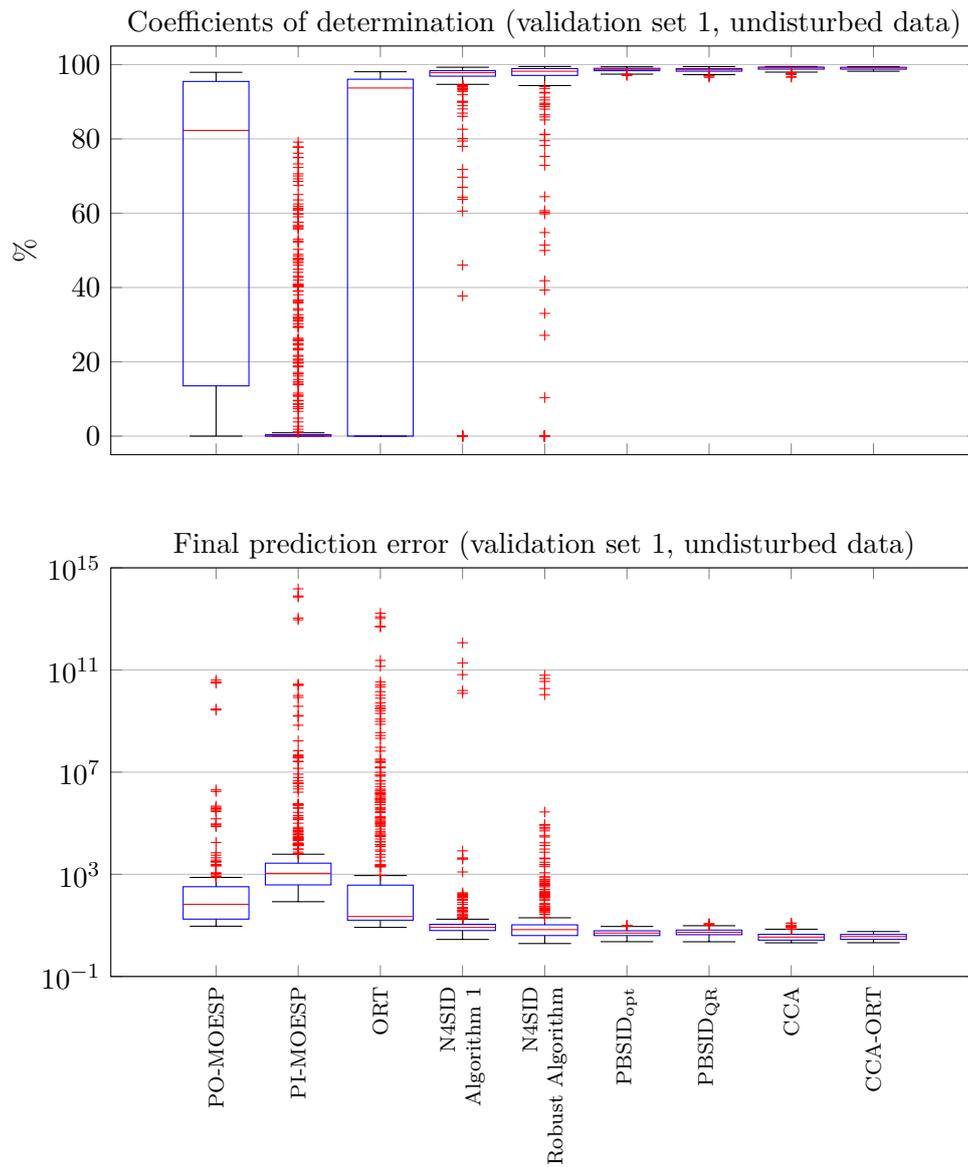
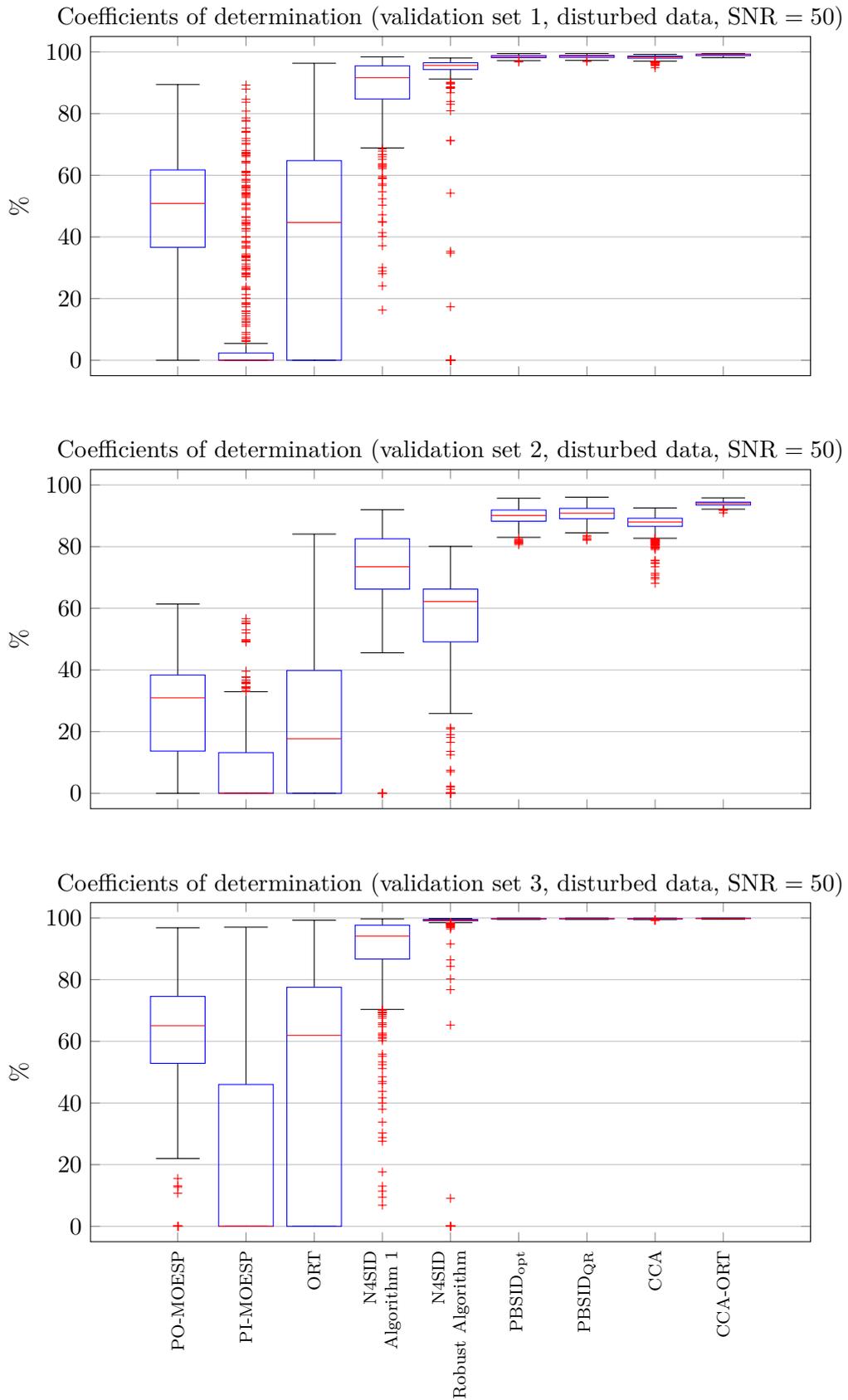Figure 5.3: Evaluation of the reactor pressure output for identifications without disturbances (validation set 1)

Coefficients of determination (validation set 1, disturbed data, SNR = 50)

Coefficients of determination (validation set 2, disturbed data, SNR = 50)

Coefficients of determination (validation set 3, disturbed data, SNR = 50)

Figure 5.4: Evaluation of the reactor pressure output for identifications with disturbances (SNR = 50) and for validation sets 1, 2, and 3
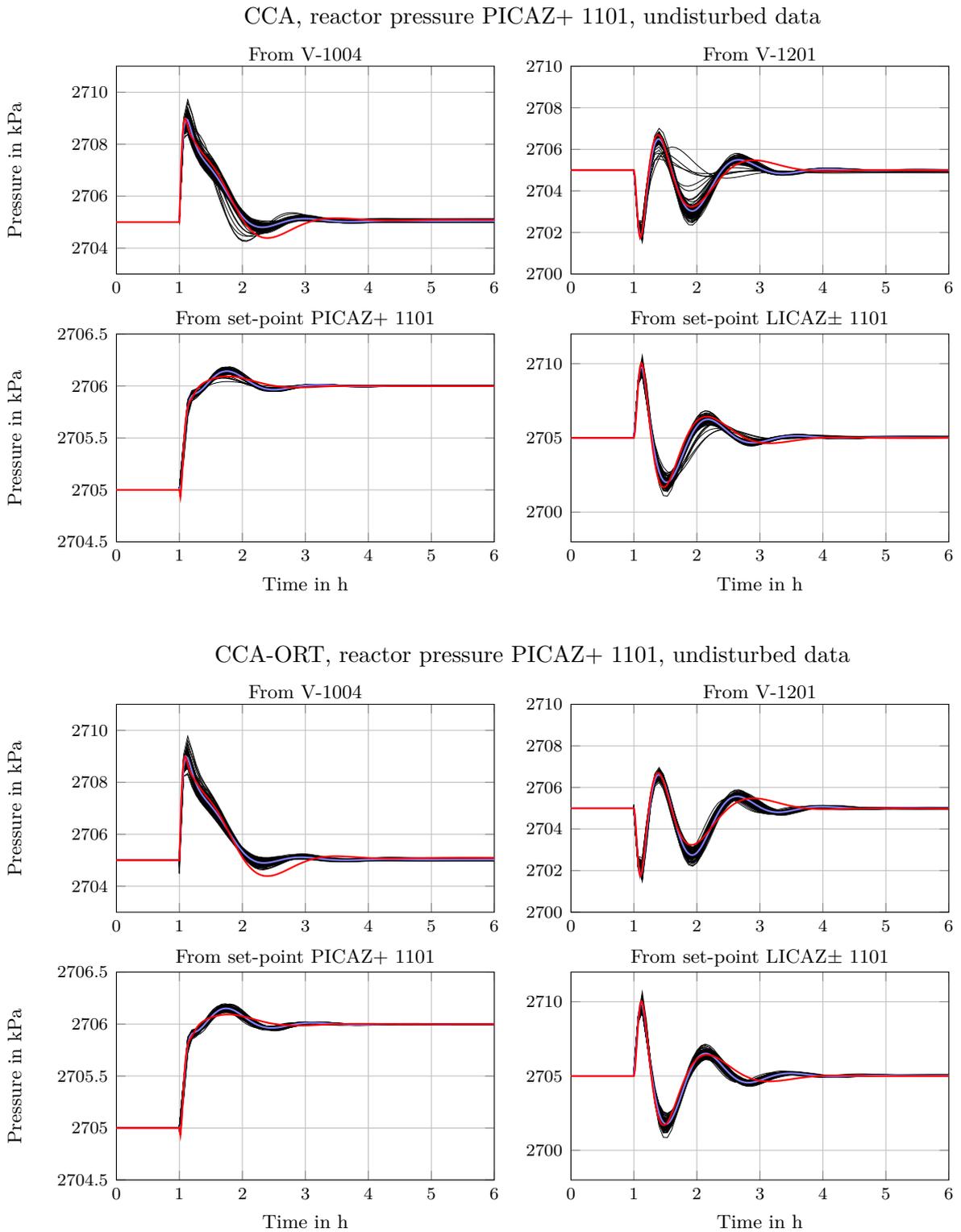
Figure 5.5: Step responses of models identified by the CCA method and the CCA-ORT algorithm using data of undisturbed process; black: individual results, blue: mean of results, red: step response of process
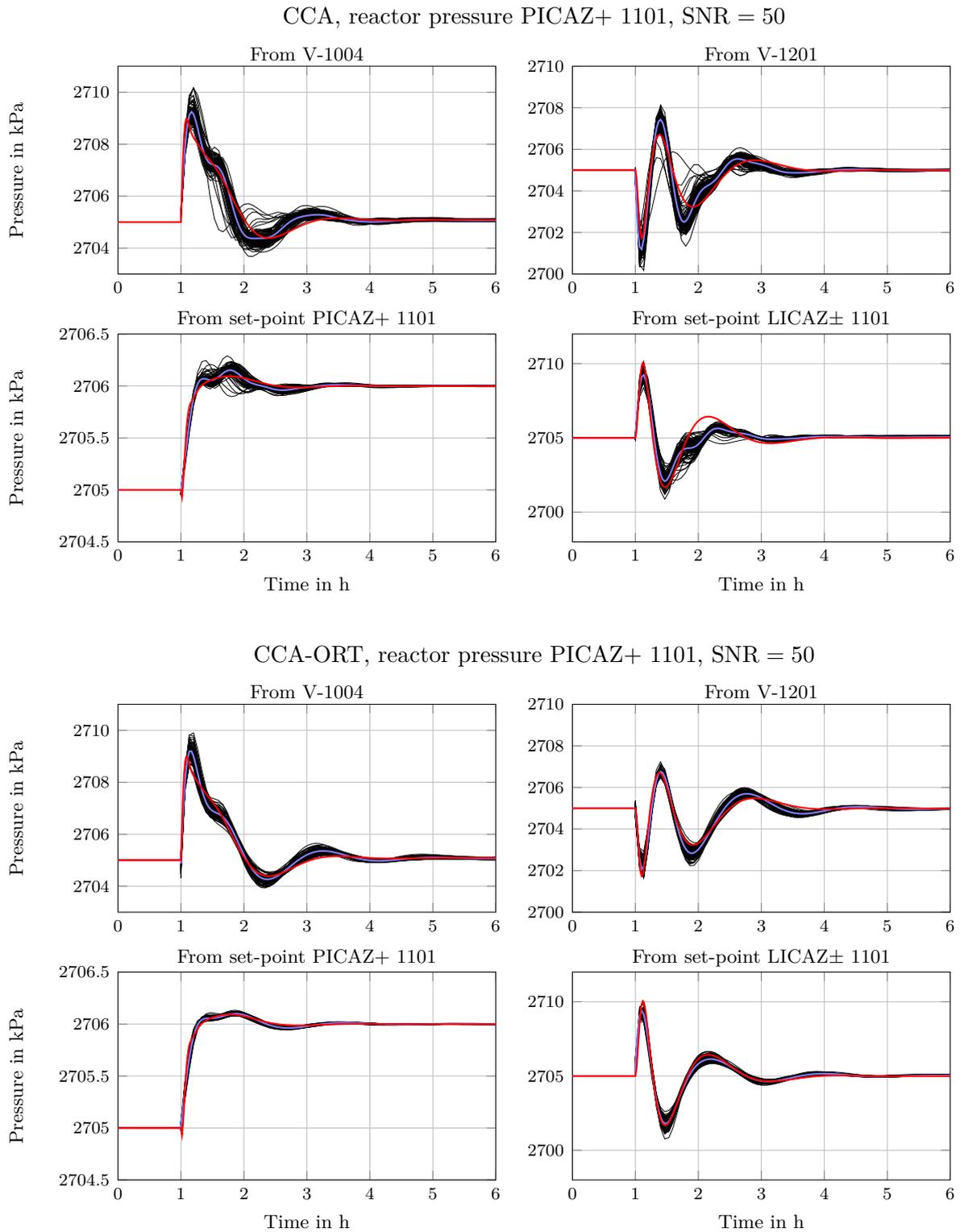
Figure 5.6: Step responses of models identified by the CCA method and the CCA-ORT algorithm using data of process subjected to disturbance (SNR = 50); black: individual results, blue: mean of results, red: step response of process
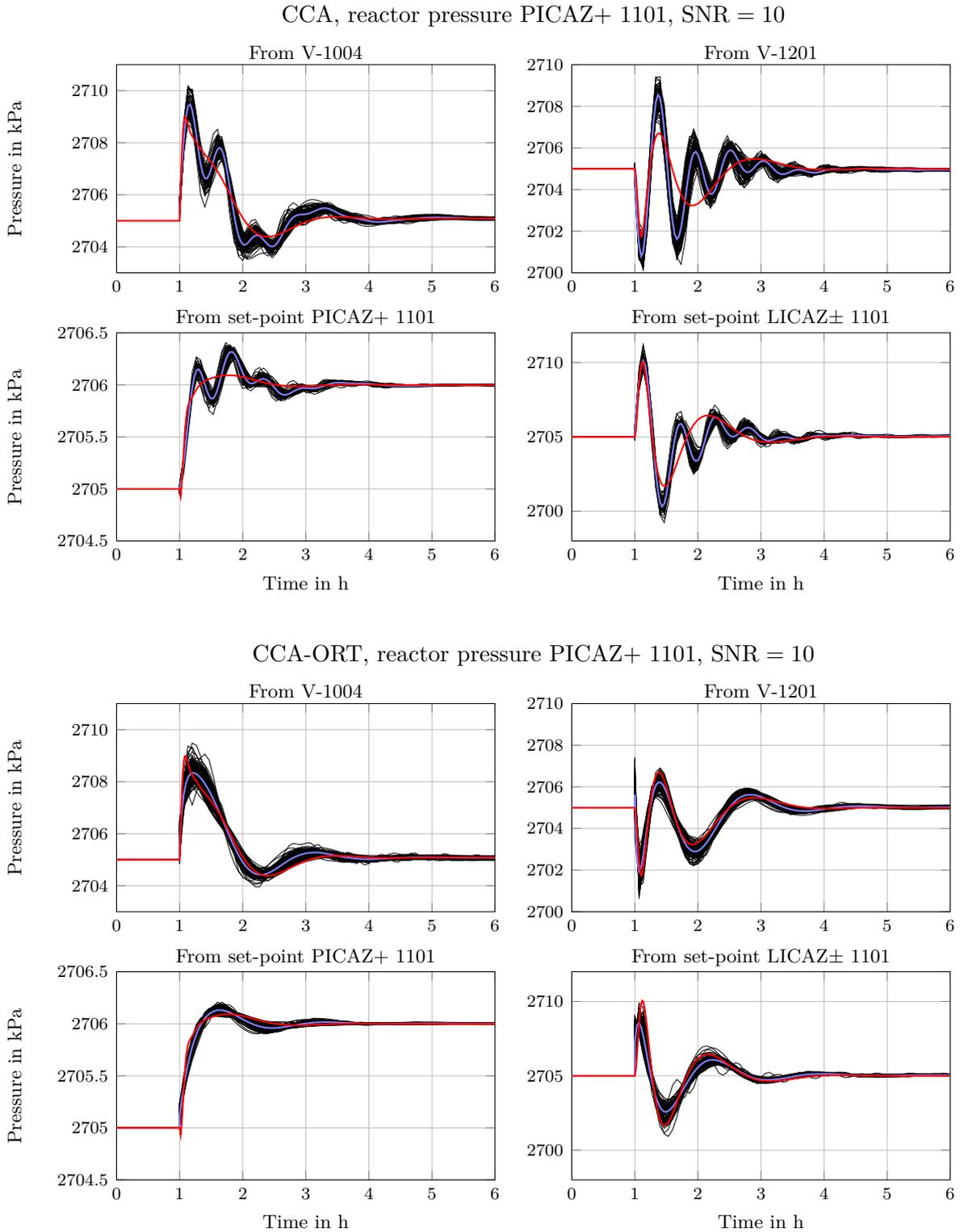
Figure 5.7: Step responses of models identified by the CCA method and the CCA-ORT algorithm using data of process subjected to disturbance (SNR = 10); black: individual results, blue: mean of results, red: step response of process

5.4 Summary

This chapter concludes the derivations of basic identification algorithms for the later derived approach to recursive subspace identification by an identification study. This study uses the realistic process model of the Tennessee Eastman Process. Results on real systems can be expected to match the here presented results. The identification tests include a variety of methods and cases considered. The methods applied to the data are the open-loop methods/-algorithms MOESP (PO-MOESP, PI-MOESP), N4SID (N4SID Algorithm 1, N4SID Robust Algorithm), CCA (state algorithm), MOESP-based ORT, and CCA-ORT as well as the algorithms $\text{PBSID}_{\text{opt}}$ and PBSID_{QR} of closed-loop method PBSID, as this method is also able to identify a model under open-loop conditions. Although the inherently unstable process model has been stabilized by some control loops, the data acquisition has been made outside of these control loops and thus under open-loop conditions. The cases considered for the identification are an undisturbed plant and plant subjected to some arbitrary colored noise of decreasing signal-to-noise ratio. The operating point was fixed for these identifications. The identifications are furthermore based on a small subprocess of the whole process and consists of 4 inputs and 2 outputs.

Recapitulation of main results

- Although there are methods for the order estimation, which work reasonably well in the case of LTI systems, they failed in this example when applied to a system of the scale of the Tennessee Eastman process.

- It turned out that the realization-based MOESP method is already challenged by the identification of the undisturbed system. With the addition of process disturbances this problem increases. This holds for every algorithm, PO-MOESP, PI-MOESP, and MOESP-based ORT algorithm, used here. The results of the state-regression-based methods N4SID, CCA, and PBSID are consistently better.

- In terms of the state-regression-based methods, the methods based on the theoretical framework and whose algorithms are rather straight forward implementations of the these theoretical descriptions have an advantage over methods/algorithms which do not exactly follow the theoretical descriptions. This follows from the comparison of the CCA and PBSID methods with the N4SID method. The N4SID method can give reliable results in an undisturbed environment but sometimes tend to yield wrong models (outliers in the boxplots). Under the influence of disturbances, the results of the N4SID method drastically worsen. Although not shown by diagrams, it has been observed that the CCA and PBSID methods are able to correctly identify the zero-output responses from three of the four inputs with respect to the condenser cooling water flow (this is only influenced by the position of its respective valve).

- In terms of the disturbance-affected identification data , the CCA-ORT algorithm yield the expected results. Of all methods and algorithm considered, this algorithm is the only one to recover the deterministic plant behavior even under the influence of severe disturbances (SNR = 10).

6 Recursive subspace identification

Up to this point, the identifications are considered to be done while the plant is in an operating point, i.e., at rest. This approach to identifying a system results in the necessity of reidentifying the model once a new operating point is reached as illustrated in Figure 4.1. The alternative is the identification of a genuine non-linear model for the whole system, e.g., a LPV model, which was deemed hard to achieve in Section 3.1 if a certain complexity of the process is surpassed, or unnecessary from a the standpoint of effort and reward, as a LTI model of the operating point basically suffices if the reidentifications are left out. To avoid this problem of a reidentification when the operating point is reached, the task of Section 3.1 reformulated this problem into an adaptive identification problem. That is, instead of performing the identification when the new operating point is reached and building a new model from scratch, the given model of the old operating point is adapted or updated during the transition phase between the two operating points. This idea is illustrated in Figure 6.1. The identification is pushed back to the transition phase between the operating points and the already identified model is not discarded but adapted. This idea of adapting a model is basically the problem formulation of recursive identification.

As discussed in Section 3.2.2, the problem of a genuine recursive formulation for subspace methods is basically concerned with the problem of reusing a previously identified predictor space. In this context, be reminded of the fact that any predictor space of a system is essentially equivalent to the model itself (Akaike, 1974, p. 669, Lindquist and Picci, 1996b). That is, any calculations of future predictor spaces that use a past predictor space of the same system, are, as a matter of principle, calculations based on the past information of the model and an update of this model.

From this conceptual idea, two important considerations follow:

1. The procedure of the recursion can be based on the projections (for stochastic and joint stochastic-deterministic systems)

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}\left\{\mathcal{Y}_{t+k}^+ \left| \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{X}_t^{+/-}\right.\right\}, \text{ or } \mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+}\left\{\mathcal{Y}_{t+k}^+ \left| \mathcal{P}_{[t,t+k)}^- \vee \mathcal{X}_t^{+/-}\right.\right\}.$$

   That is, the usual past can be replaced by any previous predictor space and the intermediate data space, i.e., by the spaces spanned by the values of y or y and u between the time points of the predictor spaces. The increment of the recursion is furthermore variable; there is no restriction to one-step-recursions or certain step sizes.

2. This scheme is especially rewarding from a practical point of view when time-varying systems are considered. Assuming that the order of the system remains the same, but the system dynamics change, a reduction of the past horizon in terms of $\mathcal{P}_t^-$ is a meaningful contribution as the calculations perform an averaging over the presented data, which, depending on the rate of the change and the lengths of the horizons, could result in an insufficient model for the current point in time. The predictor space is however an entity which is not influenced by the changing dynamics of the past, as it is equivalent to a state of a certain time. The state represents a map of past input-output
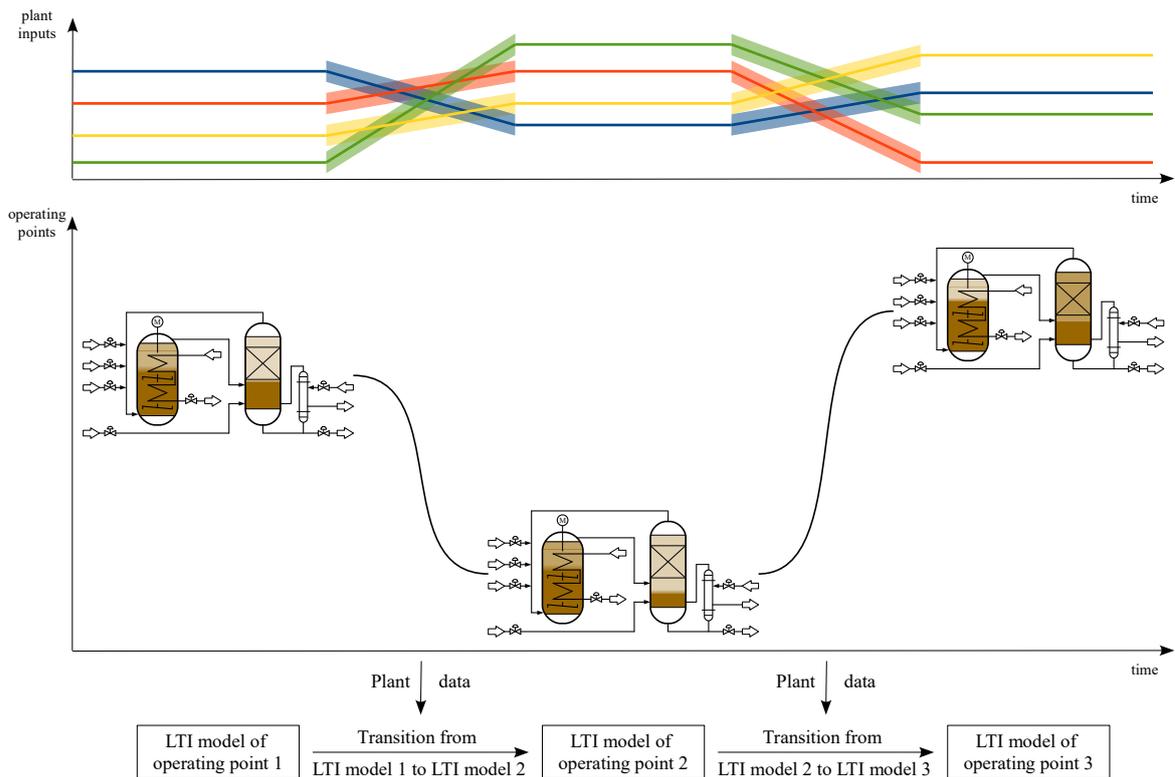
Figure 6.1: Principle of recursive identification by adapting the existing model based on the plant data during the change of an operating point; the colored areas around the inputs symbolizes the excitation for identification

data[1] to the future. This map is in turn governed by the system. That is, the state decouples the change of the system from the information which are transfered by the system from the past to the future, as it contains exactly these information. Hence, by using the predictor space, past information of the system are taken into account while avoiding interferences by the changing of system.

