Université de Lille

CRIStAL Centre de Recherche en Informatique, Signal et Automatique de Lille

cnrs

Towards Stochastic Realization Theory for Linear Switched Models

THESIS SUBMITTED TO THE UNIVERSITY OF LILLE IN THE SUBJECT OF AUTOMATIC CONTROL AND COMPUTER SCIENCE FOR THE DEGREE OF PHD

Graduate school n°631 MAthematics and DIgital Science (MADIS)
Thesis prepared in Villeneuve d'Ascq, at CRIStAL UMR 9189,
on the 24th of October 2025, by

ELIE ROUPHAEL

Defended in front of the examination board:

Laurent MEVEL Research Director, Inria, Centre Inria de l'université de Rennes	President / Reviewer
Yacouba BOUBACAR MAINASSARA Full professor, University of Hauts-de-France	Reviewer
Pauline KERGUS Research scientist, LAPLACE, University of Toulouse	Examiner
Guillaume MERCERE Full professor, LIAS – ENSIP, University of Poitiers	Examiner
Ralf PEETERS Full professor, Department of Advanced Computing Sciences, University of Maastricht	Examiner
Mihály PETRECZKY Research scientist, CRIStAL, Centrale Lille	Ph.D. co-supervisor
Lotfi BELKOURA Full professor, CRIStAL, University of Lille	Ph.D. advisor

DOCTORAL THESIS

Université de Lille

CRIStAL Centre de Recherche en Informatique, Signal et Automatique de Lille

cnrs

Contribution à la théorie de la réalisation stochastique pour les systèmes linéaires à commutation

Thèse présentée à l'Université de Lille en automatique et informatique en vue de l'obtention du grade de docteur

École doctorale n°631
Mathématiques, sciences du numérique et de leurs interactions (MADIS)
Thèse préparée à CRIStAL UMR 9189, soutenue à Villeneuve d'Ascq
le 24 Octobre 2025 , par

Elie ROUPHAEL

Soutenue devant le jury composé de :

Laurent MEVEL Directeur de recherche, Inria, Centre Inria de l'université de Rennes	Président / Rapporteur
Yacouba BOUBACAR MAINASSARA Professeur, Université de Hauts-de-France	Rapporteur
Pauline KERGUS Chargée de Recherche, LAPLACE, Université de Toulouse	Examinatrice
Guillaume MERCERE Professeur, LIAS – ENSIP, Université de Poitiers	Examinateur
Ralf PEETERS Professeur, Department of Advanced Computing Sciences, Université de Maastricht	Examinateur
Mihály PETRECZKY Chargé de recherche, CRIStAL, Centrale Lille	Co-encadrant
Lotfi BELKOURA Professeur, CRIStAL, Université de Lille	Directeur de thèse

Thèse de doctorat

A disciple is not above his teacher,
but everyone who is perfectly trained
will be like his teacher.

Luke 6:40

Acknowledgements

I would first like to express my heartfelt gratitude to my parents and my sister. Despite the physical distance between us, their constant love and support have been a guiding force throughout this journey.

I would also like to warmly thank the members of my thesis jury for accepting to evaluate my work and for their valuable time and insights. I am honored by the presence of Yacouba Boubacar Mainassara and Laurent Mevel (Reviewers), Pauline Kergus, Guillaume Mercere and Ralf Peeters (Examiners). Their expertise and feedback are deeply appreciated.

I am deeply thankful to my thesis director, Lotfi Belkoura, for his valuable guidance and encouragement, and to Laurentiu Hetel, head of the CO^2 team at CRIStAL, for his support and leadership.

I would also like to extend a special thanks to the team at SZTAKI in Budapest, Hungary. Their warm welcome and stimulating research environment significantly enriched my academic experience.

To my friend, Paul Melki — thank you for being by my side through the most difficult moments of this thesis. Your unwavering mental and emotional support made a real difference, and I am truly grateful for your presence.

Above all, I dedicate this manuscript to my supervisor, Mihály Petreczky. Everything I know today, I owe to his exceptional mentorship, patience, intellectual generosity, and commitment. His influence has been instrumental in shaping both my academic path and my way of thinking.

Abstract

This thesis develops a stochastic realization theory for Generalized Linear Switched Systems (GLSS), a model class that unifies and extends several classical and modern dynamical system classes such as Linear Switched Systems, Jump-Markov Models, LPV systems, and piecewise-linear systems. These models are suitable for capturing the behavior of nonlinear and time-varying systems through a combination of multiple linear subsystems governed by a switching signal.

The central contribution of the thesis is the formulation of a stochastic realization framework for GLSS models with inputs and i.i.d. switching. The study begins with a general decomposition result (Chapter 3), showing that any output process of a GLSS can be uniquely decomposed into deterministic and stochastic components. This separation enables the reuse of existing deterministic realization results and sets the stage for a unified treatment of stochastic GLSS realization in innovation form.