Whereas the first point will be proven in the following, the second point is a mere consequence and extension of the rationale of the recursive procedure. In fact, the proofs are based on the assumption of stationary processes – which is obviously not given if one allows for time-varying systems. Using the recursive scheme for the identification or tracking of time-varying systems, there will be some deviations from the expected values. However, as explained, these errors stem from the averaging effect over the horizons. Thus, if the changes of the dynamics of the system are slow compared to the length the horizons of the intermediate data and the future (minimal length fixed by the order of the system), the identification will give reasonable results[2].

---

[1] Keep in mind that the output is needed to retrieve the information on the unobservable input, i.e., on the innovation process e.

[2] As it will be discussed in Section 6.5.1, the accuracy of the tracking is also influenced by the column number of the tail matrices. If this problem is however circumvented by the explained approach of Section 6.5.1, the here made statement holds.

Similarly to the above considerations, which contain both theoretical aspects and its practical implications, this chapter deals with both the theoretical and practical aspect of the recursive scheme. The major theoretical contribution is given in Section 6.1 and Section 6.2, which the recursive scheme is explained and the basic points are proven in. This derivation is made for both stochastic and joint stochastic-deterministic systems. The principles of the scheme are illustrated by the stochastic systems and are extended to joint stochastic-deterministic systems in the following. The implementation of the recursive scheme in terms of genuine methods is outlined in Section 6.3. Examples in Section 6.4 illustrate the functionality of those methods and serve as a lead-in to the practical aspects discussed in Section 6.5. Parts of the following results are disclosed in Bathelt, Söffker, and Jelali (2018)[3], whereas an initial derivation for open-loop joint stochastic-deterministic systems is disclosed in Bathelt, Söffker, and Jelali (2017).

6.1 Preliminary considerations

In this section, the introductory topics are covered. Those are the assumptions for the derivations and the mutual relations of the different data spaces involved in the calculation. These relations are fundamental as they constitute the point of origin for the derivations of the recursive methodology.

6.1.1 Assumptions

The assumptions for the derivations contain basically restrictions on the processes y, x, and u. These assumptions are:

- The processes y, x, and u are stationary zero-mean Gaussian processes.

- With respect to the ambient data space, the predictor space $\mathcal{X}_t^{+/-}$ is an internal subspace. That is, in terms of stochastic systems (here restriction to the forward predictor space),

$$\mathcal{X}_t^- \subset \mathcal{Y}_t^- \ , \tag{6.1}$$

  and, in terms of joint stochastic-deterministic systems,

$$\mathcal{X}_t^- \subset \mathcal{P}_t^- \ , \tag{6.2}$$

  which makes $\mathcal{X}_t^{+/-}$ also a causal predictor space. In the end, these assumptions are needed regardless of the preceding derivations, as these internal predictor spaces are the only predictor spaces that can be extracted from the data. To some degree, this assumption is also connected to the fact that only constructible predictor spaces are considered or that the system is tacitly assumed to be reachable.

- The processes y and u fulfill the richness condition (Lindquist and Picci, 2015, p. 695)

$$\left(\mathcal{Y}_t^+ \vee \mathcal{U}_t^+\right) \cap \left(\mathcal{Y}_t^- \vee \mathcal{U}_t^-\right) = \{0\} \ . \tag{6.3}$$

---

[3]This paper contains the derivations and results for stochastic systems and joint stochastic-deterministic systems in open-loop conditions, the numerical implementation of the recursive scheme, the example regarding the influence of the column number presented in Section 6.4.2, and the input scheme of Section 6.5.1.
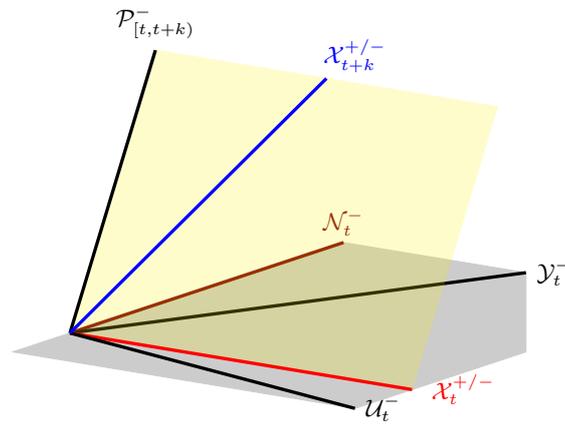
Figure 6.2: Inclusion of the future predictor space in the joint space $\mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k)}^-$

This condition is equivalent to the assumption that the joint process (y, u) has a spectrum which is bounded away from zero.

- The past and future horizons are assumed to be denoted by k_p and k_f. In the following, k denotes the step width of the update step of the recursion.

With respect to the application in connection with time-varying systems, the first assumption establishes a rather strong restriction. This can be resolved if a weaker assumption is made. As the changes of the system, i.e., changes between different operating points, are governed by the input, the input can be assumed to consist of two subprocesses u_1 and u_2. Let both subprocess be defined by the innovation systems $\Sigma_{u_1} = (A_{u_1}, K_{u_1}, C_{u_1}, I)$ and $\Sigma_{u_1} = (A_{u_2}, K_{u_2}, C_{u_2}, I)$, where the relation of the spectral radii[4] of A_{u_1} and A_{u_2} and the variances of u_1 and u_2 obey

$$\rho(A_{u_1}) << \rho(A_{u_2}), \qquad \Sigma_{u_1 u_1} << \Sigma_{u_2 u_2} . \tag{6.4}$$

Then, the identification can be made in terms of y_1 and u_1, as subprocess u_2 is only responsible for the change of the operating point, whereas u_1 is the usual excitation process. If the changes in u_2 are slow enough to be negligible over the horizons k_p and k_f, standard identification in terms of y and u can be made.

6.1.2 Subspace inclusions

The inclusions of certain subspaces form the basis for the recursive scheme, as these inclusions describe a way to define new predictor spaces. The definitions for stochastic systems and joint stochastic-deterministic systems differ due to the exogenous input present in the latter one. Hence, the definitions are separated according to these two types of systems. Figure 6.2 (in terms of the joint stochastic-deterministic case) illustrates the statements of the following lemma and corollaries.

---

[4]The spectral radius of a matrix is given by

$$\rho(A) = \max_i |\lambda_i(A)| .$$

Stochastic systems

For stochastic systems the statement of Theorem 2.6 already declares a way for the definition of new predictor spaces. However, this definition requires the knowledge of the wandering subspace $\mathcal{W}_t$. This theorem can be restated in terms of spaces spanned by y.

Lemma 6.1. *(Bathelt, Söffker, and Jelali, 2018) Let $\mathcal{X}_t^{+/-}$ be an internal forward Markovian splitting subspace. Then,*

$$\mathcal{X}_{t+1}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{Y}_t \, . \tag{6.5}$$

Proof. Theorem 2.6 defines

$$\mathcal{X}_{t+1}^{+/-} \subset \mathcal{X}_t^{+/-} \oplus \mathcal{W}_t \, ,$$

where

$$\mathcal{W}_t = \mathcal{S}_{t+1}^- \ominus \mathcal{S}_t^- \, . \tag{6.6}$$

If $\mathcal{X}_t^{+/-}$ is a internal forward Markovian splitting subspace, then

$$\mathcal{S}_t^- = \mathcal{Y}_t^- \vee \mathcal{X}_t^- = \mathcal{Y}_t^- \, ,$$

where

$$\mathcal{X}_t^- = \bigvee_{k=-\infty}^{t} \mathcal{X}_k^{+/-} \, .$$

Hence, by (6.6) the wandering subspace is given as $\mathcal{W}_t = \mathcal{S}_{t+1}^- \ominus \mathcal{S}_t^- = \mathcal{Y}_{t+1}^- \ominus \mathcal{Y}_t^-$. By (2.54a) (cf. proof of Proposition 17.3.15 in Lindquist and Picci, 2015)

$$\begin{aligned}
\mathcal{Y}_{t+1}^- \ominus \mathcal{Y}_t^- &= (\mathcal{Y}_t^- \vee \mathcal{Y}_t) \ominus \mathcal{Y}_t^- = \mathcal{Y}_t \ominus \mathcal{Y}_t^- \\
&= \mathcal{Y}_t - \hat{E}\left\{\mathcal{Y}_t \,\middle|\, \mathcal{Y}_t^-\right\} = \mathcal{Y}_t - \hat{E}\left\{\mathcal{Y}_t \,\middle|\, \mathcal{X}_t^{+/-}\right\} \\
&\subset \mathcal{Y}_t \vee \mathcal{X}_t^{+/-} \, ,
\end{aligned}$$

follows, which in turn proofs (6.5). $\qquad\square$

Note that this lemma holds only true for the case of the forward predictor space or forward Markovian splitting subspace. If $\mathcal{X}_t^{-/+}$ would be the backward predictor space, the inclusion would read

$$\mathcal{X}_{t-1}^{-/+} \subset \mathcal{X}_t^{-/} \vee \mathcal{Y}_t \, .$$

From Lemma 6.1 an extension to more than a one-step-ahead inclusion can readily deduced.

Corollary 6.1. *(Bathelt, Söffker, and Jelali, 2018) For any predictor space, which lies k steps into the future, the inclusion*

$$\mathcal{X}_{t+k}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \tag{6.7}$$

holds.

Systems with exogenous inputs

The transition to systems with exogenous inputs is now obvious and follows from Proposition 2.12, which is the equivalent to Lemma 6.1[5]. Iterating the inclusion for the predictor space, a general inclusion for any future predictor space can be concluded.

Corollary 6.2. *(Bathelt, Söffker, and Jelali, 2017) For any predictor space, which lies k steps into the future, the inclusion*

$$\mathcal{X}_{t+k}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \tag{6.8}$$

holds.

Figure 6.2 illustrates this inclusion. The past $\mathcal{P}_t^- = \mathcal{Y}_t^- \vee \mathcal{U}_t^-$ can be constructed as

$$\mathcal{P}_t^- = \mathcal{X}_t^{+/-} \oplus \mathcal{N}_t^-$$

by the predictor space and its junk space $\mathcal{N}_t^-$, i.e., the subspace containing no information on the future. No component, which belongs to $\mathcal{N}_t^-$, has an influence on $\mathcal{X}_{t+k}^{+/-}$. The minimal oblique Markovian splitting subspace $\mathcal{X}_{t+k}^{+/-}$ of time $t+k$ is hence contained only in the joint space spanned by the predictor space $\mathcal{X}_t^{+/-}$ and the data up to time $t+k$. Simply put, by reason of the Markov property of $\mathcal{X}_t^{+/-}$, only this subspace of the past influence the future.

6.2 Determination of predictor space

Now that the subspaces which contain $\mathcal{X}_{t+k}^{+/-}$ are defined, the next steps concern the definition of suitable (oblique) Markovian splitting subspaces and the determination of a minimal (oblique) Markovian splitting subspace. The latter problem follows from the former one. Hence, first some general (oblique) Markovian splitting subspaces for the stochastic system and the joint stochastic-deterministic system are defined. Subsequently, they are reduced to the respective minimal (oblique) Markovian splitting subspaces, i.e., the sought-after predictor spaces.

Stochastic systems

Based on Corollary 6.1, let a subspace $\mathcal{X}_{t+k}$ be defined as

$$\mathcal{X}_{t+k} := \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- . \tag{6.9}$$

Lemma 6.2. *(Bathelt, Söffker, and Jelali, 2018) The space $\mathcal{X}_{t+k}$ defined by (6.9) is a Markovian splitting subspace such that*

$$\left(\mathcal{Y}_{t+k}^- \vee \mathcal{X}_{t+k}^-\right) \perp \left(\mathcal{Y}_{t+k}^+ \vee \mathcal{X}_{t+k}^+\right) \mid \mathcal{X}_{t+k} , \tag{6.10}$$

---

[5]Note that the approach to stochastic systems and joint stochastic-deterministic systems is different, as the Markovian splitting subspace of stochastic systems usually contains both the forward and backward predictor spaces. For the joint stochastic-deterministic case, the derivations in Lindquist and Picci (2015) are however already restricted to internal/causal and hence forward predictor spaces.

and the respective scattering pair fulfills the invariance property (2.58).

Proof. To proof that the space defined in (6.9) is a Markovian splitting subspace, the splitting and Markov properties need to be shown. Therefore, let the respective past and future spaces of $\mathcal{X}_{t+k}$, $\mathcal{X}_{t+k}^{-}$ and $\mathcal{X}_{t+k}^{+}$, in (6.10) be defined as

$$\mathcal{X}_{t+k}^{-} = \bigvee_{i=-\infty}^{0} \mathcal{X}_{t+i}^{+/-} \vee \mathcal{Y}_{[t+i,t+k+i)}^{-} \,,$$

$$\mathcal{X}_{t+k}^{+} = \bigvee_{i=0}^{\infty} \mathcal{X}_{t+i}^{+/-} \vee \mathcal{Y}_{[t+i,t+k+i)}^{-} \,.$$

Using these spaces, the scattering pair $(\mathcal{S}_t^{-}, \mathcal{S}_t^{+})$ of $\mathcal{X}_{t+k}$ is defined as in (2.64). First, the Markov property is shown in terms of the invariance property (2.58). By assumption, $\mathcal{X}_t^{+/-}$ is a forward internal Markovian splitting subspace, which subsequently means that $\mathcal{X}_{t+k}^{-} \subset \mathcal{Y}_{t+k}^{-}$ holds. Thus, $\mathcal{S}_{t+k}^{-} = \mathcal{Y}_{t+k}^{-}$ and $\mathcal{S}_{t+k-1}^{-} \subset \mathcal{S}_{t+k}^{-}$ trivially follows. The inclusion with respect to $\mathcal{S}_{t+k}^{+}$ is given as

$$\begin{aligned}
\mathcal{S}_{t+k}^{+} &= \mathcal{Y}_{t+k}^{+} \vee \mathcal{X}_{t+k}^{+} \\
&= \mathcal{Y}_{t+k+1}^{+} \vee \mathcal{X}_{t+k+1}^{+} \vee (\mathcal{Y}_t \vee \mathcal{X}_t^{+/-}) \\
&= \mathcal{S}_{t+k+1}^{+} \vee (\mathcal{Y}_t \vee \mathcal{X}_t^{+/-}) \,.
\end{aligned}$$

Hence, $\mathcal{S}_{t+k+1}^{+} \subset \mathcal{S}_{t+k}^{+}$. The splitting property follows from

$$\hat{E}\{\mathcal{S}_t^{+} \mid \mathcal{S}_t^{-}\} = \hat{E}\{\mathcal{S}_t^{+} \mid \mathcal{X}_t^{+/-}\} \,,$$

i.e.,

$$\hat{E}\{\lambda \mid \mathcal{S}_t^{-}\} = \hat{E}\{\lambda \mid \mathcal{X}_t^{+/-}\} \; \forall \, \lambda \in \mathcal{S}_t^{+} \,.$$

As $\mathcal{S}_{t+k}^{+} \subset \mathcal{S}_t^{+}$, this means in particular

$$\hat{E}\{\mathcal{S}_{t+k}^{+} \mid \mathcal{S}_t^{-}\} = \hat{E}\{\mathcal{S}_{t+k}^{+} \mid \mathcal{X}_t^{+/-}\} \,.$$

Due to the restriction to forward internal Markovian splitting subspaces, $\mathcal{S}_{t+k}^{-} = \mathcal{S}_t^{-} \vee \mathcal{Y}_{[t,t+k)}^{-}$. Hence,

$$\begin{aligned}
\hat{E}\{\mathcal{S}_{t+k}^{+} \mid \mathcal{S}_{t+k}^{-}\} &= \hat{E}\{\mathcal{S}_{t+k}^{+} \mid \mathcal{S}_t^{-} \vee \mathcal{Y}_{[t,t+k)}^{-}\} \\
&= \hat{E}\{\mathcal{S}_{t+k}^{+} \mid \mathcal{X}_t^{+/-} + \mathcal{Y}_{[t,t+k)}^{-}\} \\
&= \hat{E}\{\mathcal{S}_{t+k}^{+} \mid \mathcal{X}_{t+k}\}
\end{aligned}$$

where the second equation follows as a consequence of the richness condition (6.3) (see also Lemma 2.7.3 in Lindquist and Picci, 2015)[6]. Thus, a subspace as defined in (6.9) fulfills the necessary properties to make it a Markovian splitting subspace. $\qquad\square$

---

[6]The whole step of the proof uses the fact proven by Lemma 2.7.3 in Lindquist and Picci (2015) that

$$\hat{E}\{\mathcal{A} \mid \mathcal{B} \vee \mathcal{C}\} = \hat{E}\{\mathcal{A} \mid \mathcal{D} \vee \mathcal{C}\} \,,$$

if $\mathcal{B} \subset \mathcal{D}$, and $\mathcal{D} \cap \mathcal{C} = \{0\}$.

In general, the subspace defined by (6.9) is not a minimal Markovian splitting subspace. This is similar to the space spanned by the whole past of a process. Although it is also a Markovian splitting subspace, it is obviously not a minimal one. This problem, i.e., the extraction of the minimal Markovian splitting subspace contained in $\mathcal{X}_{t+k}$, is dealt with by the next theorem.

Theorem 6.1. *(Bathelt, Söffker, and Jelali, 2018) The minimal forward Markovian splitting subspace contained in $\mathcal{X}_{t+k}$ is given by*

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}\left\{ \mathcal{Y}_{t+k}^{+} \middle| \mathcal{X}_{t+k} \right\} . \tag{6.11}$$

Proof. To show that $\mathcal{X}_{t+k}^{+/-}$ as defined in (6.11) is a minimal Markovian splitting subspace, it suffices to show that it is both observable and constructible.

Constructibility follows by construction as

$$\mathcal{X}_{t+k} \cap \left(\mathcal{Y}_{t+k}^{-}\right)^{\perp} = \{0\} .$$

Hence, the minimal Markovian splitting subspace $\mathcal{X}_{t+k}^{+/-}$ is given by the observable subspace of $\mathcal{X}_{t+k}$. In terms of the scattering pair $(\mathcal{S}_{t+k}^{-}, \mathcal{S}_{t+k}^{+})$, which is associated with $\mathcal{X}_{t+k}$,

$$\begin{aligned} \mathcal{X}_{t+k} &= \mathcal{S}_{t+k}^{-} \cap \mathcal{S}_{t+k}^{+} \\ &= \hat{E}\{\mathcal{Y}_{t+k}^{+} \mid \mathcal{S}_{t+k}^{-} \cap \mathcal{S}_{t+k}^{+}\} \oplus \left(\mathcal{S}_{t+k}^{-} \cap \mathcal{S}_{t+k}^{+}\right) \cap \left(\mathcal{Y}_{t+k}^{+}\right)^{\perp} \\ &= \hat{E}\{\mathcal{Y}_{t+k}^{+} \mid \mathcal{X}_{t+k}\} \oplus \mathcal{X}_{t+k} \cap \left(\mathcal{Y}_{t+k}^{+}\right)^{\perp} , \end{aligned} \tag{6.12}$$

where

$$\mathcal{N}_{t+k}^{-} := \mathcal{X}_{t+k} \cap \left(\mathcal{Y}_{t+k}^{+}\right)^{\perp}$$

defines the unobservable subspace. Let the orthogonal projection in (6.12) define a subspace of $\mathcal{X}_{t+k}$ as

$$\hat{E}\{\mathcal{Y}_{t+k}^{+} \mid \mathcal{S}_{t+k}^{-} \cap \mathcal{S}_{t+k}^{+}\} = \mathcal{X}_{t+k}^{*} = \hat{\mathcal{S}}_{t+k}^{-} \cap \hat{\mathcal{S}}_{t+k}^{+} ,$$

where $(\hat{\mathcal{S}}_{t+k}^{-}, \hat{\mathcal{S}}_{t+k}^{+})$ is the respective scattering pair of $\mathcal{X}_{t+k}^{*}$. By construction, $\left(\hat{\mathcal{S}}_{t+k}^{-} \cap \hat{\mathcal{S}}_{t+k}^{+}\right) \cap \left(\mathcal{Y}_{t+k}^{+}\right)^{\perp} = \{0\}$, and hence (cf. proof of Theorem 7.4.9 in Lindquist and Picci, 2015)

$$\mathcal{Y}_{t+k}^{+} \vee \mathcal{Y}_{t+k}^{-} = \left(\hat{\mathcal{S}}_{t+k}^{-}\right)^{\perp} \vee \left(\hat{\mathcal{S}}_{t+k}^{+}\right)^{\perp} \vee \mathcal{Y}_{t+k}^{+} = \left(\hat{\mathcal{S}}_{t+k}^{+}\right)^{\perp} \oplus \left(\left(\hat{\mathcal{S}}_{t+k}^{-}\right)^{\perp} \vee \mathcal{Y}_{t+k}^{+}\right) .$$

Thus[7],

$$\hat{\mathcal{S}}_{t+k}^{+} = (\hat{\mathcal{S}}_{t+k}^{-})^{\perp} \vee \mathcal{Y}_{t+k}^{+}$$

states that $\mathcal{X}_{t+k}^{*} \subset \mathcal{X}_{t+k}$ is observable, making it a minimal Markovian splitting subspace, and

$$\mathcal{X}_{t+k}^{+/-} := \mathcal{X}_{t+k}^{*} .$$

$\square$

---

[7] $\mathcal{H} = \left(\hat{\mathcal{S}}_{t+k}^{+}\right)^{\perp} \oplus \left(\left(\hat{\mathcal{S}}_{t+k}^{-}\right)^{\perp} \vee \mathcal{Y}_{t+k}^{+}\right) \Rightarrow \mathcal{H} \ominus \left(\hat{\mathcal{S}}_{t+k}^{+}\right)^{\perp} = \hat{\mathcal{S}}_{t+k}^{+} = \left(\hat{\mathcal{S}}_{t+k}^{-}\right)^{\perp} \vee \mathcal{Y}_{t+k}^{+}$

Remark 6.1. To show that the predictor space in (6.11) is a minimal predictor space without explicitly proving observability and constructibility, minimality follows also from Proposition 2.3. By Proposition 2.3,

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}\left\{\mathcal{Y}_{t+k}^+ \middle| \mathcal{Y}_{t+k}^-\right\}$$

is a minimal Markovian splitting subspace. As $\mathcal{X}_{t+k}$ is already shown to be a Markovian splitting subspace,

$$\hat{E}\left\{\mathcal{Y}_{t+k}^+ \middle| \mathcal{Y}_{t+k}^-\right\} = \hat{E}\left\{\mathcal{Y}_{t+k}^+ \middle| \mathcal{X}_{t+k}\right\}$$

holds.

From Theorem 6.1, the calculation of a minimal Markovian splitting subspace or minimal predictor space based on knowledge of any past predictor space and the intermediate data space is hence given by the afore claimed

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}\left\{\mathcal{Y}_{t+k}^+ \middle| \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{X}_t^{+/-}\right\} . \tag{6.13}$$

Systems with exogenous inputs

The derivation of the recursive scheme for the joint stochastic-deterministic case follows the derivation for the stochastic systems. However, due to the existence of the exogenous inputs and the possible feedback of the output, some statements change in comparison to the stochastic system. In this context, the condition

$$\mathcal{S}_t^- \cap \mathcal{F}_t^+ = \{0\} , \tag{6.14}$$

where

$$\mathcal{S}_t^- = \mathcal{P}_t^- \vee \mathcal{X}_t^-, \qquad \mathcal{F}_t^+ = \mathcal{U}_t^+ \vee \mathcal{E}_t^+$$

is needed. This condition is essential for the subsequent derivations. Only if this condition is fulfilled, the correct decomposition of the future output into its past and future components is possible. Here, $\mathcal{F}_t^+$ is defined based on the future space of the innovation process, as the wandering subspace $\mathcal{W}_t$ becomes the innovation process $\mathcal{E}_t$ under the assumption of causality of the predictor space.

Again, define a subspace

$$\mathcal{X}_{t+k} := \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- , \tag{6.15}$$

which will serve as basis for the minimal oblique Markovian splitting subspace.

Lemma 6.3. *The space $\mathcal{X}_{t+k}$ defined by* (6.15) *is an oblique Markovian splitting subspace such that*

$$\left(\mathcal{P}_{t+k}^- \vee \mathcal{X}_{t+k}^-\right) \perp \left(\mathcal{Y}_{t+k}^+ \vee \mathcal{X}_{t+k}^+\right) \middle| \mathcal{X}_{t+k} \vee \mathcal{F}_{t+k}^+ , \tag{6.16}$$

and the respective scattering pair fulfill the invariance property (2.87b)*.*

Proof. The proof is similar to that of Lemma 6.2. By showing the oblique splitting and oblique Markov properties, the space $\mathcal{X}_{t+k}$ defined in (6.15) is shown to be an oblique Markovian

splitting subspace. Let the respective past and future spaces of $\mathcal{X}_{t+k}$, $\mathcal{X}_{t+k}^-$ and $\mathcal{X}_{t+k}^+$, in (6.16) be defined as

$$\mathcal{X}_{t+k}^- = \bigvee_{i=-\infty}^{0} \mathcal{X}_{t+i}^{+/-} \vee \mathcal{Y}_{[t+i,t+k+i)}^- \vee \mathcal{U}_{[t+i,t+k+i)}^- \,,$$

$$\mathcal{X}_{t+k}^+ = \bigvee_{i=0}^{\infty} \mathcal{X}_{t+i}^{+/-} \vee \mathcal{Y}_{[t+i,t+k+i)}^- \vee \mathcal{U}_{[t+i,t+k+i)}^- \,.$$

Based on theses spaces, the scattering pair $(\mathcal{S}_{t+k}^-, \mathcal{S}_{t+k}^+)$ of $\mathcal{X}_{t+k}$ is defined as in (2.90). First, the Markov property is shown in terms of the invariance property (2.87b). By assumption, $\mathcal{X}_t^{+/-}$ is internal and causal, which subsequently means that $\mathcal{S}_{t+k}^- = \mathcal{Y}_{t+k}^- \vee \mathcal{U}_{t+k}^-$ and $\mathcal{S}_{t+k-1}^- \subset \mathcal{S}_{t+k}^-$ follows. The inclusion with respect to $\mathcal{S}_{t+k}^+$ is given as

$$\begin{aligned}
\mathcal{S}_{t+k}^+ &= \mathcal{Y}_{t+k}^+ \vee \mathcal{X}_{t+k}^+ \\
&= \mathcal{Y}_{t+k+1}^+ \vee \mathcal{X}_{t+k+1}^+ \vee (\mathcal{Y}_t \vee \mathcal{U}_t \vee \mathcal{X}_t^{+/-}) \\
&= \mathcal{S}_{t+k+1}^+ \vee (\mathcal{Y}_t \vee \mathcal{U}_t \vee \mathcal{X}_t^{+/-}) \,.
\end{aligned}$$

Hence, $\mathcal{S}_{t+k+1}^+ \subset \mathcal{S}_{t+k}^+$. The splitting property follows from

$$\hat{E}\{\mathcal{S}_t^+ \mid \mathcal{S}_t^- \vee \mathcal{F}_t^+\} = \hat{E}\{\mathcal{S}_t^+ \mid \mathcal{X}_t^{+/-} \vee \mathcal{F}_t^+\} \,,$$

which again implies by $\mathcal{S}_{t+k}^+ \subset \mathcal{S}_t^+$ that

$$\hat{E}\{\mathcal{S}_{t+k}^+ \mid \mathcal{S}_t^- \vee \mathcal{F}_t^+\} = \hat{E}\{\mathcal{S}_{t+k}^+ \mid \mathcal{X}_t^{+/-} \vee \mathcal{F}_t^+\} \,.$$

Due to the restriction to internal and forward predictor spaces, $\mathcal{S}_{t+k}^- = \mathcal{S}_t^- \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^-$. Hence,

$$\begin{aligned}
\hat{E}\{\mathcal{S}_{t+k}^+ \mid \mathcal{S}_{t+k}^- \vee \mathcal{F}_t^+\} &= \hat{E}\{\mathcal{S}_{t+k}^+ \mid \mathcal{S}_t^- \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \vee \mathcal{F}_t^+\} \\
&= \hat{E}\left\{\mathcal{S}_{t+k}^+ \,\middle|\, \mathcal{X}_t^{+/-} + \left(\mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- + \mathcal{F}_t^+\right)\right\} \\
&= \hat{E}\{\mathcal{S}_{t+k}^+ \mid \mathcal{X}_{t+k} + \mathcal{F}_t^+\} \,,
\end{aligned}$$

where the third equation follows as a consequence of the richness condition (6.3) (see also Lemma 2.7.3 in Lindquist and Picci, 2015). Hence, the subspace as defined in (6.15) is an oblique Markovian splitting subspace. $\qquad\square$

Seen in terms of the assumption $\mathcal{X}_{t+k}^- \subset \mathcal{P}_{t+k}^-$, condition (6.14) is already fulfilled by the richness condition, i.e., by

$$\mathcal{P}_{t+k}^- \cap \mathcal{F}_{t+k}^+ = \{0\} \,. \tag{6.17}$$

Thus,

$$\hat{E}_{\|\mathcal{F}_{t+k}^+}\{\mathcal{S}_{t+k}^+ \mid \mathcal{S}_{t+k}^-\} = \hat{E}_{\|\mathcal{F}_{t+k}^+}\{\mathcal{S}_{t+k}^+ \mid \mathcal{X}_{t+k}\} \tag{6.18}$$

is furthermore equivalent to

$$\hat{E}_{\|\mathcal{F}_{t+k}^+}\{\mathcal{Y}_{t+k}^+ \mid \mathcal{P}_{t+k}^-\} = \hat{E}_{\|\mathcal{F}_{t+k}^+}\{\mathcal{Y}_{t+k}^+ \mid \mathcal{X}_{t+k}\} \tag{6.19}$$

Hence, the minimal oblique Markovian splitting subspace can be determined as given by the following theorem.

Theorem 6.2. *The minimal oblique Markovian splitting subspace contained in $\mathcal{X}_{t+k}$ is given by*

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{F}_{t+k}^+} \{\mathcal{Y}_{t+k}^+ \mid \mathcal{X}_{t+k}\} \, . \tag{6.20}$$

Proof. By Lemma 6.3, $\mathcal{X}_{t+k}$ is an oblique Markovian splitting subspace, so that

$$\hat{E}_{\|\mathcal{F}_{t+k}^+} \{\mathcal{Y}_{t+k}^+ \mid \mathcal{P}_{t+k}^-\} = \hat{E}_{\|\mathcal{F}_{t+k}^+} \{\mathcal{Y}_{t+k}^+ \mid \mathcal{X}_{t+k}\} \, .$$

In terms of Theorem 2.8, the left-hand side of the above equation defines the minimal oblique Markovian splitting subspace contained in $\mathcal{P}_{t+k}^-$. Thus,

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{F}_{t+k}} \{\mathcal{Y}_{t+k}^+ \mid \mathcal{X}_{t+k}\} \, .$$

$\square$

By means of this theorem the recursive determination of the minimal predictor space for systems with exogenous inputs in the general structure, which includes feedback, is given by

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{F}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \left| \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \right. \right\} \, . \tag{6.21}$$

As $\mathcal{X}_t^{+/-}$ is assumed to be an internal predictor space, the reasoning of Theorem 6.2 holds also for the one-step-ahead oblique Markovian splitting subspace (cf. Proposition 17.3.11 in Lindquist and Picci, 2015). Hence, similar to Proposition 2.4, the oblique predictor space can be defined without taking the innovations in $\mathcal{F}_{t+k}$ explicitly into consideration.

Corollary 6.3. *Let the system be finite-dimensional with order n. Then, the oblique predictor space is given by the vector sum*

$$\mathcal{X}_{t+k}^{+/-} = \bigvee_{h=0}^{n-1} \mathcal{X}_{t+k}^h \, , \tag{6.22}$$

with

$$\mathcal{X}_{t+k}^h = \hat{E}_{\|\mathcal{U}_{t+k}}^{\mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k)}^-} \left\{ \hat{E}_{\|\mathcal{U}_{t+k+1}}^{\mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k+1)}^-} \left\{ \cdots \left\{ \hat{E}_{\|\mathcal{U}_{t+k+h}}^{\mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k+h)}^-} \{\mathcal{Y}_{t+k+h}\} \right\} \right\} \right\} \, , \tag{6.23}$$

where $\hat{E}_{\|\mathcal{U}_{t+k+i}}^{\mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k+i)}^-} \{\cdot\}$, $i = 0, \cdots h$ *is a short-hand notation for* $\hat{E}_{\|\mathcal{U}_{t+k+i}}\{\cdot \mid \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k+i)}^- \vee \mathcal{U}_{[t,t+k+i)}^-\}$.

Similar to the underlying calculations of the PBSID method as given in Lemma 2.8, the recursive implementation of this method follows from Corollary 6.3

Corollary 6.4. *Let the system be finite-dimensional with order n. Then, oblique predictor space is given by the vector sum*

$$\mathcal{X}_{t+k}^{+/-} = \bigvee_{h=0}^{n-1} \mathcal{X}_{t+k}^h \, , \tag{6.24}$$

where the subspaces $\mathcal{X}_{t+k}^h$ can be extracted as

$$\mathcal{X}_{t+k}^h = \hat{E}_{\|\mathcal{P}_{[t+k,t+k+h)}^+} \left\{ \mathcal{Y}_{t+k+h} \left| \mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k)}^- \right. \right\} . \tag{6.25}$$

If the feedback is dropped, the space of the future innovation is by assumption orthogonal to both the future inputs and the future outputs. In this case, $\mathcal{F}_{t+k}^+$ can be reduced to $\mathcal{U}_{t+k}^+$.

Corollary 6.5. *(Bathelt, Söffker, and Jelali, 2017) Let the system operate under open-loop conditions. Then, the minimal oblique Markovian splitting subspace contained in $\mathcal{X}_{t+k}$ is given by*

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \left| \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \right. \right\} . \tag{6.26}$$

This corollary is the consequence of

$$\hat{E}\{\mathcal{A}|\mathcal{B} \vee \mathcal{C}_1 \oplus \mathcal{C}_2\} = \hat{E}\{\mathcal{A}|\mathcal{B} \vee \mathcal{C}_1\} ,$$

if $\mathcal{C}_2 \perp \mathcal{A}$. A thorough proof of the corollary can be found in Bathelt, Söffker, and Jelali (2017). That proof shows that the space given by the projection in (6.26) is both observable and constructible and hence minimal.

In terms of the orthogonal decomposition, the respective determination of the deterministic predictor space follows from the open-loop calculation by reducing the intermediate data space to $\mathcal{U}_{[t,t+k)}^-$ and replacing the past predictor space by its deterministic component $\mathcal{X}_t^{d,+/-}$.

Corollary 6.6. *(Bathelt, Söffker, and Jelali, 2018) Let $\mathcal{X}_t^{d,+/-}$ be the deterministic predictor space of time t. Then, the deterministic predictor space of time $t + k$ is given by*

$$\mathcal{X}_{t+k}^{d,+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \left| \mathcal{X}_t^{d,+/-} \vee \mathcal{U}_{[t,t+k)}^- \right. \right\} . \tag{6.27}$$

Proof. Let $\mathcal{X}_t^{d,+/-}$ be the predictor space of the deterministic subsystem and $\mathcal{X}_t^{s,+/-}$ be the predictor space of the stochastic subsystem. Then the inclusion (see Proposition 2.5)

$$\mathcal{X}_t^{+/-} \subset \mathcal{X}_t^{d,+/-} \oplus \mathcal{X}_t^{s,+/-}$$

holds for the joint predictor space. This statement reflects the case in which the deterministic and stochastic system join certain dynamics. Equality holds only for disjoint dynamics. Taking this connection between $\mathcal{X}_t^{+/-}$ and $\mathcal{X}_t^{d,+/-}$, $\mathcal{X}_t^{s,+/-}$ into consideration, (6.26) can be decomposed as

$$\begin{aligned}
\mathcal{X}_{t+k}^{+/-} &= \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \left| \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \right. \right\} \\
&\subset \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \left| \left(\mathcal{X}_t^{d,+/-} \vee \mathcal{U}_{[t,t+k)}^- \right) \oplus \left(\mathcal{X}_t^{s,+/-} \vee \tilde{\mathcal{Y}}_{[t,t+k)}^- \right) \right. \right\} = \mathcal{X}_{t+k}^{d,+/-} \oplus \mathcal{X}_{t+k}^{s,+/-} ,
\end{aligned}$$

where $\tilde{\mathcal{Y}}_{[t,t+k)}^- = \mathcal{Y}_{[t,t+k)}^- - \hat{E}\left\{ \mathcal{Y}_{[t,t+k)}^- \left| \mathcal{X}_t^{d,+/-} \vee \mathcal{U}_{[t,t+k)}^- \right. \right\}$. This decomposition of (6.26) is similar to the decomposition of the oblique projection in (4.11). Hence the same reasoning leads to (6.27). $\square$

6.3 Numerical implementation

All recursive approaches discussed in Kameyama, Ohsumi, et al. (2005), Lovera, Gustafsson, and Verhaegen (2000), Mercère, Bako, and Lecuche (2008), and Oku and Kimura (1999, 2002) are based on the numerical implementations of subspace methods. Hence, each of the algorithms result in a purpose-build implementation, i.e., a special implementation, which deviate from the original algorithm of the underlying standard subspace method. In terms of the literature, the basic algorithm is usually MOESP. In contrast to these recursive algorithms, the recursive scheme proposed by Theorem 6.1 and Theorem 6.2 is completely embedded into the theoretical framework of (oblique) Markovian splitting subspaces. The resulting implication is that there is no need for special algorithms. That is, any of the proposed subspace methods can be used as a basis to implement the recursive scheme. In terms of the identification of stochastic systems, the methods outlined in Katayama (2005), Tanaka and Katayama (2006, 2007), and van Overschee and De Moor (1993) might be used. For the identification of joint stochastic-deterministic systems, the (standard) methods outlined in Katayama and Picci (1999), van Overschee and De Moor (1994), and Verhaegen (1994) might be used. Another consequence presents itself in terms of the re-identification intervals. As the past data spaces used for the construction of $\mathcal{X}_{t+k}$ in (6.9) or (6.15) are not restricted to a certain horizon, the re-identification can be carried out over varying intervals. Fixed interval length are not necessary.

In terms of the numerical implementation of the recursive scheme, i.e., the implementation of (6.13) or (6.21) (or any calculation based thereon), the only modification of the methods concerns the QR decomposition of the data matrices (for stochastic systems and joint stochastic-deterministic systems)

$$\begin{bmatrix} Y_t^- \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} , \qquad \begin{bmatrix} U_t^+ \\ P_t^- \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{21} & R_{22} & 0 \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \\ Q_3^{\mathrm{T}} \end{bmatrix} , \qquad (6.28)$$

which are replaced by

$$\begin{bmatrix} \begin{bmatrix} x_N(t_0) \\ Y_{[t_0,t)}^- \end{bmatrix} \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} , \qquad \begin{bmatrix} U_t^+ \\ x_N(t_0) \\ P_{[t_0,t)}^- \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{21} & R_{22} & 0 \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \\ Q_3^{\mathrm{T}} \end{bmatrix} , \qquad (6.29)$$

where t_0 is the time of the previous identification. Depending on the method, the structure of the data matrices on the left-hand side of both equations may vary, but the approach remains the same. The past data matrix P_t^-/Y_t^- is replaced by $x_N(t_0)$ and $P_{[t_0,t)}^-/Y_{[t_0,t)}^-$. Although the main focus was the derivation of a theoretically based approach to recursive

subspace identification and not on the reduction of the numerical load, this second aspect appears quite naturally as a consequence of the recursive scheme. As easily seen from the above equations, the number of rows in the data matrices is reduced in the wake of the replacement of Y_t^-/P_t^- by $\begin{bmatrix} x_N(t_0)^{\mathrm{T}} & \left(Y_{[t_0,t)}^- \right)^{\mathrm{T}} \end{bmatrix} / \begin{bmatrix} x_N(t_0)^{\mathrm{T}} & \left(P_{[t_0,t)}^- \right)^{\mathrm{T}} \end{bmatrix}$. Hence, the matrices are smaller and the computational load is reduced.

The identification proceeds according to the sequence explained in the following (cf. Bathelt, Söffker, and Jelali, 2017, 2018):

1. Initial identification at time t_0 using standard methods; identification of the stochastic or joint stochastic-deterministic system with respectively chosen lengths for the horizons of the past and future, i.e., for $\mathcal{Y}_{t_0}^-$ and $\mathcal{Y}_{t_0}^+$ or $\mathcal{P}_{t_0}^-$ and $\mathcal{Y}_{t_0}^+, \mathcal{U}_{t_0}^+$

2. If not already given, determination of the state $x(t_0)$ via

$$x(t_0) = \mathcal{C}_k y_t^- , \qquad x(t_0) = \mathcal{C}_k^K p_t^- ,$$

where $\mathcal{C}_k$ (stochastic systems) or $\mathcal{C}_k^K$ (joint stochastic-deterministic systems) can be determined alongside $\mathcal{O}_k$ during the sequence of the realization algorithms; see, e.g., Katayama (2005).

3. Second identification at time t_1 based on (6.13) or (6.21); the past consists of $\mathcal{Y}_{[t_0,t_1)}^-$ $\vee$ $\mathcal{X}_{t_0}^{+/-}$ or $\mathcal{P}_{[t_0,t_1)}^- \vee \mathcal{X}_{t_0}^{+/-}$

4. Iteration of identifications for t_2, t_3, $\cdots$, starting with Step 2.

The state estimation of Step 2 needs to be included if a system is identified either by the stochastic methods of Tanaka and Katayama (2006, 2007) and van Overschee and De Moor (1993) or the joint stochastic-deterministic method of Verhaegen (1994), as either the implementations or the methods itself are realization-based.

Remark 6.2. Regarding the question whether the identification should explicitly include a past state or not, the explanations in Chui and Maciejowski (2005) give a reasonable guideline. Here, a lower bound for the length of the past horizon k_p is given with (simplified)

$$k_p \geq 3n ,$$

where n is the order of the system to be identified. That is, as long as the increment between two adjacent identifications is smaller than $3n$, the recursive scheme might be used. Once the increment reaches $3n$, the past data suffices to (re-)identify the system.

6.4 Examples

The simulations given in this section are rather a proof of concept than a comparative study with the existing methods. The primary reason stems from the different key aspects and approaches to the problem – theoretically based approach for time-varying systems vs. numerically based approach for reduction of computational load. The given methods essentially update the sample covariance matrices, which the identifications are based on. The only method focusing also on the identification of time-varying systems is outlined in Kameyama, Ohsumi, et al. (2005). However, from the way this method works, it is obvious that it essentially performs a standard identification, i.e., there is no reduction of the past in terms of the horizons or any reuse of information from previous identifications except for the results of the previous QR decomposition. Thus, the existing methods do not perform a recursion in terms of information from previous identifications. In short, there is no conceptually similar approach, which the proposed recursive scheme can be compared to.

The examples will be divided into two groups. The first group is concerned with the actual proof of concept. These examples will show that the results of the numerical implementations

are in line with the theoretical derivation. Thus, the systems to be identified are LTI (SISO) systems. The important point will be that the concepts known from standard identifications will carry over to recursive identifications. The second group is concerned with the identification of time-varying systems. The focus will be on the tracking capabilities of the approach. Closed-loop identifications will be skipped in both cases.

Regarding the results of the second group, one aspect should be pointed out beforehand. It has turned out that the accuracy of the tracking of a change within the system is limited by the number of the columns of the data matrices. This effect is not completely unexpected. A sufficiently accurate estimation of the sample covariance matrices of the current time needs an interval of stationary data. Hence, the column number of the tail matrices needed for the estimation of the sample covariances is dictated by the required accuracy of this estimation. This in turn means also that every recursive method suffers from the same problem. However, due to the different approach to recursive subspace identification proposed in this thesis, it is possible to derive a solution for this problem. This will be discussed in detail in Section 6.5.1. Except for Kameyama, Ohsumi, et al. (2005), the existing algorithms exhibit furthermore a much more severe problem, as the estimation of the sample covariance matrices in these schemes sum up all the past data until the current time. Hence, a change in the system which presents itself by a change of the covariance matrices is blurred by the past data. This has led to the inclusion of a forgetting factor in the respective methods of Lovera, Gustafsson, and Verhaegen (2000) and Mercère, Bako, and Lecuche (2008).

6.4.1 LTI SISO system

The examples presented in this subsection are re-runs of the two examples of Section 4.1.4. That is, one example is the identification of an ARMAX system whereas the other one is the identification of a Box-Jenkins system. The setup and properties of identifications and simulations are the same as outlined in Section 4.1.4. The only difference here is with respect to the input signal. For each of the 100 simulations, a new input signal is generated. The disturbance is again also changed for each simulation. The methods/algorithms used for the identification are the recursive versions of CCA (state algorithm), CCA-ORT and the QR decomposition-based algorithm of the PBSID method (see Appendix B.4, Appendix B.5, and Appendix B.6 for the implementations). The interval between each identification is set to four time steps. It should be noted that a correct non-recursive identification requires for both examples a past horizon of at least $k_p = 15$ (order of $P(z)$ is $n = 5$)[8]. The initial identifications are made as outlined in Section 4.1.4.

The resulting Bode plots of the identification are shown for the ARMAX example in Figure 6.3 and for the Box-Jenkins example in Figure 6.4. Both figures show the expected results, i.e., the Bode plots indicate that the theoretical derivation of (6.26) and (6.27) can be implemented without any difficulties. Regardless of the underlying calculation, i.e., irrespective of the use of the past data of 30 time steps or the previous predictor space and past data of four time steps, the results of the identifications remain the same without showing significant deviations from each other. The small variations visible in the respective Bode plots of each method for different recursion cycles can be attributed to the numerical differences of the

---

[8]For a correct identification of the Box-Jenkins system by the CCA and PBSID methods, the model order needs actually to be set to 7, and the resulting minimal length of the past horizon is $k_p = 21$. However, the Box-Jenkins example given here aims at the illustration of an identification with unknown (non-white) disturbances on the output of the system and non-white input signals, i.e., the worst case encountered in terms of identifications.

data. Furthermore, the general behavior of the methods remains also for the recursive implementation. The identification of the systems which share the dynamics, i.e., the stochastic and deterministic subsystems have the same dynamics, is performed correctly by each of the methods. In terms of the Box-Jenkins structure, i.e., the identification of a system, which is subjected to an arbitrarily colored noise, the correct results are only given by methods based on the ORT approach (see Figure 6.4).
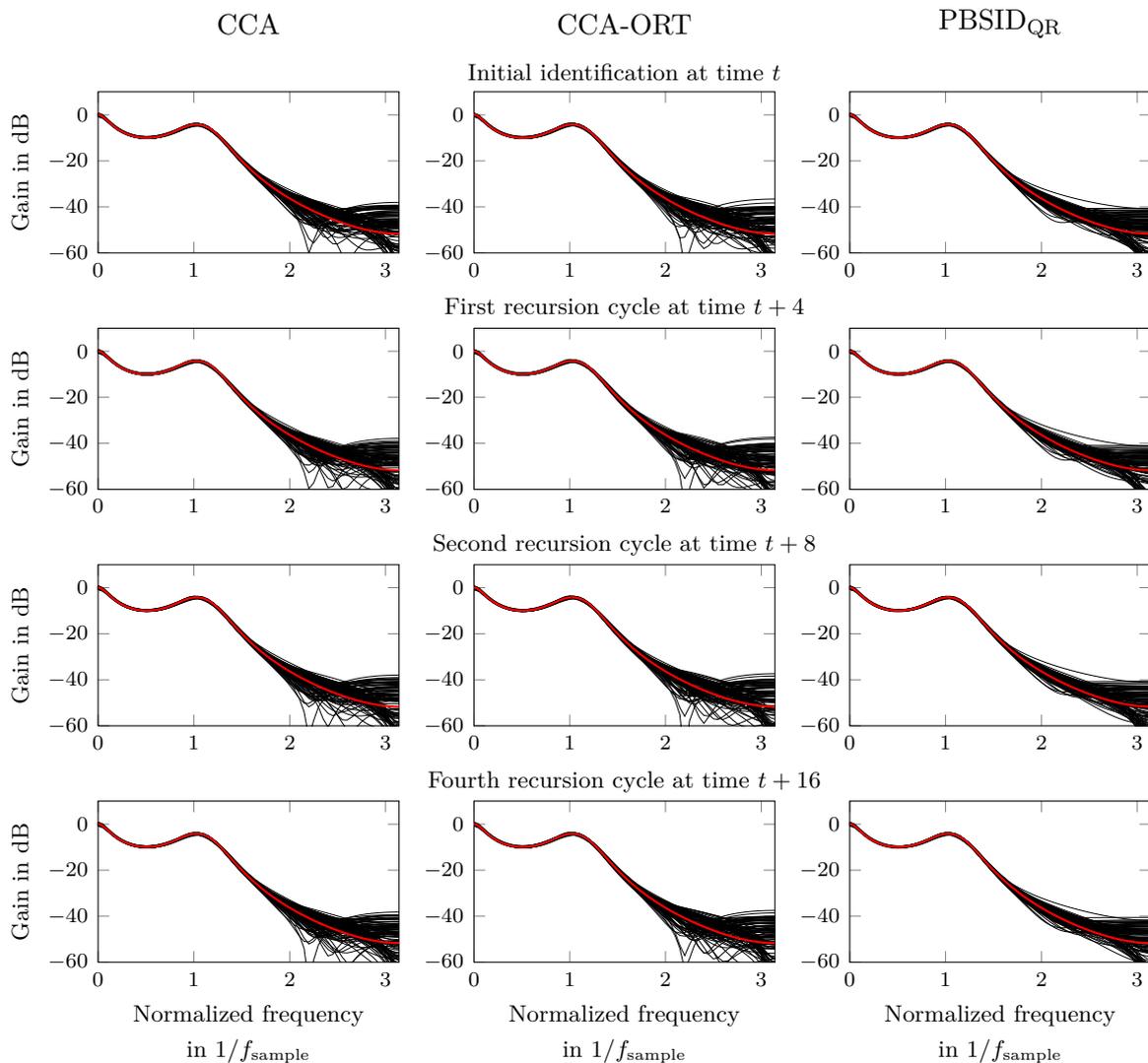