In Chapter 4, the thesis focuses on autonomous GLSS models and provides an algebraic characterization of minimality and innovation form. It establishes sufficient conditions — in terms of the rank and observability properties of the system matrices — under which a stochastic autonomous GLSS admits a minimal representation in innovation form. These results are pivotal in guaranteeing identifiability and uniqueness up to isomorphism.

Chapter 5 first introduces a covariance-based realization algorithm for GLSS with inputs, which constructs a minimal stochastic realization in innovation form using only input-output covariance data. This algorithm addresses key challenges in subspace system identification, including dimensionality reduction via nice-selections, minimality in innovation form, and uniqueness up to isomorphism. Building on this, a complete identification procedure is proposed: it estimates the required covariances from input-output data, and then applies the realization algorithm. The identification method is shown to be asymptotically consistent and computationally efficient. Finally, a gradient-based refinement step is introduced to enhance the accuracy of the estimated parameters.

The thesis concludes by summarizing these contributions and outlining several future directions, including generalizations to non-i.i.d. switching and time series applications.

Résumé de la thèse

Cette thèse développe une théorie de la réalisation stochastique pour les modèles appelés Generalized Linear Switched Systems (GLSS), une classe unificatrice englobant plusieurs modèles dynamiques classiques et modernes, tels que les systèmes linéaires commutés (LSS), les modèles de Markov à sauts (JMLS), les systèmes linéaires à paramètres variant (LPV). Ces modèles permettent de représenter des comportements non linéaires et temporellement variables en combinant plusieurs sous-systèmes linéaires gouvernés par un signal de commutation.

La contribution principale de cette thèse est la formulation d'une théorie de la réalisation stochastique pour les GLSS avec entrées et signal de commutation i.i.d. Le travail débute par un résultat de décomposition (Chapitre 3), démontrant que tout processus de sortie d'un GLSS peut être décomposé de manière unique en une composante déterministe (dépendante des entrées) et une composante stochastique (dépendante du bruit). Cette décomposition permet de tirer parti des résultats existants sur la réalisation déterministe et sert de base pour l'étude des représentations sous forme d'innovation.

Le Chapitre 4 se concentre sur les GLSS autonomes stochastiques. Il propose une caractérisation algébrique permettant de vérifier si une telle représentation est minimale et sous forme d'innovation. Les conditions proposées sont suffisantes, et s'expriment uniquement en fonction des matrices du modèle. Ces résultats permettent de garantir l'unicité (à isomorphisme près) de la représentation, ce qui est crucial pour les problèmes d'identifiabilité.

Le Chapitre 5 propose d'abord un algorithme de réalisation basé sur les covariances d'entrée-sortie, permettant de construire une représentation stochastique minimale d'un GLSS sous forme d'innovation, à partir de moments statistiques. Cet algorithme traite des enjeux fondamentaux liés à l'identification par sous-espace, notamment la réduction de dimension via des sélections appropriées, la garantie de minimalité sous forme d'innovation, ainsi que l'unicité à un isomorphisme près. Sur cette base, une procédure complète d'identification est ensuite formulée : elle consiste à estimer les covariances à partir de données d'entrée-sortie, puis à appliquer l'algorithme de réalisation. La cohérence asymptotique de l'identification est démontrée, et son efficacité computationnelle est assurée. Enfin, une phase de raffinement par descente de gradient est proposée pour améliorer la précision des paramètres estimés.

Contents

List of Acronyms and Abbreviations

aGLSS	autonomous Generalized Switched System
ARMAX	Autoregression Moving Average with eXogenous inputs model
ARX	Auto Regression with eXogenous inputs model
asGLSS	autonomous stationary Generalized Linear Switched System
CCA	Canonical Correlation Analysis
CVA	Canonical Variate Analysis
DDLC	Data-Driven Local Coordinate
dLSS	deterministic Linear Switched System
EM	Expectation-Maximization
GB	Gradient-Based
GBS	General Bilinear System
GLSS	Generalized Linear Switched System
HEV	Hybrid Electric Vehicles
i.i.d.	Independent and Identically Distributed
I/O	Input/Output
JMLS	Jump-Markov Linear Switched
KCCA	Kernelized Canonical Correlation Analysis
LMI	Linear Matrix Inequalities
LPV	Linear Parameter-Varying
LS-SVM	Least-Squares Support Vector Machine
LSS	Linear Switched System
LTI	Linear Time-Invariant
MA	Moving Average
MOESP	Multivariable Output-Error State sPace
N4SID	Numerical Subspace State-Space System IDentification
PBSID	Predictor Based Subspace IDentification
PEM	Prediction-Error Minimization

PO-MOESP	Past-Output Multivariable Output-Error State-sPace
SII	Square Integrable with respect to the switching
SS	State-Space
SVD	Singular Value Decomposition
sGLSS	stationary Generalized Linear Switched System
WSS	Wide Sense Stationary
ZMWSSI	Zero Mean Wide Sense Stationary with respect to the switching

List of Algorithms

Chapter 1

General introduction

Control of dynamical systems is a mature field whose history stretches back to the 1930's. Its objective is to change the behavior of dynamical systems by using feedback. In a nutshell, a dynamical system is a system in which variables of different kinds (inputs, disturbances, noise, etc.) interact and produce observable signals (outputs) [66]. In other words, a dynamical system determines the evolution of the observed signals in time, discretely or continuously, with internal variables describing the states of the system. Feedback is a mechanism which maps measured outputs to control inputs. It is often referred to as feedback controller, or controller for short.