Figure 6.3: Recursive identification of the LTI ARMAX system with CCA, CCA-ORT, and PBSID$_{QR}$ (from left to right): plots show the identification results of the initial identification and the first, second, and fourth recursion cycle (red: nominal value of $P(z)$, black: identified models)
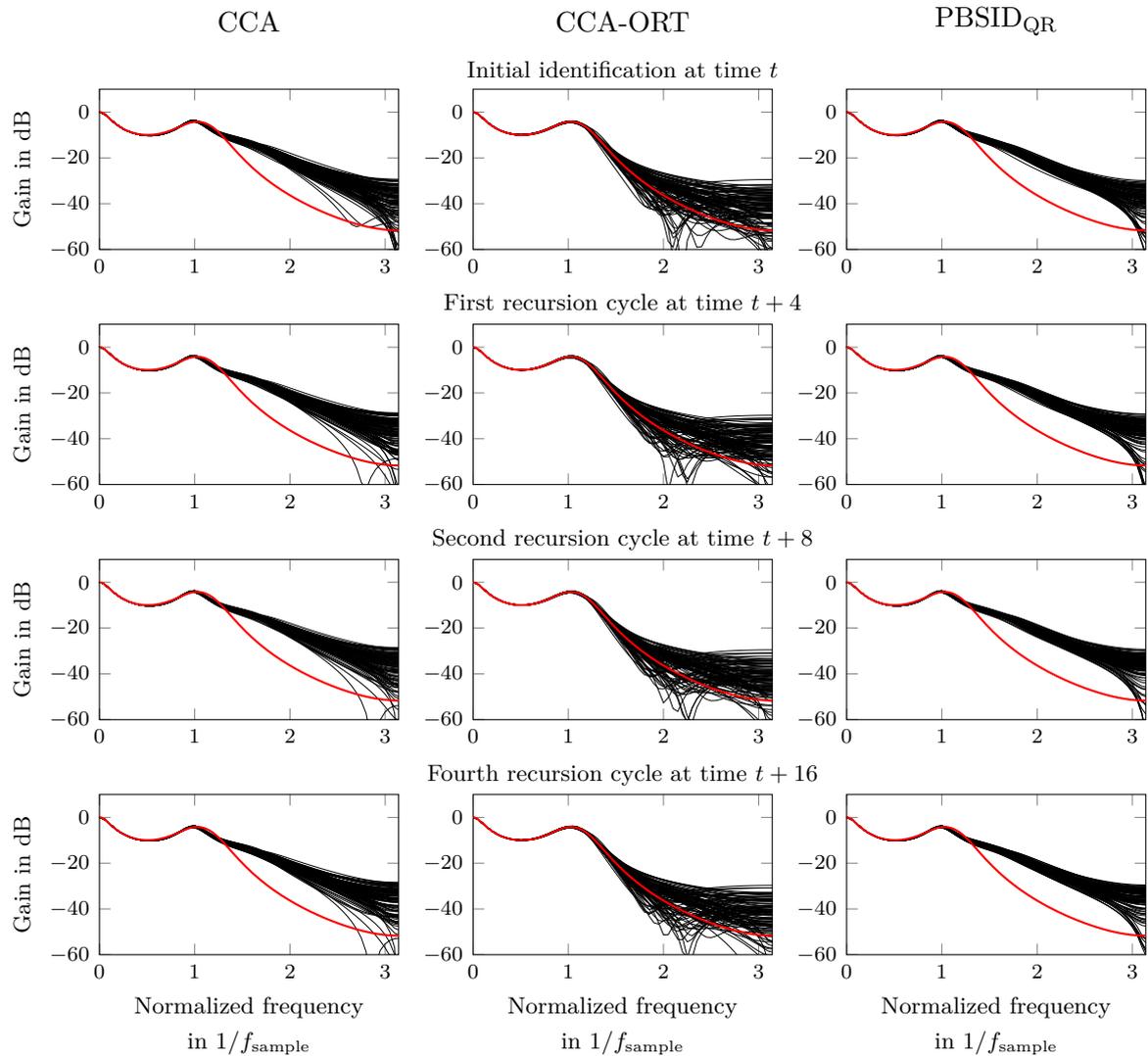
Figure 6.4: Recursive identification of the LTI Box-Jenkins system with CCA, CCA-ORT, and PBSID$_{QR}$ (from left to right): plots show the identification results of the initial identification and the first, second, and fourth recursion cycle (red: nominal value of $P(z)$, black: identified models)

6.4.2 Time-varying SISO system

In this subsection, two different examples are considered. The first is a general example illustrating the capabilities of tracking the changes of time-varying systems. The second provides a deeper look at the influence the change rate and/or the number of columns has on the results of the identification.

Standard identifications

Again, the two examples of Section 4.1.4 are considered with unchanged basic simulations and identification parameters. That is, the identifications of a time-varying ARMAX system
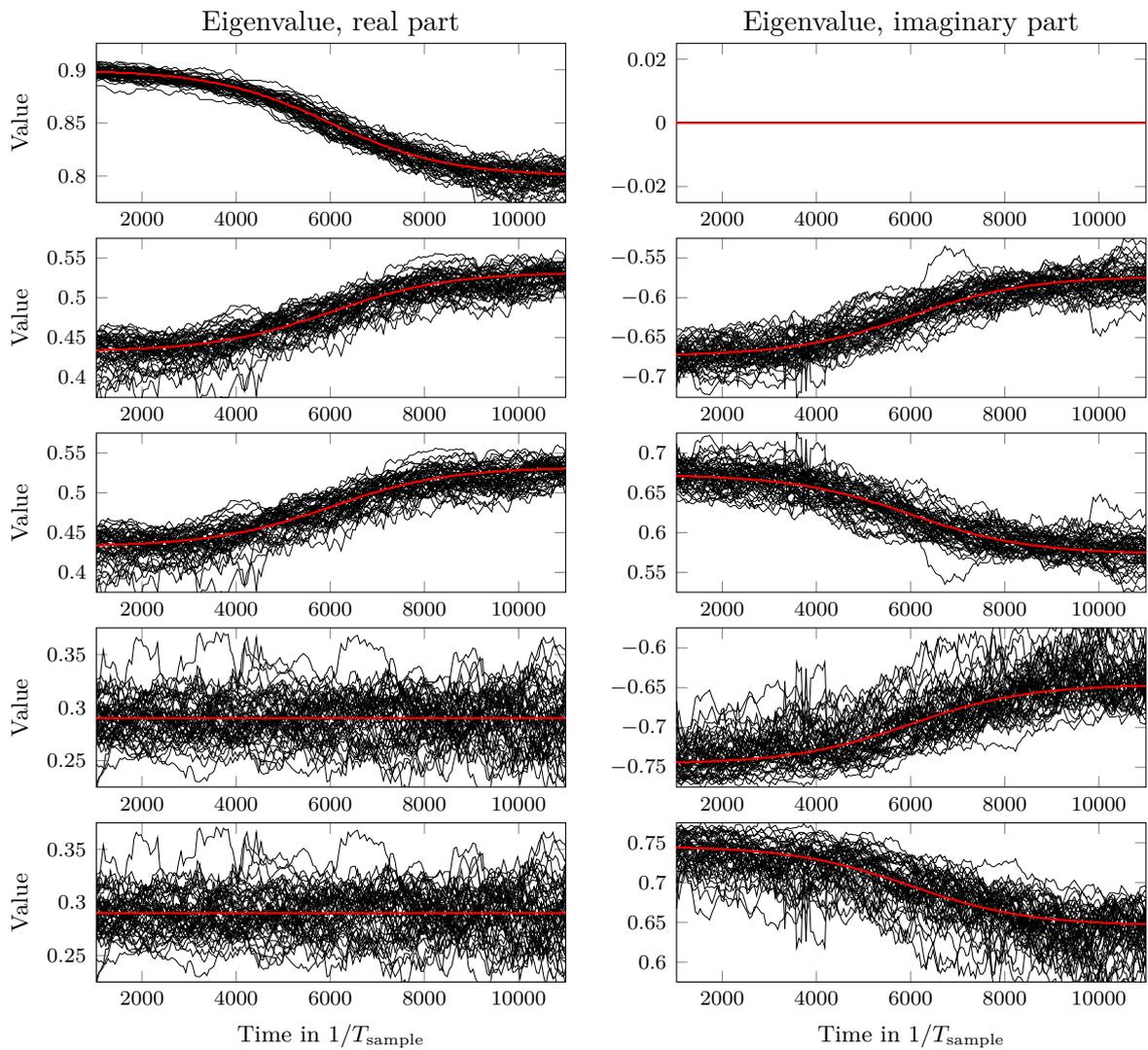
Figure 6.5: Tracking of eigenvalues z_1 to z_5 (from top to bottom) by the CCA method for the ARMAX example (red: nominal value of $P(z)$, black: value of identified models)

and a time-varying Box-Jenkins system are considered. The methods used are the recursive versions of the CCA method (state algorithm) and the CCA-ORT algorithm. The difference with respect to Section 4.1.4 is only in terms of the input and the number of simulations. For each of the 50 simulations, a new input signal is generated alongside with the change of the disturbance. The interval between each identification is 10 time steps and identifications are made over a total time interval of $[30, 12000]$, i.e., the re-identifications take place at
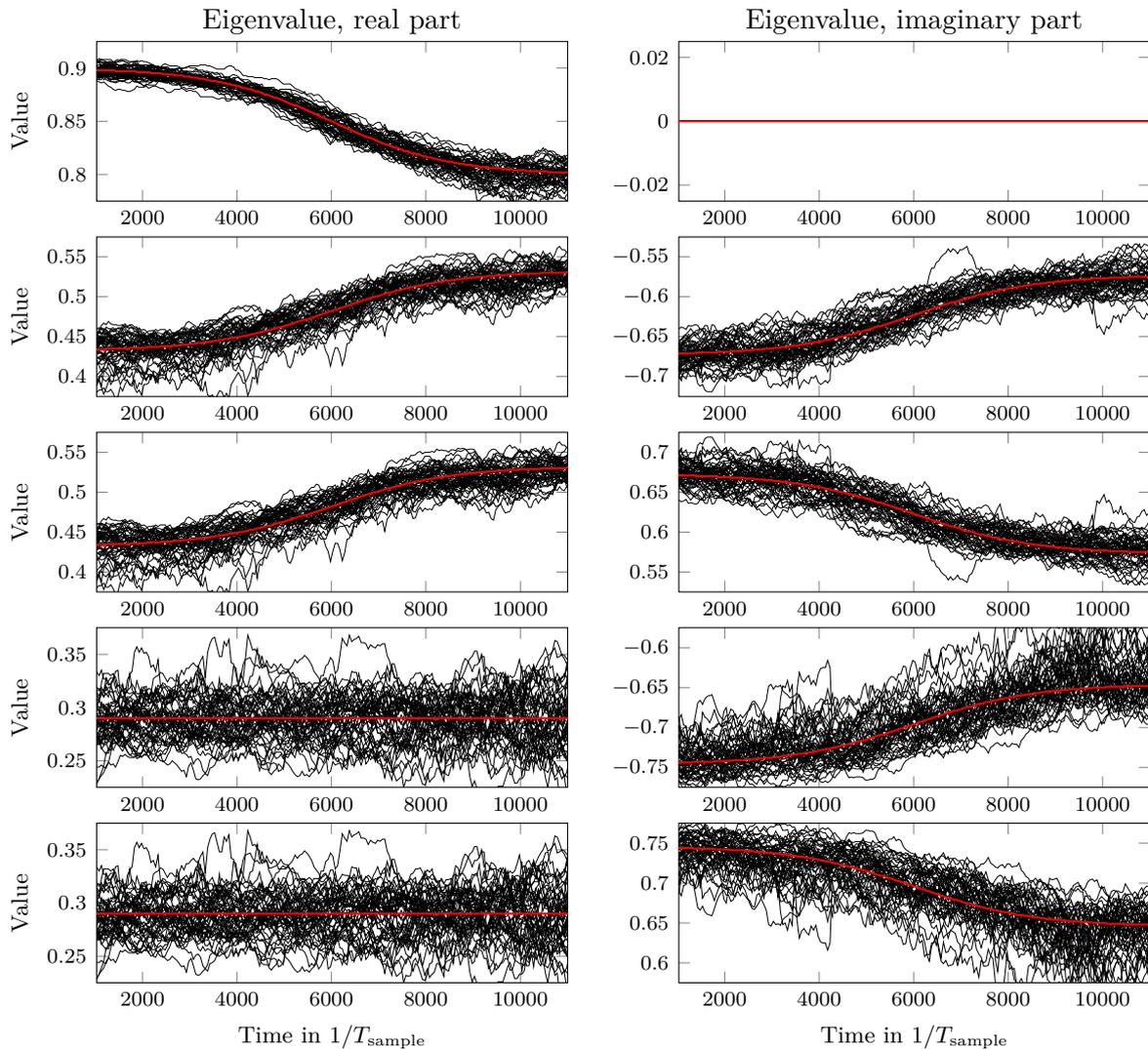
Figure 6.6: Tracking of eigenvalues z_1 to z_5 (from top to bottom) by the CCA-ORT algorithm for the ARMAX example (red: nominal value of $P(z)$, black: value of identified models)

40, 50, 60, $\cdots$, 12000, whereas the initial identification was made at 30. The time variation of the system was introduced by changing the eigenvalues (poles) according to

$$z_1(t) = 0.9000 - 0.1\frac{1}{1 + e^{\frac{6000-t}{1250}}} \ ,$$

$$z_{2/3}(t) = 0.4323 + 0.1\frac{1}{1 + e^{\frac{6000-t}{1250}}} \pm \left(0.6729 - 0.1\frac{1}{1 + e^{\frac{6000-t}{1250}}}\right) \ ,$$

$$z_{4/5}(t) = 0.2898 \pm \left(0.7458 - 0.1\frac{1}{1 + e^{\frac{6000-t}{1250}}}\right) \ .$$

The resulting progressions of the eigenvalues of the models are shown in Figure 6.5 and Figure 6.6 for the ARMAX system and in Figure 6.7 and Figure 6.8 for the Box-Jenkins system. To
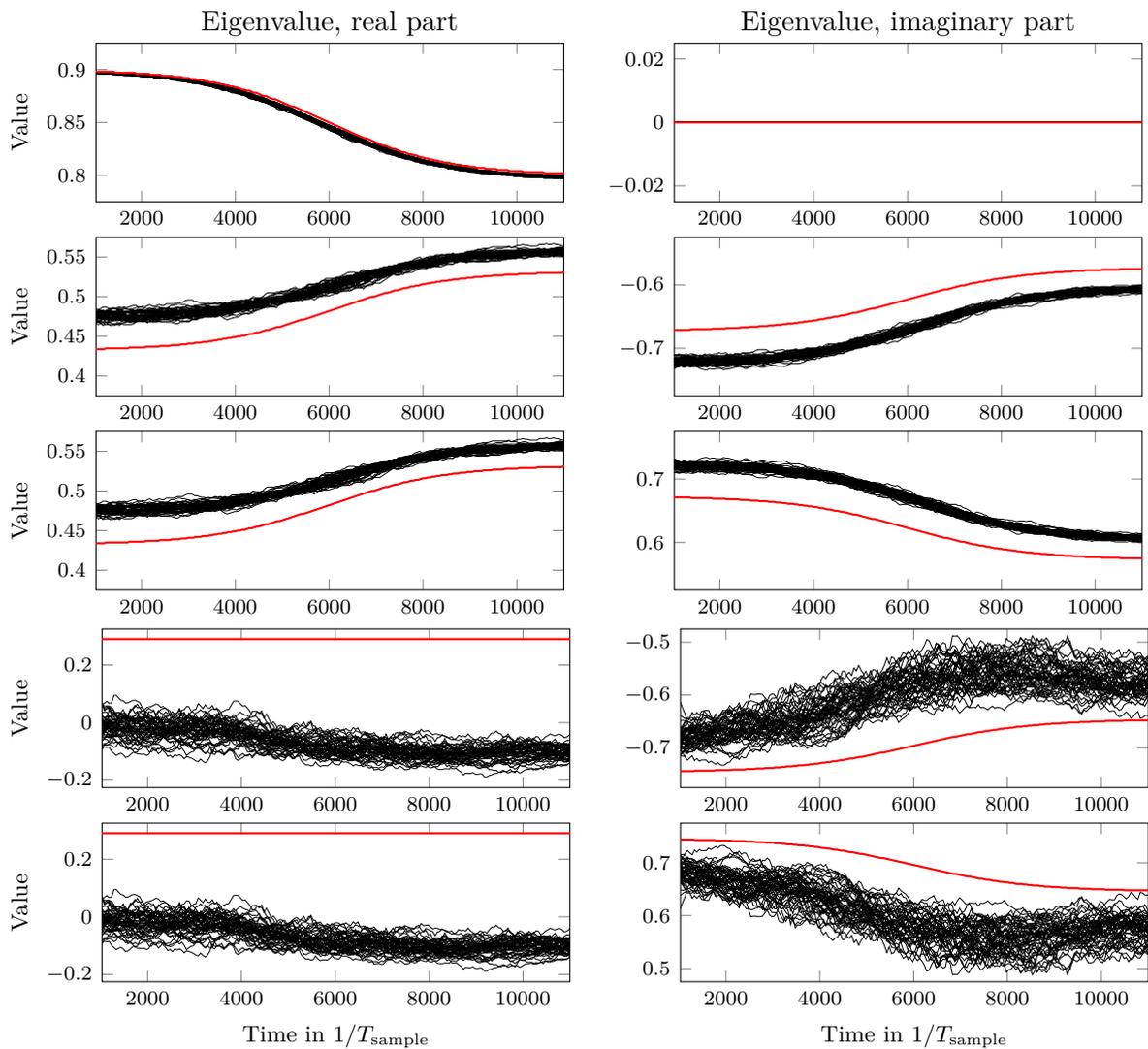
Figure 6.7: Tracking of eigenvalues z_1 to z_5 (from top to bottom) by the CCA method for the Box-Jenkins example (red: nominal value of $P(z)$, black: value of identified models)

align the data within the diagrams, the reference time of the tail matrices $y_N(t)$ and $u_N(t)$ is set to be in the middle of the tail matrices, i.e., the tail matrices are constructed as

$$y_N(t) = \left[y(t-1000) \cdots y(t-1) \ y(t) \ y(t+1) \cdots y(t+999) \right] .$$

Unless this adjustment is made the results of the identification would appear to be time shifted with respect to the actual progression of the system's eigenvalues. Nevertheless, this already reveals the first problem of the basic recursive scheme. A genuine tracking seems to be impossible because of the need for (with respect to the reference time t) future values. That is, an actual online tracking will always exhibit a time delay of $N/2$. This offset follows from an averaging of the covariances matrices over the interval $[t-N+1,t]$, which, for a constant rate of change, leads to the covariance of $t - \lfloor N/2 \rfloor$. This aspect will be further developed in Section 6.5.1. For the time being, the general behavior of the results is of interest. As shown by the diagrams, the general properties of the methods remain the same. Whereas the
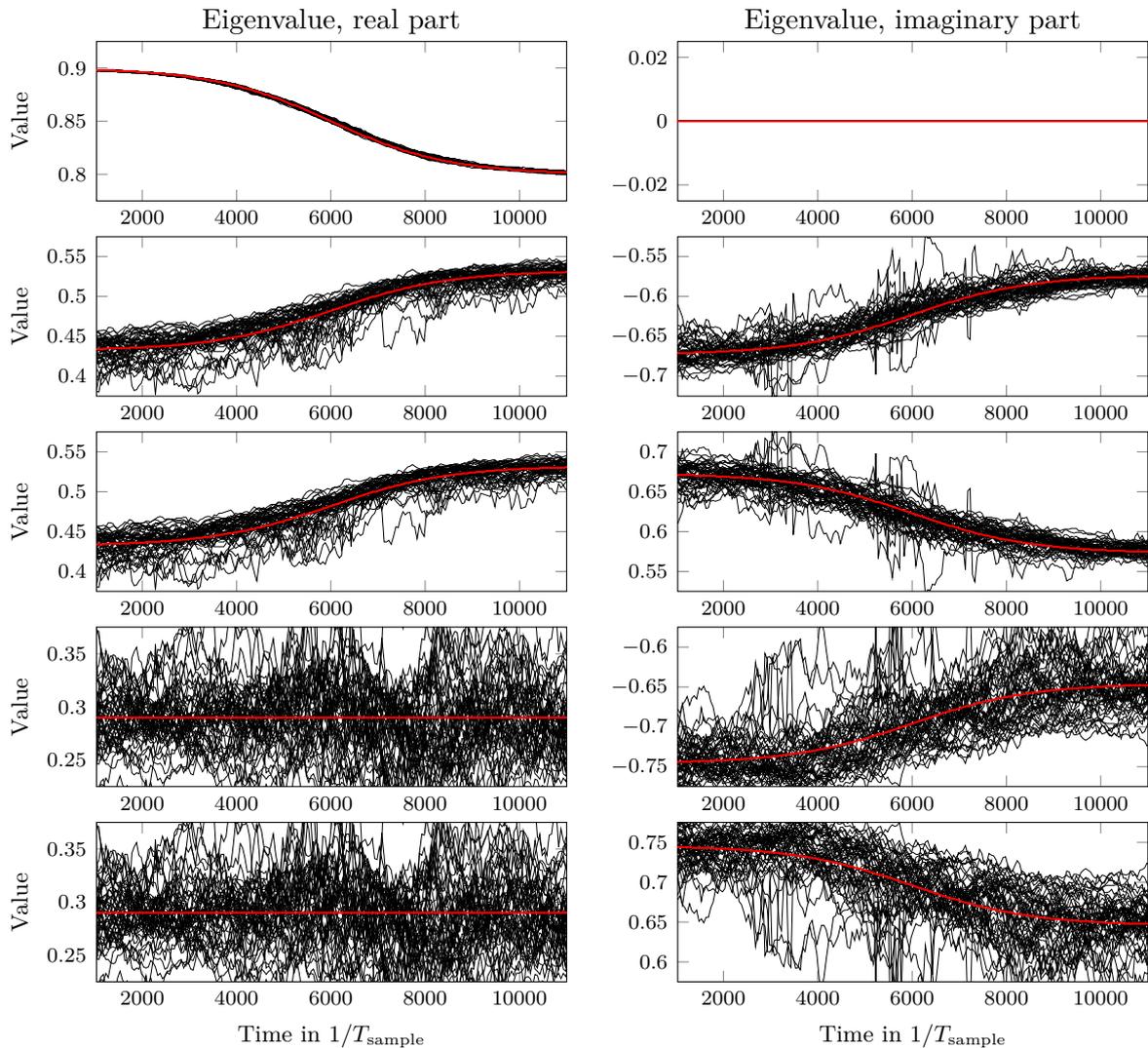
Figure 6.8: Tracking of eigenvalues z_1 to z_5 (from top to bottom) by the CCA-ORT algorithm for the Box-Jenkins example (red: nominal value of $P(z)$, black: value of identified models)

identification of the systems with joint dynamics is not a problem for either of the methods, the correct identification of a system in the presence of an arbitrarily colored noise is only achieved by the CCA-ORT algorithm. The results of the PBSID$_{\mathrm{QR}}$ algorithm are omitted as they are basically the same as for the CCA method.

Number of columns vs. change rate

In the wake of the remark regarding the influence of the tail matrices on the tracking capabilities of recursive subspace methods, the question regarding the limitations of this approach cannot be evaded. It should be noted that this is not a problem of the proposed approach but rather a general problem. Although recognized in the literature, this aspect is usually only superficially dealt with. In Lovera, Gustafsson, and Verhaegen (2000) and Mercère, Bako, and Lecuche (2008) this problem appears in terms of the calculation of sample covariance

matrices. The influence by past values is then suppressed by introducing a forgetting factor. In Kameyama, Ohsumi, et al. (2005) the influence of the past is suppressed by introducing a data window. That is, the sample covariance matrices are only calculated over a certain interval. In each iteration step the past data, which lies outside of the window, is subtracted from whereas the new data is added to the covariance estimates. However, the limitations of this approach and its cause of the limitations, which become apparent in the examples are not thoroughly addressed.

To illustrate the influence the length of the tail matrices and the change rate have on the results (and in particular to derive an approach to solve this), an identification of an undisturbed ARMAX system, i.e. an ARX system, with an initial transfer function of

$$P(z) = \frac{0.0275z^{-4} + 0.0551z^{-5}}{1 - 2.3443z^{-1} + 3.081z^{-2} - 2.5274z^{-3} + 1.2415z^{-4} - 0.3686z^{-5}} \ .$$

is considered. The initial identification is made with $k_p = k_f = 30$. The following identifications are made with a time increment of 10 and over the interval $[30, 5500]$. For each of 50 simulations, a different white-noise input is generated. In terms of the nominal identifications, i.e., the identification which the comparisons are made against, the column number N is 1000, whereas the change of the system is given by the change of the eigenvalues as

$$z_1(t) = 0.9000 - 0.1\frac{1}{1 + e^{\frac{2500-t}{350}}} \ ,$$

$$z_{2/3}(t) = 0.4323 + 0.1\frac{1}{1 + e^{\frac{2500-t}{350}}} \pm \left(0.6729 - 0.1\frac{1}{1 + e^{\frac{2500-t}{350}}}\right) \ ,$$

$$z_{4/5}(t) = 0.2898 \pm \left(0.7458 - 0.1\frac{1}{1 + e^{\frac{2500-t}{350}}}\right) \ .$$

For the following four comparative identifications, either the column number was set to 500 or 2000 or the change rate was accelerated according to

$$z_1(t) = 0.9000 - 0.1\frac{1}{1 + e^{\frac{2500-t}{175}}} \ ,$$

$$z_{2/3}(t) = 0.4323 + 0.1\frac{1}{1 + e^{\frac{2500-t}{175}}} \pm \left(0.6729 - 0.1\frac{1}{1 + e^{\frac{2500-t}{175}}}\right) \ ,$$

$$z_{4/5}(t) = 0.2898 \pm \left(0.7458 - 0.1\frac{1}{1 + e^{\frac{2500-t}{175}}}\right) \ .$$

or decelerated according to

$$z_1(t) = 0.9000 - 0.1\frac{1}{1 + e^{\frac{5000-t}{700}}} \ ,$$

$$z_{2/3}(t) = 0.4323 + 0.1\frac{1}{1 + e^{\frac{6000-t}{700}}} \pm \left(0.6729 - 0.1\frac{1}{1 + e^{\frac{5000-t}{700}}}\right) \ ,$$

$$z_{4/5}(t) = 0.2898 \pm \left(0.7458 - 0.1\frac{1}{1 + e^{\frac{5000-t}{700}}}\right) \ ,$$

with a respective adjustment of the identification interval to $[30, 11000]$ for the slower change rate.
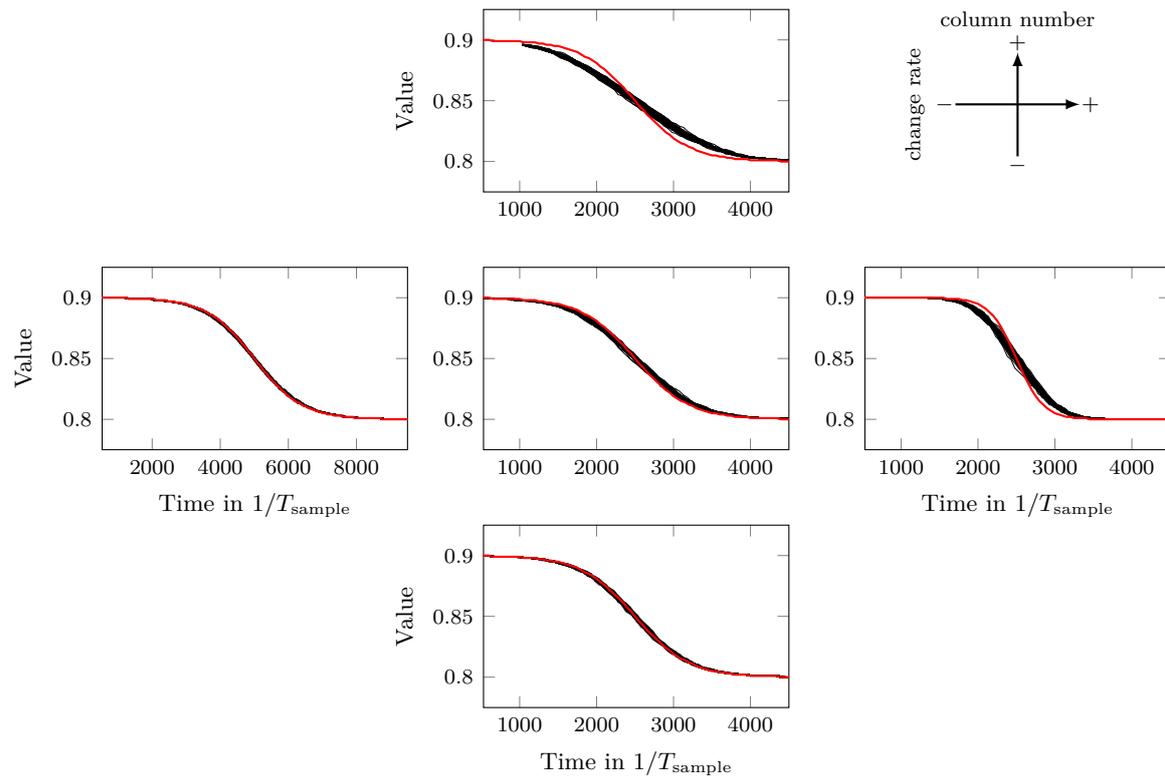
Figure 6.9: Comparison of the influence of the column number and the change rate of the system in terms of the real-valued eigenvalue z_1 of $P(z)$ (from left to right: increase of the change rate with otherwise fixed simulation and identification parameters, from bottom to top: increase of the column number with otherwise fixed simulation and identification parameters; red: nominal value of $P(z)$, black: identified models)

The resulting progressions of the real-valued eigenvalue z_1 (transition from 0.9 to 0.8) are compared in Figure 6.9. The reference times of the tail matrices are again aligned with nominal progression of the system's eigenvalues. The diagrams show the expected results. Whereas the accuracy of the tracking improves as either the column number or the change rate are reduced, the tracking becomes worse if both values are increased. These observations led to the remark in the introduction of this section regarding the yet to be solved problem of recursive subspace identification and the outline of an approach for solving this problem as presented in Section 6.5.1.

6.5 Practical aspects

Up to this point, the main focus has been on the theoretical aspect of the recursion. In this last section, two aspects which are more concerned with the implementation and application of the particular methods are highlighted. The first and foremost important aspect of making the methods actually usable deals with the influence of the column count of the tail matrices. This aspect already surfaced in someway in the previous examples. Here, the analysis and a possible remedy are discussed. The second aspect deals with the effects the re-identifications

have on the basis of the model. This results in an outline of an approach for keeping the basis of a model of a SISO system the same throughout the recursion cycles. An extension of the idea to MIMO systems is also mentioned.

6.5.1 Influence of the column number

As previously stated, the hindrance for tracking fast changes of a system stems from the tail matrices $y_N(t)$ or $u_N(t)$, which represent the stochastic variables $y(t)$ or $u(t)$. However, the underlying reasoning, which gives rise to these tail matrices, coincidentally depicts also an approach to circumvent this problem. Hence, after explaining the need for tail matrices, an approach for a possible solution is drawn. Some of the following considerations are also made in Bathelt, Söffker, and Jelali (2017, 2018).

In terms of the numerical implementation, the spaces $\mathcal{U}_t^-$, $\mathcal{Y}_t^-$, etc. are represented by the row spaces of block (row-wise) Hankel matrices. These block rows are in turn given by tail matrices of the type (here for the output y)

$$y_N(t) = \Big[y(t)\ y(t+1)\ \cdots\ y(t+N-1)\Big] \ , \tag{6.30}$$

where $y(t)$ are the values of a sample function of y. The idea to represent the stochastic variable $y(t)$ by $y_N(t)$ follows from the law of large numbers and the ergodic theorem (see, e.g., Doob, 1990, 1953; Rozanov and Feinstein, 1967). Basically, a tail matrix can be constructed from the results of repetitions of the random experiment $y(t)$. However, multiple evaluations of a stochastic processes at one point in time are practically impossible, as time can obviously not be stopped. The solution presents itself by the assumed stationarity of the processes. That is, all random variables within a stochastic process are the same (for subspace identification similarity in terms of second moments, i.e., wide-sense similarity). Hence, the necessary multiple evaluation at one point in time can be replaced by evaluations of consecutive random variables, i.e., by evaluations of $\cdots,\ y(t),\ y(t+1),\ y(t+2),\cdots$. This results in one possible sample function of y, which is then used for the construction of the tail matrix. Given these values, the tail matrices $\cdots, y_N(t),\ y_N(t+1),\ y_N(t+2),\ \cdots$ or its respective elements represent the stochastic process y through the sample covariance function

$$\hat{\Sigma}_{yy}(l) = \frac{1}{N} y_N(t+l) y_N^\mathrm{T}(t) = \frac{1}{N} \sum_{k=0}^{N-1} y(t+l+k) y^\mathrm{T}(t+k) \ , \tag{6.31}$$

which replaces the ensemble covariance function in terms of the numerical implementations, see Katayama (2005) and Lindquist and Picci (2015). For $N \to \infty$,

$$\Sigma_{yy}(l) = \hat{\Sigma}_{yy}(l) = \lim_{N\to\infty} \frac{1}{N} y_N(t+l) y_N^\mathrm{T}(t) \tag{6.32}$$

holds. In terms of the numerical implementation, the number N of the columns is dictated by both the signal-to-noise ratio and by the complexity of the process. Thus, the number of columns of the tail matrices and hence of the block Hankel matrices is usually larger ($\gtrsim 10$ times larger) than the lengths of the past and future horizons, i.e., the number of its block rows.

Considering the typical application of recursive identification means to consider time-varying systems. In such cases, the stationarity of y is however no longer given. Abandoning the
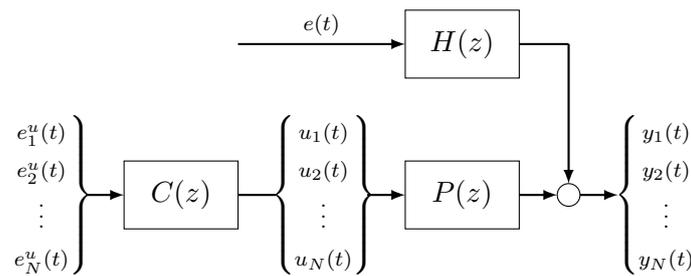
Figure 6.10: Construction of input signals consisting of multiple subprocesses; e_i are mutually orthogonal

assumption of stationarity has in turn major implications regarding on the elements of the tail matrices $\cdots, y_N(t), y_N(t+1), y_N(t+2), \cdots$. As the underlying stochastic process y is now allowed to change in terms of its stochastic properties, the single random variables $\cdots, y(t), y(t+1), y(t+2), \cdots$ do not exhibit the same stochastic properties. Thus, $\hat{\Sigma}_{yy}(l)$ in (6.31) is no longer an estimator of the $\Sigma_{yy}(l)$, as $\hat{\Sigma}_{yy}(l)$ essentially represents the mean of the covariances over the interval $[t, t + N - 1]$ (cf. ergodic theorem and central limiting theorem in Rozanov and Feinstein, 1967, pp. 156–161, pp. 190–198). This averaging effect has already been mentioned during the explanation of the examples and has led to the shift of the results. Technically speaking and respecting the data collection during on-line operation, this effect already renders a genuine tracking close to impossible as the tail matrices need to be constructed under on-line conditions as

$$ y_N(t) = \left[y(t - N + 1) \ y(t - N + 2) \ \cdots \ y(t) \right] , $$

which in turn results together with the averaging effect over the interval of $y_N(t)$ in a delayed tracking of the system's changes. This strict restriction was relaxed during the simulation by placing the reference time of the tail matrices in the middle of the interval of tail matrices. Similar considerations were also made in Kameyama, Ohsumi, et al. (2005). The other problem resulting from this is the direct influence of the averaging. Unless one assumes that the processes is close to stationarity over the $[t - N + 1, t]$ (called stationarity interval in Kameyama, Ohsumi, et al., 2005), results as shown by the top and right diagrams of Figure 6.9 follow, i.e., there will be deviations from the exact value even if the reference times of the tail matrices are aligned. As however the length of this stationarity interval is usually required to be much larger than the past and future horizons, the process is assumed to be slow enough that there is practically no merit of a recursive approach (except for run time benefits). That is, standard methods can be used without compromising the accuracy of the results. Due to the relation between the tail matrices and the sample covariances, this issue appears not only in terms of the algorithm of Kameyama, Ohsumi, et al. (2005) but also the in terms of the algorithms of Lovera, Gustafsson, and Verhaegen (2000) and Mercère, Bako, and Lecuche (2008).

The solution of these problems can be concluded from the idea of the tail matrices. Multiple momentary evaluations of the stochastic processes y and u at any time must be facilitated. This requirement can be achieved by constructing the input process u as a sum of multiple input subprocesses u_i, which are the same in a wide sense, i.e., in terms of their second moments. Under this condition, the resulting subprocesses y_i of the output are also the

same. Then, tail matrices can be constructed from the sample function of these subprocesses as

$$y_N(t) = \left[y_1(t) \; y_2(t) \; \cdots \; y_N(t) \right] , \tag{6.33}$$

or as

$$y_N(t) = \left[y_1(t) \; \cdots \; y_M(t) \; y_1(t+1) \; \cdots \; y_M(t+k-1) \; y_1(t+k) \; \cdots \; y_M(t+k) \right]$$

if the number of subprocesses is smaller than the number of columns. This idea for the generation of an input is illustrated in Figure 6.10. For example, the introduction of 10 subprocesses reduces the length of the interval of the tail matrices to one tenth of its original length. The reduction of the interval in turn allows change rates which are, with respect to the original tail matrix, 10 times faster than original. This is a direct consequence of the reduction of the averaging effects if the tail matrices are constructed as in (6.33). This averaging vanish completely if all elements stem from the same point in time and hence exhibit the same stochastic properties. These properties are governed by the system $P(z)$. To facilitate a separation of the output or input into the individual subprocesses y_i and u_i, these subprocesses need to be mutually orthogonal. This property can be in turn inherited from the driving white noise of the input process, as

$$E\{u_1(t)u_2^{\mathrm{T}}(t)\} = C(z)E\left\{ e_1^u(t)(e_2^u(t))^{\mathrm{T}} \right\} C^{\mathrm{T}}(z) . \tag{6.34}$$

Hence, if the driving white-noise processes e_i are mutually orthogonal, so are the individual subprocesses y_i and u_i. The separation of y and u into its subprocesses is then given as follows. Whereas the input separation simply follows with

$$u_i(t) = \hat{E}\left\{ u(t) \Big| \mathcal{E}_{t+1}^{i,-} \right\} , \tag{6.35}$$

the separation of the output process needs to take the disturbances, which, by assumption, are orthogonal to the e_i, into consideration. That is, first, the disturbances need to be determined by

$$y_s(t) = y(t) - \hat{E}\left\{ y(t) \left| \bigvee_{i=1}^{h} \mathcal{E}_{t+1}^{i,-} \right. \right\} , \tag{6.36}$$

where h is the number of subprocesses e_i. The disturbance $y_s(t)$ is then added to each subprocess when the separation is made by

$$y_i(t) = \hat{E}\left\{ y(t) \Big| \mathcal{E}_{t+1}^{i,-} \right\} + y_s(t) . \tag{6.37}$$

The core of this proposed scheme is based on the underlying relation between the second moments of a stochastic process and a system. These second moments are governed by the filter $C(z)$ and the system $P(z)$. Thus, the identifications only need these second moments of a stochastic processes to be the same – in general, the same holds also for classical identifications. The general stochastic properties of the stochastic process or the underlying subprocesses of the entries in the tail matrices (6.33) are thus irrelevant as long the second moments are equal. A proof of this idea would be based on the central limiting theorem, which is however beyond the scope of this thesis.
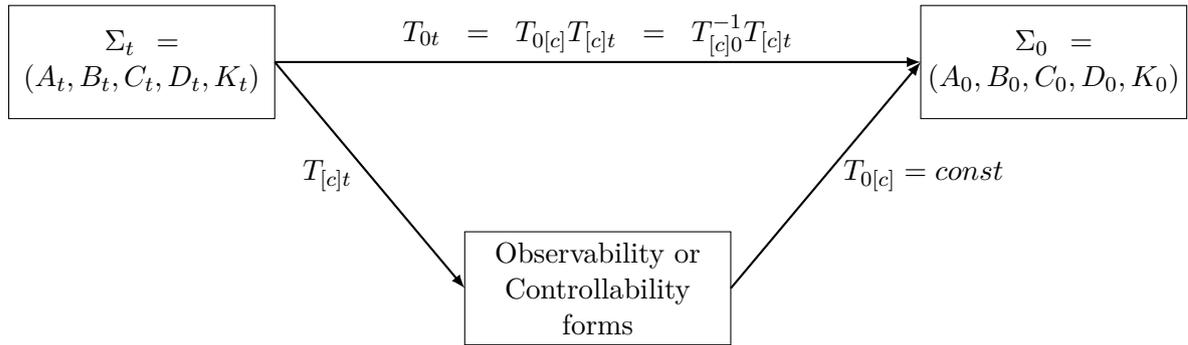
$$\Sigma_t = (A_t, B_t, C_t, D_t, K_t)$$

$$T_{0t} = T_{0[c]}T_{[c]t} = T_{[c]0}^{-1}T_{[c]t}$$

$$\Sigma_0 = (A_0, B_0, C_0, D_0, K_0)$$

$$T_{[c]t}$$

$$T_{0[c]} = const$$

Observability or Controllability forms

Figure 6.11: Preservation of the basis of model by similarity transformations based on canonical forms; the subscript c is to be replaced by either $\mathcal{C}$ or $\mathcal{O}$, indicating the controllability or observability forms

6.5.2 Preservation of basis

The eponymous operation of subspace methods is the extraction of the predictor space. That does actually not include the extraction of a certain basis of this predictor space. In fact, different methods might determine different bases for the same predictor space. In terms of the recursive methodology, this leads to the nuisance of a possible basis change with every re-identification. Hence, a method is needed to keep the basis of the model either to a predefined basis or to the basis of the initially identified model.

An approach to this problem includes similarity transformations between two systems $\Sigma_1 = (A_1, B_1, C_1, D_1, K_1)$ and $\Sigma_2 = (A_2, B_2, C_2, D_2, K_2)$ which are given by

$$A_2 = T_{21}^{-1}A_1T_{21}, \quad B_2 = T_{21}^{-1}B_1, \quad C_2 = C_1T_{21}, \quad D_2 = D_1, \quad K_2 = T_{21}^{-1}K_1 .$$

As on the one hand, the transformation matrix T_{21} is determined (Kailath, 1980) as

$$T_{21} = \mathcal{C}_1\mathcal{C}_2^{-1} = \mathcal{O}_1^{-1}\mathcal{O}_2 ,$$

but on the other hand the matrices of the realization, which is to be transformed into, are unknown, a procedure using an intermediate step via canonical forms is introduced. In terms of SISO systems and using the controllability or observability forms, the transformation matrices reduce to

$$T_{\mathcal{C}1} = \mathcal{C}_1, \qquad \text{or} \qquad T_{\mathcal{O}1} = \mathcal{O}_1^{-1},$$

as the respective controllability and observability matrices are the unity matrices, see Kailath (1980, p. 128). Based thereon, the procedure of the transformation of the identified models into realizations having the same basis is as follows:

1. Determination of the transformation matrix $T_{0[c]}$, where the subscript c is to be replaced by either $\mathcal{C}$ or $\mathcal{O}$, indicating a transformation from the controllability or observability forms to a basis realization Σ_0, i.e., the model of the initial identification cycle. This transformation matrix is the inverse of $T_{[c]0}$, the transformation from Σ_0 to the respective canonical form.

2. Determination of the transformation matrix $T_{[c]t}$ for the transformation from the model, identified in in each recursion cycle, to the controllability or observability forms

3. Construction of the global transformation matrix T_{0t} for the transformation from the identified model to the basis of the basis realization as $T_{0[c]}T_{[c]t}$

4. Transformation of the identified model

This procedure is illustrated in Figure 6.11. The artifice of the transformation procedure lies in the fact that the transformation matrix $T_{0[c]}$ remains the same independent of the numerical values of the canonical forms. The changes of the canonical forms or the realizations in general stem from the changes of the system dynamics but not from changes of the state basis, cf. LPV models. Hence, once the transformation matrix $T_{0[c]}$ is determined by inverting the transformation matrix from Σ_0 to the controllability or observability forms, it can be used in every identification cycle. The determination of the missing transformation matrix T_{ct} is straightforward as only the realization of the model of the current recursion cycle is needed. It should be noted that this procedure is however restricted to SISO models.

The problem appearing in terms of the MIMO system is the loss of the uniqueness of the controllability or observability forms (Kailath, 1980, p. 426), which is however crucial for this procedure. Without uniqueness, the transformation $T_{[c]t}$ might give a canonical realization which differs from the realization, which $T_{0[c]}$ is based on. A solution to this problem is given by the approach outlined in Mercère and Bako (2011) or the overall unique echelon form, see Hannan and Deistler (2012).

A general issue that should be kept in mind is the question regarding the correct extraction of the order. The idea outlined in this section is based on the tacit assumption of a known and fixed order of the system, as it would for example be the case with LPV systems. An overall change of the basis can then be ruled out, as the order of the system is constant and exactly known. In all other cases, a possible change of the predictor space as a whole might be possible.

6.6 Summary

In this chapter, the work with respect to the theoretical contribution of the thesis is discussed. The approach to recursive subspace identification is based on the fundamental methodological relations of subspace identification instead of the compression of some data matrices as it has been done before. Hence, the approach outlined here is the first of its kind. The theoretical foundation for the approach to recursive subspace identification is derived based on the coordinate-free framework. In this framework, the space of the state of a system is given by the (oblique) Markovian splitting subspace of some past and future data spaces. It is shown that these Markovian splitting subspaces can be recursively extracted. This leads to an inherent recursive scheme for the identification by subspace methods. That is, given a Markovian splitting subspace of any point in the past and the intermediate data, the current Markovian splitting subspace can be determined. As a minimal Markovian splitting subspace or minimal predictor space provides the necessary past information while being the smallest possible subspace of the past to do so, the minimal predictor space gives also the best possible data compression. That is, there is no redundant data or information propagated to the future. As this scheme is furthermore theoretically motivated, the resulting methods/algorithms are not purpose-build but rather simple extensions of existing methods. The numerical examples verified the functioning of the approach. Further studies show that the tracking capability and hence quality of the results are limited by the length of the interval needed to construct the tail matrices, which represent the stochastic variables, i.e., the longer the tail matrices

are, the slower the change rate of the system has to be to retain an accurate tracking of the system. If this is not taken into account, this problem manifests itself in terms of a time shift or a blurring of the changes of the system. Although this problem also concerns the methods discussed in the literature, it has not sufficiently been addressed before. In terms of the approach proposed here, this emerging problem is given a possible solution by an input scheme, which facilitates the construction of tail matrices based on the data of reduced intervals or even of a single point in time.

Recapitulation of main results

- The future minimal (oblique) Markovian splitting subspaces are contained in the joint spaces spanned by the current minimal (oblique) Markovian splitting subspaces and the respective intermediate data spaces (stochastic case and joint stochastic-deterministic case) as

$$\mathcal{X}_{t+k}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \,, \qquad \mathcal{X}_{t+k}^{+/-} \subset \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \,.$$

- The determination of minimal (oblique) Markovian splitting subspaces follows as (stochastic case and general joint stochastic-deterministic case)

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}\left\{\mathcal{Y}_{t+k}^+ \,\middle|\, \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^-\right\} \,,$$
$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{F}_t^+}\left\{\mathcal{Y}_{t+k}^+ \,\middle|\, \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^-\right\} \,.$$

- From the derivation of the minimal oblique Markovian splitting subspace for the joint stochastic-deterministic case, the formulation for the open-loop case is achieved by reducing $\mathcal{F}_t^+$ to $\mathcal{U}_t^+$, yielding

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+}\left\{\mathcal{Y}_{t+k}^+ \,\middle|\, \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^-\right\} \,.$$

  This is further extended to the identification of the deterministic subsystem by a reduction to the deterministic component $\mathcal{X}_t^{d,+/-} \vee \mathcal{U}_{[t,t+k)}^-$. This gives

$$\mathcal{X}_{t+k}^{d,+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+}\left\{\mathcal{Y}_{t+k}^+ \,\middle|\, \mathcal{X}_t^{d,+/-} \vee \mathcal{U}_{[t,t+k)}^-\right\} \,.$$

  Similar reasoning leads to a version for the predictor-based approach of the PBSID method (which is supposed to also work under closed-loop conditions) with

$$\mathcal{X}_{t+k}^{+/-} = \bigvee_{h=0}^{n-1} \mathcal{X}_{t+k}^h \,,$$

  where

$$\mathcal{X}_{t+k}^h = \hat{E}_{\|\mathcal{P}_{[t+k,t+k+h)}^+}\left\{\mathcal{Y}_{t+k+h} \,\middle|\, \mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k)}^-\right\} \,.$$

- The numerical implementation of the recursive methodology requires only minor changes to existing methods. In terms of the basic QR decompositions, the numerical data matrices of the past, Y_t^- (stochastic case) or P_t^- (joint stochastic-deterministic case), are replaced by $\left[x_N^T(t_0) \; (Y_{[t_0,t)}^-)^T\right]^T$ or $\left[x_N^T(t_0) \; (P_{[t_0,t)}^-)^T\right]^T$. The remainder of the met-

hods remains the same, except for, where necessary, an additional determination of the current state estimate $\hat{x}_N(t)$. That is, the recursive approach can be embedded into existing method without changing the actual identification algorithm.

- The influence of the interval which the data of the tail matrices is collected over, is reduced by constructing an input consisting of several mutually orthogonal subprocess. By constructing the tail matrices based on the sample functions of these subprocess, the interval length of the tail matrices is reduced or avoided altogether. This allows for the identification of systems with higher change rates.

7 Conclusions and continuative work

This last chapter concludes the thesis. In Section 7.1, the achieved results are summarized and assessed in terms of the goals set in Chapter 3. Ideas and recommendations for future work are given in Section 7.2.

7.1 Summary and conclusions

In this thesis, an approach to recursive subspace identification is proposed. This is mainly motivated by the pursuit of simplifying the identification of complex industrial plants, so that LTI models can be used instead of LPV models. These LTI models are then supposed to be adapted during a change of the operating point. In this context, not only a recursive approach is needed but also methods or algorithms suitable for the industrial environment. That is, methods or algorithms are needed that facilitate the recursive use in open-loop and closed-loop settings, while also be able to suppress disturbances and yield sufficiently accurate results in an industrial environment.

Open-loop methods

The basic idea of embedding the ORT approach into the calculation procedure of the CCA method is based on the fact that the determination of the deterministic predictor space can be achieved by

$$\mathcal{X}_t^{d,+/-} = \hat{E}_{||\mathcal{U}_t^+} \{ \mathcal{Y}_t^+ | \mathcal{U}_t^- \} \, .$$

This results from merging the two-step procedure of calculating first the deterministic component of the output and then the respective predictor space of this component. Based on the above oblique projection, the state-regression algorithm of the CCA method is modified to yield the model of the deterministic part of the system. This algorithm is hence called CCA-ORT. Comparing the results of this algorithm in terms of academic examples with the CCA method, N4SID method, PO/PI-MOESP algorithms, and the original MOESP-based ORT algorithm, the proposed CCA-ORT algorithm yields the same results or in comparison with the original MOESP-based ORT algorithm slightly better results. However, comparing the methods in terms of the more realistic benchmark example of the Tennessee Eastman Process, the CCA-based methods/algorithms yield the overall best results. This becomes even more pronounced once the outputs of the system are subjected to disturbances. Although the results of the CCA-based ORT algorithm starts to slightly loose its accuracy once the signal-to-noise ratio increases (SNR > 50), The results remain however the best, as the loss of accuracy is not as severe as it is the case with other methods/algorithms. Hence, the proposed algorithm outperforms the other methods/algorithms and satisfies the set goal.

Closed-loop methods

The candidate method for the identification in a closed-loop setting is the PBSID method, which is categorized as a direct approach method, i.e., neither the knowledge of the controller transfer function (indirect approach) nor an entire identification of the closed-loop system (joint input-output approach) are needed. As the algorithmic core of this method is a VARX estimation, the major issue is first and foremost concerned with an algorithm for the estimation which avoids any pre-estimation of a VARX model. In this context, an algorithm directly implementing the theoretical basis of the PBSID method is proposed. Instead of using a least-squares approach or a VARX estimation, the calculation of the predictor space is implemented according to

$$\mathcal{X}_t^{+/-} = \bigvee_{h=0}^{k_f-1} \mathcal{X}_t^h \, ,$$

where

$$\mathcal{X}_t^h = \hat{E}_{\|\mathcal{P}_{[t,t+h)}} \{ y(t+h) \mid \mathcal{P}_t^- \} \, .$$

This is achieved by rearranging the data in the numerical data matrix of the QR decomposition, which in turn facilitates the implementation of the projections of the predictor subspaces $\mathcal{X}_t^h$, $h = 0, \cdots, k_f - 1$ as given in the above equation. This algorithm is hence called PBSID$_{QR}$. However, problems in terms of either the numerical examples or the implementation of the methods – the QR decomposition-based PBSID method and benchmark methods – forbid a thorough comparison of the accuracy of the methods/algorithms. None of the methods/algorithms is able to achieve the set goal of identifying ARX, ARMAX or Box-Jenkins systems in a closed-loop setting. These problems appear although the benchmark methods are implemented as given in the literature. The cause of this problem needs further investigation, as it is not clear if it is only a parametrization error (the exact implementations and complete parametrization of the examples are not disclosed in the literature) or something more fundamental. Identifications of systems subjected to white-noise disturbances (output-error systems) in a closed-loop setting and open-loop identifications (the direct approach methods can be used in both settings) produce suitable results. Here, both the VARX-based and the QR decomposition-based PBSID algorithms are the best, with a slight advantage for the VARX-based algorithm. As furthermore, the fundamental assumptions of the ORT approach forbid a direct embedding of this approach into closed-loop methods, only the main goal, i.e., deriving a method which allows the recursive use, was achieved. Thus, the adaption of the algorithm for the use within an industrial environment, which would require a disturbance suppression, needs further work. In terms of Tennessee Eastman Process study, the PBSID algorithms were the second best algorithms, only surpassed by the CCA method and the CCA-ORT algorithm.

Recursive subspace identification

The goal of deriving a methodological basis for recursive subspace identification is achieved. It is given in terms of the modeling of stochastic systems by

$$\mathcal{X}_{t+k}^{+/-} = \hat{E} \left\{ \mathcal{Y}_{t+k}^+ \, \middle| \, \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \right\} \, ,$$

and in terms of the modeling of joint stochastic-deterministic systems by

$$\mathcal{X}_{t+k}^{+/-} = \hat{E}_{\|\mathcal{F}_t^+} \left\{ \mathcal{Y}_{t+k}^+ \, \middle| \, \mathcal{X}_t^{+/-} \vee \mathcal{Y}_{[t,t+k)}^- \vee \mathcal{U}_{[t,t+k)}^- \right\} \, .$$

The recursion is given by the calculation of a future minimal predictor space based on a past minimal predictor space, which are both equal to the model of a system (Akaike, 1974, p. 669, Lindquist and Picci, 1996b). The calculations for the deterministic minimal predictor space and for the theoretical basis of the PBSID method follow as corollaries with

$$\mathcal{X}_{t+k}^{d,+/-} = \hat{E}_{\|\mathcal{U}_{t+k}^+} \left\{ \mathcal{Y}_{t+k}^+ \left| \mathcal{X}_t^{d,+/-} \vee \mathcal{U}_{[t,t+k)}^- \right. \right\}$$

and

$$\mathcal{X}_{t+k}^{+/-} = \bigvee_{h=0}^{n-1} \mathcal{X}_{t+k}^h \,,$$

where

$$\mathcal{X}_{t+k}^h = \hat{E}_{\|\mathcal{P}_{[t+k,t+k+h)}^+} \left\{ \mathcal{Y}_{t+k+h} \left| \mathcal{X}_t^{+/-} \vee \mathcal{P}_{[t,t+k)}^- \right. \right\} \,.$$

In terms of the numerical implementation, the approach is however limited by the representation of stochastic values by tail matrices. As the covariance matrices, which are needed for the calculation of the projections, are based on these tail matrices, the change rate of a time-varying system needs to be slow, so that the ergodic theorem still holds to some degree and the sample covariance matrices are similar to the ensemble covariance matrices. Such a limitation of the change rate is however theoretically not needed. Thus, the numerical results of the identification – but not the underlying derivation of the approach – are similar to the ones of existing algorithms. A possible solution of this problem is discussed in terms of an input scheme, which can increase the permissible change rate. The open task is mainly the thorough proof of the idea and the subsequent implementation. It should be kept in mind that the correctness of the recursive approach remains unaltered even if the application of this approach to time-varying systems means a violation of the assumed stationarity, which the derivation is based on.

7.2 Continuative work

The following list contains recommendations for continuative work, which results from the work covered by this thesis. The order is with respect to the importance of the individual points.

Derivation and implementation of the input scheme

As mentioned in the conclusions, the most important next step is the derivation and implementation of the input scheme. The basic approach is already given by the outline in Section 6.5.1. In this context, the first step is to thoroughly prove the discussed idea, which states that the tail matrices, which are constructed from the sample functions of different orthogonal subprocesses, yield the same covariance estimation as the tail matrices of one sample function. This basically includes an approach via the central limiting theorem, see Doob (1990, 1953), Katayama (2005, pp. 79–81), and in particular the discussions in chapter 11 of Rozanov and Feinstein (1967). The second and more involved step is the implementation of the projections for the decomposition of y. Here, the approach needs to take into consideration that a projection is basically a linear least-squares estimation and that the coefficients associated with past random variables are not influenced by the future changes of the system. Only coefficient associated with the current and future random variables carry

the information regarding the present and future changes of the system. The application of this input scheme is not limited to recursive algorithms. It can also be used for standard algorithms.

Fixed state basis

The idea regarding the fixation of the state basis for MIMO systems needs to be further investigated. If it is true that such can be achieve similar to the SISO case by using the echolon form (see Hannan and Deistler, 2012, pp. 55–61), the question regarding the direct inclusion of this approach into the recursive scheme appears. The determination of the echolon form is based on the Kronecker indices and the resulting selection of certain parts of the Hankel matrix

$$H = \begin{bmatrix} G_1 & G_2 & G_3 & G_4 & \cdots \\ G_2 & G_3 & G_4 & G_5 & \cdots \\ G_3 & G_4 & G_5 & G_6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where its entries are given by the impulse response coefficients (for deterministic realizations and joint stochastic-deterministic realizations)

$$G_i = CA^{i-1}B , \qquad \text{or} \qquad G_i = CA^{i-1}\begin{bmatrix} B & K \end{bmatrix} .$$

As these Hankel matrices appear within the oblique projections yielding the predictor space, the question is whether the basis can be fixed by directly embedding the selection procedure into the calculations of these projections.

Closed-loop problems

The most pressing problem in this context is regarding the substandard results of all closed-loop methods as witnessed in the examples. It is not clear whether the issue might be an implementation related problem or if the chosen parameters, i.e., horizon lengths, have been outright wrong – something that cannot be ruled out, as such has not been disclosed in the respective literature. Further investigations and identification tests are needed, to get a deeper understanding.

In terms of improvements of the proposed PBSID_{QR} algorithm, the inclusion of a canonical correlation analysis approach, as discussed in Chiuso (2006) or done in Jansson (2003), might be a worthwhile endeavor. Another idea is with respect to the QR decomposition. As the data structure basically permits some kind of back-propagation of previous results, cf. Qin and Ljung (2003b), it might be possible to actually achieve a complete consistent estimation, i.e., enforcing the Toeplitz structure of $L_{\mathcal{P}_t^+}$, see page 92 and preceding derivations. See also Qin, Lin, and Ljung (2005) for a similar discussion in terms of PARSIM-E.

Closed-loop ORT

Although an approach for the orthogonal decomposition in a closed-loop setting is given by the joint input-output methods outlined in Katayama (2005) and Katayama and Tanaka

(2007), the question is whether it is possible to find also a way for direct-approach methods. Here, the main issue is regarding the necessity of the freedom of feedback needed for the ORT approach. If this assumption can be relaxed, it might be possible to find a similar approach for the closed-loop setting. In this context, the starting point is the fundamental work related to feedback in Anderson and Gevers (1982), Caines and Chan (1975, 1976), and Gevers and Anderson (1981) as well as related to feedback-free processes in Caines (1976) and Gevers and Anderson (1982).

Inclusion into process monitoring or control performance monitoring algorithms

The long-term objective[1] is to derive methods for process monitoring or control performance monitoring (see, e.g., Ding, 2013; Jelali, 2013) based on recursive subspace identification. Hence, once the previous issues are addressed, this idea might be resumed and algorithms for both fields should be derived. In particular, the process monitoring schemes outlined in Ding (2013) present a basis for such, as its algorithmic basis is intriguingly similar to subspace identification. In this context, the calculation of an initial state for the recursion by an observer reduces also the necessity of the initial identification.

Number of invariant parameters in subspace identification and order estimation

Although this is an important aspect for the implementation of order estimation schemes, the publications referring to this question are rather rare. Whereas in Hannan and Deistler (2012) the number of invariant parameters of a state-space system is specified in Theorem 2.5.3 and later on p. 69 with (assuming no feed-through)

$$d(n) = n(m + p) + np \,,$$

it is concluded in Candy, Bullock, and Warren (1979) that the number of invariant parameters of a stochastic system is given with

$$d(n) = 2np + 1/2p(p + 1) \,,$$

which takes the invariant parameters in A and C as well as in Q, R and S into account. From the standpoint of the Kalman filter representation of (2.4) and the explanations of Hannan and Deistler (2012), this is the number of invariant parameters of A, K, (C), and Σ_{ee}. As hence the specifications of the number of invariant parameters vary across the papers without one common reference, it would be a useful to find a common specification of the number of invariant parameters. Combining the explanations of Hannan and Deistler (2012) and Candy, Bullock, and Warren (1979), the initial idea would be

$$d(n) = np + n(m + p) + mp + 1/2p(p + 1) \,,$$

which takes the invariant parameters of the system matrices of (2.4) and the number of invariant parameters of the as well estimated Σ_{ee} into account. Considering furthermore the identification of the deterministic subsystem, the number of invariant parameters should be

$$d(n) = np + nm + mp \,.$$

---

[1]Or rather the actual initial idea of this thesis, which eventually turned out to be way to ambitious to be reached within the period of time.

This last equation is however restricted to systems being completely reachable by the input.

Although the above equations might be correct, the whole issue needs a thorough treatment and derivation, as those equations are rather results of assumptions than derivations. These assumptions follow from the combination of the mentioned references. Such a derivation and discussion should however be well-recieved as a combined presentation is missing and the parameter number is mostly written down without further discussion to why it is chosen in this particular way. The most ambiguous case concerns the inclusion of the innovation matrix. The question is whether it should be included into the parameter number or not.

Furthermore, the problem of order estimation might need additional treatment. Even if the methods of Bauer (2001) and Fujikoshi and Veitch (1979) produce good results in a linear setting, they tend to fail if the system in question becomes nonlinear or the estimation problem is handled in a rather straight-forward way as it was done in the Tennessee Eastman Process example (the horizon lengths were simply chosen based on a "the longer the better" judgment). In terms of the methods proposed in Bauer (2001) and considering linear systems, it should be further noted that the results depend heavily on the right choice of $C(T)$ or even the way the identification method/algorithm works. In the case of N4SID, the results have been satisfying only after the algorithm was adjusted to use a CVA weighting scheme (see van Overschee and De Moor, 1996, pp. 80–81, p. 114).

Implemenational aspects of open-loop methods

Lastly, an observation, which has been made during the Tennessee Eastman Process study, should be pointed out. Although the CCA-based algorithms performed the best in terms of a general assessment by the R^2 measure or the FPE measure, the N4SID algorithms yielded (on average) the best identification of the transient phase of the step responses (not shown by diagrams). The CCA-based algorithms are however superior in terms of the steady state estimation and percentage of usable models. When applied to the identification data, the N4SID algorithms partially tend to determine unstable models, which are clearly wrong as the system itself is stable. The CCA-based algorithms (and algorithms of the PBSID method) do not show this behavior. This determination of unstable models is reflected by outliers of the N4SID algorithms in the respective boxplots. Hence, the question is if it is possible to combine the good approximation of the transient phase of the N4SID methods with the good identification of the steady state and the reliability achieved by the CCA-based algorithms.

Bibliography

Akaike, H. (1974). "Stochastic Theory of Minimal Realization". In: *IEEE Transactions on Automatic Control* 19.6, pp. 667–674.

Akaike, H. (1976). "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion". In: *System Identification*. Ed. by R. K. Mehra and D. G. Lainiotis. Vol. 126. Mathematics in science and engineering. New York: Academic Press, pp. 27–96.

Akçay, H. (2011). "Frequency Domain Subspace-Based Identification of Discrete-Time Power Spectra from Uniformly Spaced Measurements". In: *Automatica* 47.2, pp. 363–367.

Anderson, B. D. O. and Gevers, M. R. (1982). "Identifiability of Linear Stochastic Systems Operating Under Linear Feedback". In: *Automatica* 18.2, pp. 195–213.

Åström, K. J. and Eykhoff, P. (1971). "System identification - A survey". In: *Automatica* 7.2, pp. 123–162.

Bathelt, A. and Jelali, M. (2014). "Comparative study of subspace identification methode on the Tennessee Eastman Process under disturbance effects". In: *Proceedings of the 5th International Symposium on Advanced Control of Industrial Processes*. Ed. by IEEE, pp. 31–36.

Bathelt, A., Ricker, N. L., and Jelali, M. (2015). "Revision of the Tennessee Eastman Process Model". In: *Proceedings of the 9th IFAC Symposium on Advanced Control of Chemical Processes*. Ed. by IFAC. Vol. 48, pp. 309–314.

Bathelt, A., Söffker, D., and Jelali, M. (2015). "An algorithm combining the subspace identification methods ORT and CCA". In: *Proceedings of the 54th Conference on Decision and Control, 2015*. Ed. by IEEE, pp. 3361–3366.

Bathelt, A., Söffker, D., and Jelali, M. (2017). "An Approach to Recursive Subspace Identification". In: *Proceedings of the 56th Conference on Decision and Control, 2017*. Ed. by IEEE, pp. 4638–4643.

Bathelt, A., Söffker, D., and Jelali, M. (2018). "Recursive Subspace Identification based on Stochastic Realization Theory". In: *Automatica* in preparation.

Bauer, D. (2001). "Order estimation for subspace methods". In: *Automatica* 37.10, pp. 1561–1573.

Bauer, D. (2005). "Asymptotic properties of subspace estimators". In: *Automatica* 41.3, pp. 359–376.

Bauer, D. and Ljung, L. (2002). "Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms". In: *Automatica* 38.5, pp. 763–773.

Caines, P. E. (1976). "Weak and Strong Feedback Free Processes". In: *IEEE Transactions on Automatic Control* 21.5, pp. 737–739.

Caines, P. E. and Chan, C. W. (1975). "Feedback between Stationary Stochastic Processes". In: *IEEE Transactions on Automatic Control* 20.4, pp. 498–508.

Caines, P. E. and Chan, C. W. (1976). "Estimation, Identification and Feedback". In: *Mathematics in Science and Engineering* 126, pp. 349–405.

Candy, J. V., Bullock, T. E., and Warren, M. E. (1979). "Invariant system description of the stochastic realization". In: *Automatica* 15.4, pp. 493–495.

Chiuso, A. (2006). "Asymptotic Variance of Closed-Loop Subspace Identification Methods". In: *IEEE Transactions on Automatic Control* 51.8, pp. 1299–1314.

Chiuso, A. (2007a). "On the Relation Between CCA and Predictor-Based Subspace Identification". In: *IEEE Transactions on Automatic Control* 52.10, pp. 1795–1812.

Chiuso, A. (2007b). "The role of vector autoregressive modeling in predictor-based subspace identification". In: *Automatica* 43.6, pp. 1034–1048.

Chiuso, A. and Picci, G. (1999). "Subspace Identification by Orthogonal Decomposition". In: *Proceedings of the 14th World Congress*. Ed. by T. Y. Chai and H.-F. Chen. Kidlington: Pergamon.

Chiuso, A. and Picci, G. (2001). "Some Algorithmic Aspects of Subspace Identification with Inputs". In: *International Journal of Applied Mathematics and Computer Science* 11.1, pp. 55–75.

Chiuso, A. and Picci, G. (2003). "Geometry of oblique splitting, minimality and Hankel operators". In: *Directions in mathematical systems theory and optimization*. Ed. by A. Rantzer, C. I. Byrnes, and A. Lindquist. Vol. 286. Lecture Notes in Control and Information Science. Berlin, New York: Springer, pp. 85–124.

Chiuso, A. and Picci, G. (2004a). "Asymptotic variance of subspace methods by data orthogonalization and model decoupling: a comparative analysis". In: *Automatica* 40.10, pp. 1705–1717.

Chiuso, A. and Picci, G. (2004b). "Numerical Conditioning and Asymptotic Variance of Subspace Estimates". In: *Automatica* 40.4, pp. 677–683.

Chiuso, A. and Picci, G. (2004c). "On the Ill-Conditioning of Subspace Identification with Inputs". In: *Automatica* 40.4, pp. 575–589.

Chiuso, A. and Picci, G. (2004d). "Subspace Identification by Data Orthogonalization and Model Decoupling". In: *Automatica* 40.10, pp. 1689–1703.

Chiuso, A. and Picci, G. (2004e). "The Asymptotic Variance of Subspace Estimates". In: *Journal of Econometrics* 118.1-2, pp. 257–291.

Chiuso, A. and Picci, G. (2005). "Consistency analysis of some closed-loop subspace identification methods". In: *Automatica* 41.3, pp. 377–391.

Chui, N. L. C. and Maciejowski, J. M. (2005). "Criteria for informative experiments with subspace identification". In: *International Journal of Control* 78.5, pp. 326–344.

Dahlén, A. and Scherrer, W. (2004). "The relation of the CCA subspace method to a balanced reduction of an autoregressive model". In: *Journal of Econometrics* 118.1-2, pp. 293–312.

De Moor, B. L. R. (1993). "The Singular Value Decomposition and Long and Short Spaces of Noisy Matrices". In: *IEEE Transactions on Signal Processing* 41.9, pp. 2826–2838.

Ding, S. X. (2013). "Data-driven design of monitoring and diagnosis systems for dynamic processes: A review of subspace technique based schemes and some recent results". In: *Journal of Process Control* 24.2, pp. 431–449.

Doob, J. L. (1990, 1953). *Stochastic processes*. 2nd ed. Wiley Classics Library. New York: John Wiley & Sons.

Downs, J. J. and Vogel, E. F. (1993). "A plant-wide industrial process control problem". In: *Computers & Chemical Engineering* 17.3, pp. 245–255.

Favoreel, W., De Moor, B. L. R., and van Overschee, P. (1999). "Subspace identification of bilinear systems subject to white inputs". In: *IEEE Transactions on Automatic Control* 44.6, pp. 1157–1165.

Fujikoshi, Y. and Veitch, L. G. (1979). "Estimation of dimensionality in canonical correlation analysis". In: *Biometrika* 66.2, pp. 345–351.

Gevers, M. R. and Anderson, B. D. O. (1981). "Representations of Jointly Stationary Stochastic Feedback Processes". In: *International Journal of Control* 33.5, pp. 777–809.

Gevers, M. R. and Anderson, B. D. O. (1982). "On Jointly Stationary Feedback-free Stochastic Processes". In: *IEEE Transactions on Automatic Control* 27.2, pp. 431–436.

Gustafsson, T. (1998). "Instrumental variable subspace tracking using projection approximation". In: *IEEE Transactions on Signal Processing* 46.3, pp. 669–681.

Gustafsson, T. (2002). "Subspace-based System Identification: Weighting and Pre-filtering of Instruments". In: *Automatica* 38.3, pp. 433–443.

Hannan, E. J. and Deistler, M. (2012). *The statistical theory of linear systems.* SIAM ed. Vol. 70. Classics in applied mathematics. Philadelphia, Pa: Society for Industrial and Applied Mathematics (SIAM 3600 Market Street Floor 6 Philadelphia PA 19104).

Hinnen, K., Verhaegen, M., and Doelman, N. (2005). "Robust spectral factor approximation of discrete-time frequency domain power spectras". In: *Automatica* 41.10, pp. 1791–1798.

Huang, B., Ding, S. X., and Qin, S. J. (2005). "Closed-loop subspace identification: an orthogonal projection approach". In: *Journal of Process Control* 15.1, pp. 53–66.

Huang, B. and Kadali, R. (2008). *Dynamic Modeling, Predictive Control and Performance Monitoring: A Data-driven Subspace Approach.* Vol. 374. Lecture notes in control and information sciences. London: Springer London.

Huang, D. and Katayama, T. (2004). "A Closed-loop Subspace Identification Method for Continuous-time Systems based on δ-Operator Model". In: *Asian Journal of Control* 6.3, pp. 341–352.

Jansson, M. (2003). "Subspace Identification and ARX Modeling". In: *System identification (SYSID '03).* Ed. by IFAC. Oxford: Elsevier.

Jansson, M. and Wahlberg, B. (1996). "A Linear Regression Approach to State-Space Subspace System Identification". In: *Signal Processing* 52.2, pp. 103–129.

Jansson, M. and Wahlberg, B. (1998). "On Consistency of Subspace Methods for System Identification". In: *Automatica* 34.12, pp. 1507–1519.

Jelali, M. (2013). *Control performance management in industrial automation: Assessment, diagnosis and improvement of control loop performance.* Advances in industrial control. London and New York: Springer.

Juricek, B. C., Seborg, D. E., and Larimore, W. E. (2001). "Identification of the Tennessee Eastman Challenge Process with Subspace Methods". In: *Control Engineering Practice* 9.12, pp. 1337–1351.

Kailath, T. (1980). *Linear systems.* Prentice-Hall information and system sciences series. Englewood Cliffs, NJ: Prentice-Hall.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1, pp. 35–45.

Kameyama, K., Ohsumi, A., et al. (2005). "Recursive 4SID-based identification algorithm with fixed input-output data size". In: *International Journal of Innovative Computing, Information and Control* 1.1, pp. 17–33.

Katayama, T. (2005). *Subspace Methods for System Identification.* Communications and Control Engineering. London: Springer-Verlag London Limited.

Katayama, T. (2010). "Subspace identification of combined deterministic-stochastic systems by LQ decomposition". In: *Proceedings of the 2010 American Control Conference.* Ed. by IEEE. New York: IEEE Press Books, pp. 2941–2946.

Katayama, T., Kawauchi, H., and Picci, G. (2005). "Subspace identification of closed loop systems by the orthogonal decomposition method". In: *Automatica* 41.5, pp. 863–872.

Katayama, T. and Picci, G. (1999). "Realization of stochastic systems with exogenous inputs and subspace identification methods". In: *Automatica* 35.10, pp. 1635–1652.

Katayama, T. and Tanaka, H. (2007). "An Approach to Closed-Loop Subspace Identification by Orthogonal Decomposition". In: *Automatica* 43.9, pp. 1623–1630.

Knudsen, T. (2001). "Consistency analysis of subspace identification methods based on a linear regression approach". In: *Automatica* 37, pp. 81–89.

Knudsen, T. (2002). "Subspace identification - Reducing uncertainty on the stochastic part". In: *Proceedings of the 15th World Congress of the International Federation of Automatic Control.* Ed. by IFAC, p. 448.

Kroll, A. and Schulte, H. (2014). "Benchmark problems for nonlinear system identification and control using Soft Computing methods: Need and overview". In: *Applied Soft Computing* 25, pp. 496–513.

Larimore, W. E. (1983). "System identification, reduced-order filtering and modeling via canonical variate analysis". In: *Proceedings of the 1983 American Control Conference.* Ed. by IEEE, pp. 445–451.

Larimore, W. E. (1990). "Canonical variate analysis in identification, filtering, and adaptive control". In: *Proceedings of the 29th IEEE Conference on Decision and Control.* Ed. by IEEE Control Systems Society. IEEE, 596–604 vol.2.

Larimore, W. E. and Baillieul, J. (1990). "Identification and filtering of nonlinear systems using canonical variate analysis". In: *Proceedings of the 29th IEEE Conference on Decision and Control.* Ed. by IEEE Control Systems Society. IEEE, 635–640 vol.2.

Lindquist, A. and Picci, G. (1996a). "Canonical Correlation Analysis, Approximate Covariance Extension, and Identification of Stationary Time Series". In: *Automatica* 32.5, pp. 709–733.

Lindquist, A. and Picci, G. (1996b). "Geometric Methods for State Space identification". In: *Identification, Adaptation, Learning: Proceedings of the NATO Advanced Study Institute From Identification to Learning, held in Como, Italy, August 22 - September 2, 1994.* Ed. by S. Bittanti and G. Picci. Vol. 153. NATO ASI series Series F, Computer and Systems Sciences. Berlin: Springer, pp. 1–69.

Lindquist, A. and Picci, G. (2015). *Linear stochastic systems: A geometric approach to modeling, estimation and identification.* Vol. v. 1. Series in contemporary mathematics. Heidelberg: Springer.

Ljung, L. (2009). *System identification: Theory for the user.* 2. ed., 11. printing. Prentice Hall information and system sciences series. Upper Saddle River, NJ: Prentice Hall PTR.

Ljung, L. and McKelvey, T. (1996a). "A least squares interpretation of sub-space methods for system identification". In: *Proceedings of the 35th IEEE Conference on Decision and Control.* Ed. by IEEE Control Systems Society. Vol. 1. Piscataway, NJ: IEEE, pp. 335–342.

Ljung, L. and McKelvey, T. (1996b). "Subspace identification from closed loop data". In: *Signal Processing* 52.2, pp. 209–215.

Lovera, M., Gustafsson, T., and Verhaegen, M. (2000). "Recursive Subspace Identification of Linear and Non-linear Wiener State-space Models". In: *Automatica* 36.11, pp. 1639–1650.

McAvoy, T. J. and Ye, N. (1994). "Base control for the Tennessee Eastman problem". In: *Computers & Chemical Engineering* 18.5, pp. 383–413.

McAvoy, T. J., Ye, N., and Gang, C. (1995). "An Improved Base Control for the Tennessee Eastman Problem". In: *Proceedings of the 1995 American Control Conference.* Ed. by IEEE. Vol. 1. Piscataway, NJ: IEEE, pp. 240–244.

McKelvey, T., Akçay, H., and Ljung, L. (1996). "Subspace-based multivariable system identification from frequency response data". In: *IEEE Transactions on Automatic Control* 41.7, pp. 960–979.

Mercère, G. and Bako, L. (2011). "Parameterization and Identification of Multivariable State-space Systems: A canonical approach". In: *Automatica* 47.8, pp. 1547–1555.

Mercère, G., Bako, L., and Lecuche, S. (2008). "Propagator-based methods for recursive subspace model identification". In: *Signal Processing* 88.3, pp. 468–491.

Mercère, G., Lecoeuche, S., and Vasseur, C. (2003). "A new recursive method for subspace identification of noisy systems: EIVPM". In: *System identification (SYSID '03)*. Ed. by IFAC. Oxford: Elsevier.

Moonen, M., De Moor, B. L. R., et al. (1989). "On- and off-line identification of linear state-space models". In: *International Journal of Control* 49.1, pp. 219–232.

Ohta, Y. (2011). "Stochastic System Transformation using Generalized Orthonormal Basis Functions with Applications to Continuous-time System Identification". In: *Automatica* 47.5, pp. 1001–1006.

Oku, H. and Kimura, H. (1999). "A recursive 4SID from the input-output point of view". In: *Asian Journal of Control* 1.4, pp. 258–269.

Oku, H. and Kimura, H. (2002). "Recursive 4SID algorithms using gradient type subspace tracking". In: *Automatica* 38.6, pp. 1035–1043.

Peternell, K., Scherrer, W., and Deistler, M. (1996). "Statistical analysis of novel subspace identification methods". In: *Signal Processing* 52.2, pp. 161–177.

Picci, G. and Katayama, T. (1996a). "A simple subspace identification method with exogenous inputs". In: *Proceedings of the 13th IFAC World Congress*. Ed. by J. J. Gertler.

Picci, G. and Katayama, T. (1996b). "Stochastic realization with exogenous inputs and 'subspace-methods' identification". In: *Signal Processing* 52.2, pp. 145–160.

Qin, S. J. (2006). "An overview of subspace identification". In: *Computers & Chemical Engineering* 30.10-12, pp. 1502–1513.

Qin, S. J., Lin, W., and Ljung, L. (2005). "A novel subspace identification approach with enforced causal models". In: *Automatica* 41.12, pp. 2043–2053.

Qin, S. J. and Ljung, L. (2003a). "Closed-loop Subspace Identification with Innovation Estimation". In: *System identification (SYSID '03)*. Ed. by IFAC. Oxford: Elsevier.

Qin, S. J. and Ljung, L. (2003b). "Parallel QR implementation of subspace identification with parsimonious models". In: *System identification (SYSID '03)*. Ed. by IFAC. Oxford: Elsevier.

Ricker, N. L. (1995). "Optimal steady-state operation of the Tennessee Eastman challenge process". In: *Computers & Chemical Engineering* 19.9, pp. 949–959.

Ricker, N. L. (2005). *Tennessee Eastman Challenge Archive*. URL: `http://depts.washington.edu/control/LARRY/TE/download.html` (visited on 02/09/2018).

Rozanov, Y. A. and Feinstein, A. (1967). *Stationary random processes*. Holden-Day series in time series analysis. San Francisco: Holden-Day.

Söderström, T. and Stoica, P. G. (2001). *System identification*. 2001 ed. Prentice Hall International series in systems and control engineering. New York [u.a.]: Prentice Hall International.

Sugie, T., Inoue, K., and Maruta, I. (2017). "Closed-loop Subspace Identification with long Data based on Nuclear Norm Minimization". In: *Proceedings of the 56th Conference on Decision and Control, 2017*. Ed. by IEEE.

Takei, Y., Imai, J., and Wada, K. (2001). "A new interpretation of subspace methods by using Schur complement". In: *Proceedings of the 40th IEEE Conference on Decision and Control*. Ed. by IEEE Control Systems Society. Piscataway, NJ: IEEE, pp. 3924–3929.

Takei, Y., Nanto, H., et al. (2006). "Recursive Updating of Error Covariance Matrix in Subspace Methods". In: *Proceedings of the 14th IFAC Symposium on System Identification, 2006 (SYSID 2006)*. Ed. by I. Mareels and B. Ninness, pp. 285–290.

Tanaka, H. and Katayama, T. (2005). "Stochastic Subspace Identification Method Guaranteeing Stability and Minimum Phase". In: *Proceedings of the 16th World Congress of the International Federation of Automatic Control*. Ed. by M. ebek and P. Zítek. Elsevier Science Ltd., pp. 152–157.

Tanaka, H. and Katayama, T. (2006). "A Stochastic Realization Algorithm via Block LQ Decomposition in Hilbert Space". In: *Automatica* 42.5, pp. 741–746.

Tanaka, H. and Katayama, T. (2007). "Minimum Phase Properties of Finite-Interval Stochastic Realization". In: *Automatica* 43.9, pp. 1495–1507.

van Overschee, P. and De Moor, B. L. R. (1993). "Subspace Algorithms for the Stochastic Identification Problem". In: *Automatica* 29.3, pp. 649–660.

van Overschee, P. and De Moor, B. L. R. (1994). "N4SID: Subspace algorithms for the Identification of combined deterministic-stochastic systems". In: *Automatica* 30.1, pp. 75–93.

van Overschee, P. and De Moor, B. L. R. (1996). *Subspace identification for linear systems: Theory, implementation, applications*. Boston: Kluwer Academic Publishers.

van Overschee, P., De Moor, B. L. R., et al. (1997). "A subspace algorithm for the identification of discrete time frequency domain power spectra". In: *Automatica* 33.12, pp. 2147–2157.

van Wingerden, J.-W. and Verhaegen, M. (2009). "Subspace Identification of Bilinear and LPV Systems for Open- and Closed-loop data". In: *Automatica* 45.2, pp. 372–381.

Verdult, V. and Verhaegen, M. (2002). "Subspace identification of multivariable linear parameter-varying systems". In: *Automatica* 38.5, pp. 805–814.

Verhaegen, M. (1993a). "Application of a subspace model identification technique to identify LTI systems operating in closed-loop". In: *Automatica* 29.4, pp. 1027–1040.

Verhaegen, M. (1993b). "Subspace Model Identification: Part 3. Analysis of the ordinary output-error state-space model identification algorithm". In: *International Journal of Control* 58.3, pp. 555–586.

Verhaegen, M. (1994). "Identification of the Deterministic Part of MIMO State Space Models given in Innovations Form from Input-Output Data". In: *Automatica* 30.1, pp. 61–74.

Verhaegen, M. and Dewilde, P. (1992a). "Subspace Model Identification: Part 1. The output-error state-space model identification class of algorithms". In: *International Journal of Control* 56.5, pp. 1187–1210.

Verhaegen, M. and Dewilde, P. (1992b). "Subspace Model Identification: Part 2. Analysis of the elementary output-error state-space model identification algorithm". In: *International Journal of Control* 56.5, pp. 1211–1241.

Verhaegen, M. and Hansson, A. (2014). "Nuclear norm subspace identification (N2SID) for short data batches". In: *Proceedings of the 19th World Congress of the International Federation of Automatic Control, 2014*. Ed. by IFAC. IFAC.

Verhaegen, M. and Hansson, A. (2016). "N2SID: Nuclear norm subspace identification of innovation models". In: *Automatica* 72.October 2016, pp. 57–63.

Viberg, M. (1995). "Subspace-based Methods for the Identification of Linear Time-invariant Systems". In: *Automatica* 31.12, pp. 1835–1851.

Wang, J. and Qin, S. J. (2002). "A New Subspace Identification Approach based on Principal Component Analysis". In: *Journal of Process Control* 12.8, pp. 841–855.

Wang, J. and Qin, S. J. (2006). "Closed-loop Subspace Identification using the Parity Space". In: *Automatica* 42.2, pp. 315–320.

Willems, J. C., Rapisarda, P., et al. (2005). "A note on persistency of excitation". In: *Systems & Control Letters* 54.4, pp. 325–329.

Wu, P., Yang, et al. (2008). "A Novel RSMI Based on Regression and Natural Power Method". In: *Proceedings of the 17th World Congress of the International Federation of Automatic Control, 2008*. Ed. by IFAC. Curran Associates, Inc., pp. 7474–7479.

Yang, B. (1995). "Projection approximation subspace tracking". In: *IEEE Transactions on Signal Processing* 43.1, pp. 95–107.

Publications

This thesis is in parts based on the results published in the following publications and was presented on the corresponding conferences. If based on the respective publication, the thesis deepens or goes beyond the results disclosed in the publication.

Bathelt, A. and Jelali, M. (2014). "Comparative study of subspace identification methode on the Tennessee Eastman Process under disturbance effects". In: *Proceedings of the 5th International Symposium on Advanced Control of Industrial Processes.* Ed. by IEEE, pp. 31–36.

Bathelt, A., Ricker, N. L., and Jelali, M. (2015). "Revision of the Tennessee Eastman Process Model". In: *Proceedings of the 9th IFAC Symposium on Advanced Control of Chemical Processes.* Ed. by IFAC. Vol. 48, pp. 309–314.

Bathelt, A., Söffker, D., and Jelali, M. (2015). "An algorithm combining the subspace identification methods ORT and CCA". In: *Proceedings of the 54th Conference on Decision and Control, 2015.* Ed. by IEEE, pp. 3361–3366.

Bathelt, A., Söffker, D., and Jelali, M. (2017). "An Approach to Recursive Subspace Identification". In: *Proceedings of the 56th Conference on Decision and Control, 2017.* Ed. by IEEE, pp. 4638–4643.

Bathelt, A., Söffker, D., and Jelali, M. (2018). "Recursive Subspace Identification based on Stochastic Realization Theory". In: *Automatica* in preparation.

Student theses

During the work a the Technische Hochschule Köln, the following student theses have been co-supervised together with Prof. Dr.-Ing. Mohieddine Jelali. The content of these student theses are however not related to the work presented in this doctoral thesis.

Bajo, A. (2013). "Implementierung einer Steuerung und Regelung mit SPS zur Positionierung eines Bandes im einem Walzprozess". Masterprojekt. Köln: Fachhochschule Köln.

Farbischewski, V. (2014). "Implementierung eines Zustandsreglers mit Beobachter unter MAT-LAB als S-Function mit C". Masterprojekt. Köln: Fachhochschule Köln.

Gürke, S. (2015). "Erweiterung des Tennessee Eastman Prozess Modelles". Masterprojekt. Köln: Fachhochschule Köln.

Heiloun, K. (2017). "Erstellung von S-Functions zur Einbindung der Tennessee Eastman Simulation Klasse". Masterprojekt. Köln: Technische Hochschule Köln.

Hufenbecher, D. and Behr, C. (2017). "Modellierung der Störprozesse des Tennessee Eastman Prozesses mittels Subspace Methoden". Masterprojekt. Köln: Technische Hochschule Köln.

Papadopoulos, K. (2014). "Überarbeitung und Erweiterung der Implementierung einer Systemidentifikation für den offenen und geschlossenen Kreis". Masterprojekt. Köln: Fachhochschule Köln.

A Tennessee Eastman Process

In the following chapter, an overview of the Tennessee Eastman Process model is given. It has been first described in Downs and Vogel (1993). The initial simulation model created by Downs and Vogel has been a FORTRAN program (without simulation loop). Later, several conversions of the FORTRAN code to MATLAB/Simulink versions have been released, of which the most commonly used is the S-function model of the Tennessee Eastman Challenge Archive of Ricker (2005). Issues with the S-function-based simulation model have resulted in a revision of the model, see Bathelt, Ricker, and Jelali (2015). The revised model and the original FORTRAN sources are available at Ricker (2005). The following overview is divided into the description of the process and the description of the simulation model.

A.1 Process description

The process is shown in terms of its piping and instrument diagram (P&ID) in Figure A.1. It is a multi-component process as it consists of four (gaseous) reactants (A, C, D, E), two
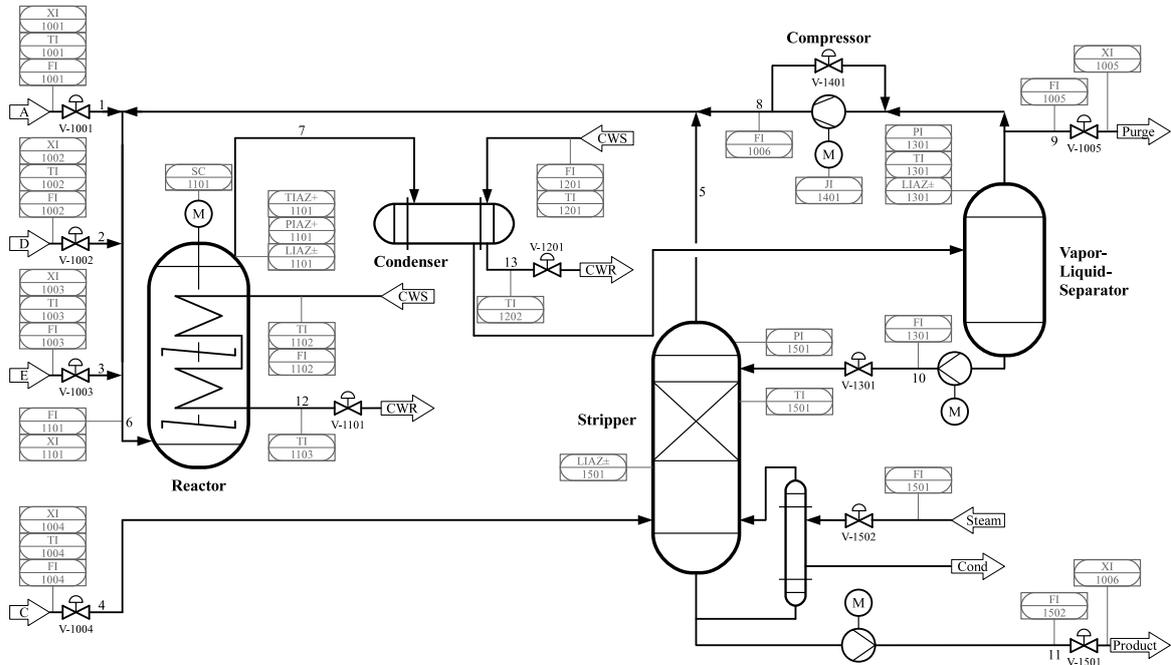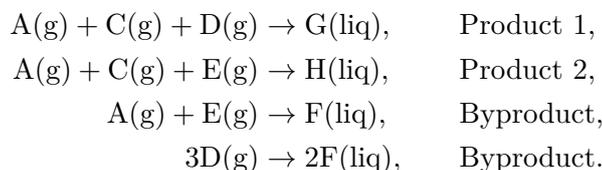

Figure A.1: P&ID of the process model (Bathelt, Ricker, and Jelali, 2015); numbers refer to the respective stream numbers

(liquid) products (G, H), one (liquid) byproduct (F) and one (gaseous) inert (B). The reaction as given in Downs and Vogel (1993) are

$$A(g) + C(g) + D(g) \rightarrow G(liq), \qquad \text{Product 1,}$$
$$A(g) + C(g) + E(g) \rightarrow H(liq), \qquad \text{Product 2,}$$
$$A(g) + E(g) \rightarrow F(liq), \qquad \text{Byproduct,}$$
$$3D(g) \rightarrow 2F(liq), \qquad \text{Byproduct.}$$

At any time, each of the component is present in all of the five major unit operations (reactor, product condenser, vapor-liquid separator, recycle compressor, product stripper) .

The reactants enter the process through streams 1 through 4 and are fed into the reactor, where the above given reactions take place. Stream 4, the feed of component C, enters the reactor not directly but through the stripper and its reactor feedback. Stream 4 is furthermore a multi-component stream as it contains not only component C but also components A and B (the feed streams of components A, D, and E contain only these components). The reactor posses a gas phase and a liquid phase. The reactions take place in the gas phase and are supported by an nonvolatile catalyst. The heat of the exothermic reactions are removed by the cooling bundle of the reactor (stream 12). Governed by the reaction kinetics, the temperature and pressure of the reactor behave contrary to each other, i.e., an increase of the temperature results in a decrease of the reactor pressure and vice versa. The products leave the reactor as vapor blended with unreacted feeds. This blend enters the condenser where, as a result of the cooling of the blend, the products become liquid. In the vapor-liquid-separator, non-condensed products as well as major parts of the unreacted feeds and the inert[1] are drawn from the blend and fed back to the reactor via the compressor. Finally, the blend, which consists of liquid D, E, F, G, and H, enters the stripper where the remaining feed and byproduct fractions are separated from the products G and H. The product mix leaves the process through stream 11. As the stripper is not only fed by the bottom fraction of the separator but also by stream 4, the product stream consist also of small fractions of components A, B, and C. Stream 9 is used to purge the system, i.e., to dispose of the inert B and the byproduct F which would otherwise accumulate in the process over time.

In Table A.1, the 73 measurements of the revised process as described in Bathelt, Ricker, and Jelali (2015) are given. The first 41 measurements stem from the original process and the remainder are measurements added to the process in the wake of the revision. Additional information on the components are given by the respective tables in Downs and Vogel (1993).

---

[1] Whereas A, B, and C are practically non-condensible, D and E are condensible and hence partly condense in the condenser.

Table A.1: Extended measurements of the Tennessee Eastman Process; indexing in terms of the output of the simulation model (Table 4 and Table 5 of Downs and Vogel, 1993 and Table 1 of Bathelt, Ricker, and Jelali, 2015)

Index	Meas. point	Description	Base value	Unit
1	FI1001	Flow A feed (stream 1)	0.25052	kscmh
2	FI1002	Flow D feed (stream 2)	3664.0	kg/h
3	FI1003	Flow E feed (stream 3)	4509.3	kg/h
4	FI1004	Flow A & C feed (stream 4)	9.3477	kscmh
5	FI1006	Recycle flow (stream 8)	26.902	kscmh
6	FI1101	Reactor feed rate (stream 6)	42.339	kscmh
7	PIAZ+1101	Reactor pressure	2705.0	kPa gauge
8	LIAZ±1101	Reactor level	75.0	%
9	TIAZ+1101	Reactor temperature	120.4	°C
10	FI1005	Purge rate (stream 9)	0.33712	kscmh
11	TI1301	Product separator temperature	80.109	°C
12	LIAZ±1301	Product separator level	50	%
13	PI1301	Product separator pressure	2633.7	kPa gauge
14	FI1301	Product separator underflow (stream 10)	25.160	m^3/h
15	LIAZ±1501	Stripper level	50	%
16	PI1501	Stripper pressure	3102.2	kPa gauge
17	FI1502	Stripper underflow (stream 11)	22.949	m^3/h
18	TI1501	Stripper temperature	65.731	°C
19	FI1501	Stripper steam flow	230.31	kg/h
20	JI1401	Compressor work	341.43	kW
21	TI1102	Reactor cooling water outlet temperature	94.599	°C
22	TI1202	Condenser cooling water outlet temperature	77.297	°C
23 - 28	XI1101	Reactor feed analysis (stream 6); components A through F	*)	mol%
29 - 36	XI1005	Purge gas analysis (stream 9); components A through H	*)	mol%
37 - 41	XI1006	Product analysis (stream 11); components A through F	*)	mol%
42	TI1001	Temperature A feed (stream 1)	45	°C
43	TI1002	Temperature D feed (stream 2)	45	°C
44	TI1003	Temperature E feed (stream 3)	45	°C
45	TI1004	Temperature A & C feed (stream 4)	45	°C
46	TI1102	Reactor cooling water inlet temperature	35	°C
47	FI1102	Reactor cooling water flow	93.37	m^3/h
48	TI1201	Condenser cooling water inlet temperature	40	°C
49	FI1201	Condenser cooling water flow	49.37	m^3/h
50 - 55	XI1001	Composition of A feed (stream 1); components A through F	*)	mol%
56 - 61	XI1002	Composition of D feed (stream 2); components A through F	*)	mol%
62 - 67	XI1003	Composition of E feed (stream 3); components A through F	*)	mol%
68 - 73	XI1004	Composition of A & C feed (stream 4); components A through F	*)	mol%

*) for base values see Table 1 of Downs and Vogel (1993, p. 247)

Table A.2: Process disturbances (Table 8 of Downs and Vogel, 1993, p. 250 and Table 5 of Bathelt, Ricker, and Jelali, 2015)

Number	Description	Type
IDV(1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV(2)	B composition, A/C feed ratio constant (stream 4)	Step
IDV(3)	D feed temperature (stream 2)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss (stream 1)	Step
IDV(7)	C header pressure loss – reduced availability (stream 4)	Step
IDV(8)	A, B, and C feed composition (stream 4)	Random variation
IDV(9)	D feed temperature (stream 2)	Random variation
IDV(10)	C feed temperature (stream 2)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	(unknown) Deviation of heat transfer within heat exchanger of stripper	(unknown) approx. random variation
IDV(17)	(unknown) Deviation of heat transfer within reactor	(unknown) approx. random variation
IDV(18)	(unknown) Deviation of heat transfer within condenser	(unknown) approx. random variation
IDV(19)	(unknown) Re-cycle valve of compressor, underflow separator (stream 10), underflow stripper (stream 11), and steam valve stripper	(unknown) sticking
IDV(20)	(unknown)	(unknown) random variation
IDV(21)	A feed temperature (stream 1)	Random variation
IDV(22)	E feed temperature (stream 3)	Random variation
IDV(23)	A feed pressure (stream 1)	Random variation
IDV(24)	D feed pressure (stream 2)	Random variation
IDV(25)	E feed pressure (stream 3)	Random variation
IDV(26)	A, B, and C feed pressure (stream 4)	Random variation
IDV(27)	pressure fluctuation in the cooling water re-circulating unit of the reactor	Random variation
IDV(28)	pressure fluctuation in the cooling water re-circulating unit of the condenser	Random variation

A.2 Simulation model description

The latest simulation model available at Ricker (2005) is the revised version described in Bathelt, Ricker, and Jelali (2015). In addition to the basic chemical-physical simulation, the model provides a number of internal process disturbances, which are listed in Table A.2. With these internal process disturbance, the process model provide a versatile range of application, like identification, monitoring, or disturbance estimation. There are three basic types of disturbances – step-like disturbances, random variation disturbances and valve sticking. Note that the random-variation disturbances are not stochastic processes but rather piece-wise defined polynomials of time, whose re-calculation interval and end-values are determined by

Table A.3: Process operating limits and shutdown constraints (Table 6 of Downs and Vogel, 1993, p. 249)

Process variable	Normal operating limits		Shutdown constraints	
	Low	High	Low	High
Reactor pressure	–	2895 kPa	–	3000 kPa
Reactor level	50% (11.8 m^3)	100% (21.3 m^3)	2.0 m^3	24 m^3
Reactor temperature	–	150 °C	–	175 °C
Product separator level	30% (3.3 m^3)	100% (9.0 m^3)	1.0 m^3	12.0 m^3
Stripper level	30% (3.5 m^3)	100% (6.6 m^3)	1.0 m^3	8.0 m^3

an internal random number generator. The variations of the heat transfer coefficients and the reaction kinetics are also basically random variations but with different characteristics. The disturbances marked with "(unknown)" were not disclosed in the original paper by Downs & Vogel. The activation and scaling of the disturbances are made by either an external input or a parameter of the simulation model. That is, there are two S-functions, one for each version of providing the activation and scaling flags of the disturbances.

As the model does not provide a complete process description, i.e., it does not include start-up and shutdown capabilities, the model is limited to a certain range of operating points, which are given in Downs and Vogel (1993, p. 247). These different operating points are called Modes, as they essentially define the amount of product (components G and H) produced and what the ratio between the product components is. As a consequence of the limitation of the operating range, the simulation model exhibits furthermore simulation constraints, which, upon violation, terminate the simulation. These shutdown constraints are given in Table A.3. For a meaningful simulation, some operating limits are also given. As these limits are coupled to measurement points of the process, the respective names of these measuring points are extended by "A" (operating limit), "Z" (shutdown limit) and "+"/"-" for the respective direction of the limit and constraint (see Table A.1).

In order to specify the initial values of the internal state, to initialize the internal random number generator, and to setup the internal structure of the process simulation, the respective parametrization of the simulation model must be made in terms of model parameters. These parameters are listed in Table A.4. The second parameter is only necessary for the S-function which not facilitates the activation and scaling of the disturbances by an additional input. In Ricker (1995), the initial values for the different Modes of the process are provided. If the parameters are omitted the simulation model switches to internal default values.

Table A.4: Parameter list of the S-functions (Bathelt, Ricker, and Jelali, 2015)

Parameter	Description
1	Array of the initial values for the 50 states of the simulation model; if empty, the default values of Mode 1 (see Downs and Vogel, 1993) are used
2	(only for S-function without additional disturbance input) Array of the activation flags of the 28 disturbances; if empty, the disturbances are switched off
3 (2)*	Initial value (seed) of the state of the random generator
4 (3)*	Model structure flag

Bit	Description
0	Activation of additional measurement points
1	Activation of monitoring outputs of the values subjected to random variations
2	Activation of monitoring outputs of the reaction and process
3	Activation of monitoring outputs of the component's concentration
4	Deactivation of measurement noise
5	Random generator uses different state variables for the process disturbances and measurement noise
6	Activation of solver-independent calculation of the process disturbances
7	Activation of disturbance scaling by the value of the activation flags; values in the interval $[0, 1]$
15	Reset structure of simulation model to original structure

∗ parameter number for S-function with additional disturbance input

B Program code of identification methods

The functions shown here are part of a superordinate simulation and identification program. Thus the parameters and return values are standardized. The construction of the Hankel matrices and evaluation of the return values is made by the superordinate program.

B.1 CCA-ORT

The implementation of the stochastic identification shown here has not been used during the examples of Chapter 4, and is hence untested.

```matlab
function sys = ccaort(U, Y, m, p, n, k)
if length(n)>1
  ndet   = n(1);              nstoch = n(2);
else
  ndet   = n;                 nstoch = n;
end;

km = size(U, 1)/2;           kp = size(Y, 1)/2;
N  = size(U, 2);

Up = U(1:km, :);             Uf = U(km + 1:2*km, :);
Yp = Y(1:kp, :);             Yf = Y(kp + 1:2*kp, :);
[Q,R] = qr([Uf; Up; Yp; Yf]',0);
L     = R';
Qdet  = Q(1:N,1:2*km);

%                 km   km   kp   kp        N
%   | Uf |    | L11   0    0    0  | | Q1' | km
%   |    |    |                      | |     |
%   | Up |    | L21  L22   0    0  | | Q2' | km
%H = |    | =  |                      | |     |
%   | Yp |    | L31  L32  L33   0  | | Q3' | kp
%   |    |    |                      | |     |
%   | Yf |    | L41  L42  L43  L44 | | Q4' | kp
L22 = L(  km +      1:2*km,              km +      1:2*km);
L33 = L(2*km +      1:2*km +   kp, 2*km +      1:2*km +   kp);
L41 = L(2*km + kp + 1:2*km + 2*kp,              1:  km);
L42 = L(2*km + kp + 1:2*km + 2*kp,   km +      1:2*km);
L43 = L(2*km + kp + 1:2*km + 2*kp, 2*km +      1:2*km +   kp);
L44 = L(2*km + kp + 1:2*km + 2*kp, 2*km + kp + 1:2*km + 2*kp);
```

```matlab
%Identification of deterministic subsystem
Rff = L42*L42' + L43*L43' + L44*L44';
Rpp = L22*L22';
Rfp = L42*L22';

[Uff, Sff, Vff] = svd(Rff); [Upp, Spp, Vpp] = svd(Rpp);
Sff = sqrtm(Sff);           Spp = sqrtm(Spp);
Lffi = Vff/Sff*Uff';        Lppi = Vpp/Spp*Upp';

OC = Lffi*Rfp*Lppi';
[~,Soc,Voc] = svd(OC);
if ndet > size(Soc,1)
  ndet = size(Soc,1);
end;
S1 = Soc(1:n, 1:n);         V1 = Voc(:, 1:n);
X  = sqrtm(S1)*V1'*Lppi*Up;

X  = X(:,1:N—1);            XX = X(:,2:N);
U = Uf(1:m,1:N—1);          Y = [L41 L42]*Qdet(1:N—1,:)';
ABCD = [XX; Y]/[X; U];
A = ABCD (1:ndet,1:ndet);
B = ABCD (1:ndet,ndet+1:ndet+m);
C = ABCD (ndet+1:ndet+p,1:ndet);
D = ABCD (ndet+1:ndet+p,ndet+1:ndet+m);

%CCA—type identifcation of stochastic subsystem (see Katayama, 2005)
Rfp = 1/N*L43*L33';
Rff = 1/N*(L43*L43' + L44*L44');
Rpp = 1/N*(L33*L33');

[Uf,Sf,Vf] = svd(Rff);      [Up,Sp,Vp] = svd(Rpp);
Sf = sqrtm(Sf);             Sp = sqrtm(Sp);
L = Uf*Sf*Vf;               M = Up*Sp*Vp';
Linv = Vf/Sf*Uf';           Minv = Vp/Sp*Up';

OC = Linv*Rfp*Minv;
[U,Sigma,V] = svd(OC);
if nstoch > size(Sigma,1)
  nstoch = size(Sigma,1);
end;
S = Sigma(1:nstoch,1:nstoch);

Ok = L*U(:,1:nstoch)*sqrtm(S);
Ck = sqrtm(S)*V(:,1:nstoch)'*M';

Lambda = Rpp(1:p,1:p);
As = Ok(1:k*p—p,:)\Ok(p+1:k*p,:);
Cs = Ok(1:p,:);             Cb = Ck(:,(k — 1)*p + 1:k*p)';
R = Lambda — Cs*S*Cs';      K = (Cb'—As*S*Cs')/R;
```

```matlab
sys = {A, B, C, D, As, K, Cs, R};
end
```

B.2 PBSID-QR

```matlab
function sys = pbsid_qr_D0(U, Y, m, p, n, k)
km = size(U, 1)/2;          kp = size(Y, 1)/2;
N  = size(U, 2);

Up = U(1:km, :);            Uf = U(km + 1:2*km, :);
Yp = Y(1:kp, :);            Yf = Y(kp + 1:2*kp, :);
Pf = zeros(kp+km,N);
for I = 1:k
    Pf(((I—1)*(m+p) + 1):I*(m+p),:) = ...
                    [Yf(((I—1)*p + 1):I*p,:); Uf(((I—1)*m + 1):I*m,:)];
end; %for I = 1:k

[Q,R] = qr([Up; Yp; Pf]',0);
L     = R';
Lp    = L(1:(kp+km),1:(kp+km));
Qp    = Q(1:N,1:(kp+km));
Lpf   = L(1:(k+(k—1))*(m+p),1:(k+(k—1))*(m+p));
Ly    = zeros(kp,(k+(k—1))*(m+p));
for I = 1:k
  Ly(((I—1)*p + 1):I*p,1:(I+k—1)*(m+p)) = ...
          L((k + I — 1)*(m+p)+1:((k + I — 1)*(m+p)+p),1:(I+k—1)*(m+p));
end; %for I = 1:k
Gamma = Ly/Lpf;
OC = Gamma(:,1:(kp+km));



[~,Sigma,VV] = svd(OC*Lp);
if n > size(Sigma,1)
  n = size(Sigma,1);
end; %if n > size(Sigma,1)
S1 = Sigma(1:n, 1:n);       V1 = VV(:, 1:n);
X  = sqrtm(S1)*V1'*Qp';

XX = X(:,2:N);              X  = X(:,1:N—1);
Ut = Uf(1:m,1:N—1);         Yt = Yf(1:p,1:N—1);
AB = XX/[X; Ut];
A = AB (1:n,1:n);           B = AB (1:n,n+1:n+m);
C = Yt/X;                   D = zeros(p,m);

W = XX — A*X — B*Ut;        E = Yt — C*X;
SigWE = [W;E]*[W;E]'/(N — 1);
```

```
QQ = SigWE(1:n,1:n);          RR = SigWE(n+1:n+p,n+1:n+p);
SS = SigWE(1:n,n+1:n+p);


[~,~,G,~] = dare(A',C',QQ,RR,SS);
K = G';


sys = {A, B, C, D, A, K, C, RR};
end
```

B.3 PBSID-VARX

Regarding the SVD step and state estimation, the implementation given in the following does
not exactly conform with the explanations in Chiuso (2007b), but yields better results.

```
function sys = pbsid_varx_D0(U, Y, m, p, n, k)
km = size(U, 1)/2;
kp = size(Y, 1)/2;
N  = size(U, 2);

Yt  = Y(kp + 1:kp + p, N−1);
Ut  = U(km + 1:km + m, N−1);
Ytv = [Y(kp+1:kp+p, :),Y((2∗k−1)∗p+1:2∗kp,(N−k+2):N)];
Zpv = zeros(km+kp,N+k−1);   Zp  = zeros(km+kp,N);
for I=1:k
  Zpv((I − 1)∗(p + m) + 1:I∗(p + m),:) = ...
                        [Y((I − 1)∗p + 1:I∗p, :), ...
                         Y((I + k − 2)∗p + 1:(I + k −1)∗p,(N−k+2):N);
                         U((I − 1)∗m + 1:I∗m, :), ...
                         U((I + k − 2)∗m + 1:(I + k −1)∗m,(N−k+2):N)];
  Zp((I − 1)∗(p + m) + 1:I∗(p + m),:) = ...
                        [Y((I − 1)∗p + 1:I∗p, :);
                         U((I − 1)∗m + 1:I∗m, :)];
end;


VARX = Ytv/Zpv;


OC  = zeros(kp−p,km+kp);
for I = 1:(k−1)
  OC((I − 1)∗p + 1:I∗p,(I − 1)∗(p + m) + 1:(km+kp)) = ...
                        VARX(:,1:(km+kp) − (I − 1)∗(p + m));
end;


[Q, R] = qr(Zp');
[~,Sigma,VV] = svd(OC∗R');
if n > size(Sigma,1)
  n = size(Sigma,1);
end; %if n > size(Sigma,1)
S1 = Sigma(1:n, 1:n);         V1 = VV(:, 1:n);
```

```matlab
X  = sqrtm(S1)*V1'*Q';

XX = X(:,2:N);                  X  = X(:,1:N−1);
AB = XX/[X; Ut];
A = AB (1:n,1:n);               B = AB (1:n,n+1:n+m);
C = Yt/X;                       D = zeros(p,m);

W = XX − A*X − B*Ut;           E = Yt − C*X;
SigWE = [W;E]*[W;E]'/(N − 1);
QQ = SigWE(1:n,1:n);           RR = SigWE(n+1:n+p,n+1:n+p);
SS = SigWE(1:n,n+1:n+p);

[~,~,G,~] = dare(A',C',QQ,RR,SS);
K = G';

sys = {A, B, C, D, A, K, C, RR};
end
```

In order to conform with the explanations in Chiuso (2007b), the following code depicts the necessary changes to the above implementation.

```matlab
function sys = pbsid_varx_D0(U, Y, m, p, n, k)

...

Yt  = Y(kp + 1:kp + p, :);
Ut  = U(km + 1:km + m, :);

...

Zp1 = [Zp; Y(kp + 1:kp+p, :); U(km + 1:km+m, :)];

...

OC  = zeros(kp−p,km+kp);    OC1 = zeros(kp−p,km+m+kp+p);
for I = 1:(k−1)
  OC((I − 1)*p + 1:I*p,(I − 1)*(p + m) + 1:(km+kp)) = ...
                        VARX(:,1:(km+kp) − (I − 1)*(p + m));
  OC1((I − 1)*p + 1:I*p,I*(p + m) + 1:(km+m+kp+p)) = ...
                        VARX(:,1:(km+kp) − (I − 1)*(p + m));
end;

...

[UU,Sigma,~] = svd(OC);
if n > size(Sigma,1)
  n = size(Sigma,1);
end; %if n > size(Sigma,1)
S1 = Sigma(1:n, 1:n);
U1 = UU(:, 1:n);
```

```
%State Estimation according to (3.6) and (3.17) in Chiuso (2007b)
Ok = U1*sqrtm(S1);
X  = Ok\OC*Zp;              XX = Ok\OC1*Zp1;


...


SigWE = [W;E]*[W;E]'/N;


...


end
```

B.4 Recursive CCA

```
function [Sys, X] = cca_rec(X, U, Y ,m ,p ,n ,k)

kfm = k*m;                 kfp = k*p;
kpm = size(U,1) - kfm;     kpp = size(Y,1) - kfp;
N   = size(U, 2);


Up = U(1:kpm, :);
Uf = U(kpm + 1:kpm + kfm, :);
Yp = Y(1:kpp, :);
Yf = Y(kpp + 1:kpp + kfp, :);

if isempty(X) == 1
  [~,R] = qr([Uf; Up; Yp; Yf]',0);
  L = R';

  L22 = L(kfm + 1:              kfm + (kpm + kpp),....
          kfm + 1:              kfm + (kpm + kpp));
  L32 = L(kfm + (kpm + kpp) + 1:kfm + (kpm + kpp) + kfp,...
          kfm + 1:              kfm + (kpm + kpp));
  L33 = L(kfm + (kpm + kpp) + 1:kfm + (kpm + kpp) + kfp, ...
          kfm + (kpm + kpp) + 1:kfm + (kpm + kpp) + kfp);
else
  %                       n
  %             kfm  kpm  kfp       N
  %                  kpp
  %     | Uf |   | L11   0    0 | | Q1' | kfm
  %     |——|   |                | |     |
  %     | X  |   |                | |     |
  %     |    |   |                | |     |
  % H = | Up |  = | L22  L22       | | Q2' | n + kpm + kpp
  %     |    |   |                | |     |
  %     | Yp |   |                | |     |
```

```matlab
  %     |——|   |                | |     |
  %     | Yf |   | L31  L32  L33 | | Q4' | kfp
  [~,R] = qr([Uf; X; Up; Yp; Yf]',0);
  L = R';

  L22 = L(kfm + 1:                    kfm + (n + kpm + kpp),....
          kfm + 1:                    kfm + (n + kpm + kpp));
  L32 = L(kfm + (n + kpm + kpp) + 1:kfm + (n + kpm + kpp) + kfp,...
          kfm + 1:                    kfm + (n + kpm + kpp));
  L33 = L(kfm + (n + kpm + kpp) + 1:kfm + (n + kpm + kpp) + kfp, ...
          kfm + (n + kpm + kpp) + 1:kfm + (n + kpm + kpp) + kfp);
end;


Rff = L32*L32' + L33*L33';
Rpp = L22*L22';
Rfp = L32*L22';



[Uff, Sff, Vff] = svd(Rff); [Upp, Spp, Vpp] = svd(Rpp);
Sff = sqrtm(Sff);           Spp = sqrtm(Spp);
Lffi = Vff/Sff*Uff';        Lppi = Vpp/Spp*Upp';

OC = Lffi*Rfp*Lppi';
[~,SS,VV] = svd(OC);
if n > size(Sigma,1)
  n = size(Sigma,1);
end; %if n > size(Sigma,1)
S1 = SS(1:n, 1:n);
V1 = VV(:, 1:n);

if isempty(X) == 1
  X  = sqrtm(S1)*V1'*Lppi*[Up; Yp];
else
  X  = sqrtm(S1)*V1'*Lppi*[X; Up; Yp];
end;

Xt  = X(:,1:N—1);           Xtt = X(:,2:N);
U = Uf(1:m,1:N—1);          Y = Yf(1:p,1:N—1);
ABCD = [Xtt; Y]/[Xt; U];
A = ABCD (1:n,1:n);         B = ABCD (1:n,n+1:n+m);
C = ABCD (n+1:n+p,1:n);     D = ABCD (n+1:n+p,n+1:n+m);

%Calculation of Kalman gain
W = Xtt — A*Xt — B*U;       E = Y   — C*Xt — D*U;
SigWE = [W;E]*[W;E]'/(N — 1);
QQ = SigWE(1:n,1:n);        RR = SigWE(n+1:n+p,n+1:n+p);
SS = SigWE(1:n,n+1:n+p);

[~,~,G,~] = dare(A',C',QQ,RR,SS);
```

```
K = G';


Sys = {A, B, C, D, A, K, C, RR};
d
```

B.5 Recursive CCA-ORT

The implementation of the stochastic identification shown here has not been used during the examples of Chapter 6, and is hence untested.

```
function [Sys X] = ccaort_rec(X, U, Y ,m ,p ,n ,k)
if length(n)>1
  ndet   = n(1);              nstoch = n(2);
else
  ndet   = n;                 nstoch = n;
end;


kfm = k*m;                   kfp = k*p;
kpm = size(U,1) − kfm;       kpp = size(Y,1) − kfp;
N   = size(U, 2);


Up = U(1:kpm, :);
Uf = U(kpm + 1:kpm + kfm, :);
Yp = Y(1:kpp, :);
Yf = Y(kpp + 1:kpp + kfp, :);


if isempty(X) == 1
  [Q,R] = qr([Uf; Up; Yp; Yf]',0);
  L     = R';
  Qdet  = Q(1:N,1:(kfm + kpm));


  %                 kfm  kpm  kpp  kfp        N
  %     | Uf |    | L11   0    0    0  | | Q1' | kfm
  %     |    |    |                      | |     |
  %     | Up |    | L21  L22   0    0  | | Q2' | kpm
  % H = |    | = |                        | |     |
  %     | Yp |    | L31  L32  L33   0  | | Q3' | kpp
  %     |    |    |                      | |     |
  %     | Yf |    | L41  L42  L43  L44 | | Q4' | kfp
  L22 = L(kfm +             1:kfm + kpm,...
          kfm +             1:kfm + kpm);
  L33 = L(kfm + kpm +       1:kfm + kpm + kpp,...
          kfm + kpm +       1:kfm + kpm + kpp);
  L41 = L(kfm + kpm + kpp + 1:kfm + kpm + kpp + kfp,...
          1:kfm);
  L42 = L(kfm + kpm + kpp + 1:kfm + kpm + kpp + kfp,...
          kfm +             1:kfm + kpm);
  L43 = L(kfm + kpm + kpp + 1:kfm + kpm + kpp + kfp,...
```

```matlab
              kfm + kpm +        1:kfm + kpm + kpp);
   L44 = L(kfm + kpm + kpp + 1:kfm + kpm + kpp + kfp,...
              kfm + kpm + kpp + 1:kfm + kpm + kpp + kfp);
else
   Xdet    = X(1:ndet,:);    Xstooch = X(ndet + 1:ndet + nstoch,:);
   [Q,R] = qr([Uf; Xdet; Up; Xstoch; Yp; Yf]',0);
   L      = R';
   Qdet  = Q(1:N,1:(kfm + ndet + kpm));

   %                 kfm    kpm    kpp    kfp       N
   %                       ndet   nstoch
   %     | Uf |    | L11     0      0      0 | | Q1' |  kfm
   %     |——|    |                          | |     |
   %     | Xd |    |                          | |     |
   %     |    |    | L21    L22     0      0 | | Q2' |  kpm + ndet
   %     | Up |    |                          | |     |
   % H = |——| = |                          | |     |
   %     | Xs |    |                          | |     |
   %     |    |    | L31    L32    L33     0 | | Q3' |  kpp + nstoch
   %     | Yp |    |                          | |     |
   %     |——|    |                          | |     |
   %     | Yf |    | L41    L42    L43    L44 | | Q4' |  kfp
   L22 = L(kfm +                     1:kfm + ndet + kpm,...
           kfm +                     1:kfm + ndet + kpm);
   L33 = L(kfm + ndet + kpm +        1:kfm + ndet + kpm + nstoch + kpp,...
           kfm + ndet + kpm +        1:kfm + ndet + kpm + nstoch + kpp);
   L41 = L(kfm + ndet + kpm  + nstoch + kpp + 1:...
           kfm + ndet + kpm  + nstoch + kpp + kfp,...
           1:kfm);
   L42 = L(kfm + ndet + kpm  + nstoch  + kpp + 1:
           kfm + ndet + kpm  + nstoch + kpp + kfp,...
           kfm +                     1:kfm + ndet + kpm);
   L43 = L(kfm + ndet + kpm + nstoch + kpp + 1:...
           kfm + ndet + kpm + nstoch + kpp + kfp,...
           kfm + ndet + kpm +        1:kfm + ndet + kpm  + nstoch + kpp);
   L44 = L(kfm + ndet + kpm + nstoch + kpp + 1:...
           kfm + ndet + kpm + nstoch + kpp + kfp,...
           kfm + ndet + kpm + nstoch + kpp + 1:...
           kfm + ndet + kpm + nstoch + kpp + kfp);
end;

%Identification of deterministic subsystem
Rff = L42*L42' + L43*L43' + L44*L44';
Rpp = L22*L22';
Rfp = L42*L22';

[Uff, Sff, Vff] = svd(Rff); [Upp, Spp, Vpp] = svd(Rpp);
Sff = sqrtm(Sff);          Spp = sqrtm(Spp);
Lffi = Vff/Sff*Uff';       Lppi = Vpp/Spp*Upp';
```

```matlab
OC = Lffi*Rfp*Lppi';
[~,Soc,Voc] = svd(OC);
if ndet > size(Sigma,1)
  ndet = size(Sigma,1);
end; %if n > size(Sigma,1)
S1 = Soc(1:ndet, 1:ndet);
V1 = Voc(:, 1:ndet);

if isempty(X) == 1
  Xdet  = sqrtm(S1)*V1'*Lppi*Up;
else
  Xdet  = sqrtm(S1)*V1'*Lppi*[Xdet; Up];
end;

Xt  = Xdet(:,1:N-1);         Xtt = Xdet(:,2:N);
U = Uf(1:m,1:N-1);           Y = [L41 L42]*Qdet(1:N-1,:)';
ABCD = [Xtt; Y]/[Xt; U];
A = ABCD (1:ndet,1:ndet);
B = ABCD (1:ndet,ndet+1:ndet+m);
C = ABCD (ndet+1:ndet+p,1:ndet);
D = ABCD (ndet+1:ndet+p,ndet+1:ndet+m);

%CCA-type identifcation of stochastic subsystem (see Katayama, 2005)
Rfp = 1/N*L43*L33';
Rff = 1/N*(L43*L43' + L44*L44');
Rpp = 1/N*(L33*L33');

[Uf,Sf,Vf] = svd(Rff);       [Up,Sp,Vp] = svd(Rpp);
Sf = sqrtm(Sf);              Sp = sqrtm(Sp);
L = Uf*Sf*Vf;               M = Up*Sp*Vp';
Sfi = inv(Sf);              Spi = inv(Sp);
Linv = Vf*Sfi*Uf';          Minv = Vp*Spi*Up';

OC = Linv*Rfp*Minv;
[U,Sigma,V] = svd(OC);
if nstoch > size(Sigma,1)
  nstoch = size(Sigma,1);
end; %if n > size(Sigma,1)
Lambda = Rpp(1:p,1:p);
S = Sigma(1:nstoch,1:nstoch);

Ok = L*U(:,1:nstoch)*sqrtm(S);
Ck = sqrtm(S)*V(:,1:n)'*M';
Xstoch = sqrtm(S)*V(:,1:n)'*Minv*L33*Q3';

As = Ok(1:k*p-p,:)\Ok(p+1:k*p,:);
Cs = Ok(1:p,:);
Cb = Ck(:,(k - 1)*p + 1:k*p)';
```

```
R = Lambda − Cs∗S∗Cs';
K = (Cb'−As∗S∗Cs')/R;


Sys = {A, B, C, D, As, K, Cs, R};
X   = [Xdet; Xstoch];
end
```

B.6 Recursive PBSID-QR

```
function [Sys, X] = pbsid_qr_D0_rec(X, U, Y ,m ,p ,n ,k)
kfm = k∗m;                  kfp = k∗p;
kpm = size(U,1) − kfm;      kpp = size(Y,1) − kfp;
N   = size(U, 2);

Up = U(1:kpm, :);
Uf = U(kpm + 1:kpm + kfm, :);
Yp = Y(1:kpp, :);
Yf = Y(kpp + 1:kpp + kfp, :);

Pf = zeros(kfp + kfm,N);
for I = 1:k
  Pf(((I−1)∗(m + p) + 1):I∗(m + p),:) = ...
  [Yf(((I−1)∗p + 1):I∗p,:); Uf(((I−1)∗m + 1):I∗m,:)];
end; %for I = 1:k

if isempty(X) == 1
  [Q,R] = qr([Up; Yp; Pf]',0);
  L = R';
  Lp = L(1:(kpp + kpm),1:(kpp + kpm));
  Qp = Q(1:N,1:(kpp + kpm));

  Lpf = L(1:(k+(k−1))∗(m+p),1:(k+(k−1))∗(m+p));
  Ly = zeros(kfp,(k+(k−1))∗(m+p));
  for I = 1:k
    Ly(((I−1)∗p + 1):I∗p,1:(I+k−1)∗(m+p)) = ...
         L((k + I − 1)∗(m+p)+1:((k + I − 1)∗(m+p)+p),1:(I+k−1)∗(m+p));
  end; %for I = 1:k
  Gamma = Ly/Lpf;
  OC = Gamma(:,1:(kpp + kpm));
else
  [Q,R] = qr([X; Up; Yp; Pf]',0);
  L     = R';
  Lp    = L(1:(n + kpp + kpm),1:(n + kpp + kpm));
  Qp    = Q(1:N,1:(n + kpp + kpm));

  Lpf = L(1:n + kpp + kpm + (k−1)∗(m+p),1:n + kpp + kpm + (k−1)∗(m+p));
```

```matlab
  Ly = zeros(kfp,n + kpp + kpm + (k—1)*(m+p));
  for I = 1:k
    Ly(((I—1)*p + 1):I*p,1:n + kpp + kpm + (I—1)*(m+p)) = ...
          L( n + kpp + kpm + (I — 1)*(m+p) + 1:...
            (n + kpp + kpm + (I — 1)*(m + p) + p),...
            1:n + kpp + kpm + (I — 1)*(m+p));
  end; %for I = 1:k
  Gamma = Ly/Lpf;
  OC = Gamma(:,1:(n + kpp + kpm));
end; %if isempty(X) == 1

[~,Sigma,VV] = svd(OC*Lp);
if n > size(Sigma,1)
  n = size(Sigma,1);
end; %if n > size(Sigma,1)
S1 = Sigma(1:n, 1:n);
V1 = VV(:, 1:n);
X  = sqrtm(S1)*V1'*Qp';

Xt  = X(:,1:N—1);          Xtt = X(:,2:N);
Ut  = Uf(1:m,1:N—1);       Yt  = Yf(1:p,1:N—1);
AB = Xtt/[Xt; Ut];         C  = Yt/Xt;
A = AB (1:n,1:n);          B = AB (1:n,n+1:n+m);
D = zeros(p,m);

%Calculation of Kalman gain
W = Xtt — A*Xt — B*Ut;     E = Yt  — C*Xt;
SigWE = [W;E]*[W;E]'/(N — 1);
QQ = SigWE(1:n,1:n);       RR = SigWE(n+1:n+p,n+1:n+p);
SS = SigWE(1:n,n+1:n+p);

[~,~,G,~] = dare(A',C',QQ,RR,SS);
K = G';

Sys = {A, B, C, D, A, K, C, RR};
end
```