The design of a feedback controller relies directly or indirectly on models. In particular, the quality of the feedback controller is significantly influenced by the quality of these models. For this reason, to be able to compute feedback controllers in a scientifically sound way, the class of dynamical systems under consideration has to be restricted to those that can be adequately described by suitable models. Below we will discuss a number of model classes that are widely used in control theory.

The most popular model class used in controller synthesis is the class of *linear systems*. Linear systems are models which specify a *linear* relationship between the various signals. Mathematically, they can be described by linear difference/differential equations. A particularly convenient class of linear models are *Linear Time-Invariant* (LTI) state-space (SS) models. There is a rich and complete literature on designing feedback controllers for LTI-SS models, for various control objectives.

However, linear models, including LTI-SS, are very limited when discussing practical, realistic applications due to the nonlinear behavior of these applications. Therefore, many models were developed for describing nonlinear dynamic systems, such as bilinear models, polynomial models, input-affine models and many others. Nevertheless, while these models are accurate representations of the underlying system, they are difficult to use for synthesizing feedback controllers. Hence, in order to combine the accuracy of nonlinear models with the favorable properties of linear ones, researchers have proposed various classes of models which are composed of several linear ones. From these models we mention Linear Switched Systems (LSS), Linear Parameter-Varying (LPV) models and piecewise-linear models.

In addition, in the field of control engineering, there are many challenges that are characterized by irregular random fluctuations in the observed signals. Examples include environmental variables such as temperature, humidity, air quality or climate control; traffic flow in urban areas, including fluctuations in vehicle speeds, densities, etc.; fluctuations in heart rate variability or neural activity in the brain in the biological systems; energy consumption patterns in smart grid; economic indicators; and many others. These fluctuations can be modeled mathematically by a stochastic control system. In system theory, many theorems and concepts are offered in order to formulate stochastic control systems. Difference/differential equations with stochastic noise (and possible random parameters) are mathematical models developed in the interest of describing stochastic linear systems, and applying stochastic controllers. There is a rich theory for linear stochastic systems, but linear stochastic systems are not always accurate, and nonlinear stochastic systems are difficult to use. Jump-Markov Linear Switched (JMLS) models offer a compromise between linear and nonlinear stochastic systems: they are approximations of nonlinear systems but still inherit some nice properties of the linear ones.

From the above discussion, it becomes rather clear that high-quality models are essential for the design of feedback controllers. While models based on domain knowledge (i.e., first-principle models) are ideal, their development is often challenging or even impractical due to their complexity. An alternative is to learn models of dynamical systems from data. The latter problem is the subject of the branch of control theory known as system identification. Algorithms for identifying stochastic LTI systems are well understood [14], but system identification of stochastic nonlinear models is not well-developed in the literature.

This thesis aims at contributing to system identification of a class of models referred to as Generalized Linear Switched Systems (GLSS). The models belonging to this class can capture nonlinear behavior, which makes them a good choice for modeling nonlinear dynamical systems with the aim of designing feedback controllers. More precisely, a GLSS model is considered to be a good model for nonlinear systems, because it maintains the linear relation between the input and the output of the system while conserving the nonlinear behavior. In turn, the preserved linearity between inputs and outputs allows the extensions of proven techniques from controller design of linear systems and robust control. In fact, the existing literature proposes computationally efficient and theoretically sound control design methods for several subclasses of GLSS.

However, there remain many open problems in system identification of GLSS, especially for GLSS with stochastic noise. As the first step towards solving these open problems, one needs to develop the so called *realization theory* of stochastic GLSS. In a nutshell, realization theory is an idealized version of the learning problem studied in system identification, where it is assumed that the input/output behavior of the underlying system can be completely measured. That is, it is assumed that we can make an infinite amount of arbitrary experiments with the system.

In this thesis we develop stochastic realization theory for the GLSS class and we demonstrate its utility by using it for formulating a theoretically sound system identification algorithm.

1.1 Informal representation of the GLSS framework

Generalized linear switched systems is a class of nonlinear systems. GLSS class includes various well-known classes of models such as LSS, JMLS, LPV with affine dependency on the scheduling signal, piecewise-linear systems and many other subclasses. The latter classes are all popular classes of models used in control theory. In general, a GLSS is an extension of LSS that allows for more general structure and has piece-wise linear dynamics: it consists of multiple, possibly infinite, linear subsystems, and the system can switch between them according to a predefined or state-dependent switching rule. Figure 1.1 illustrates the GLSS framework. All the subclasses of the GLSS class can be found by Fig 1.1, however, further restrictions on the switching signal are required for some of these subgroups.

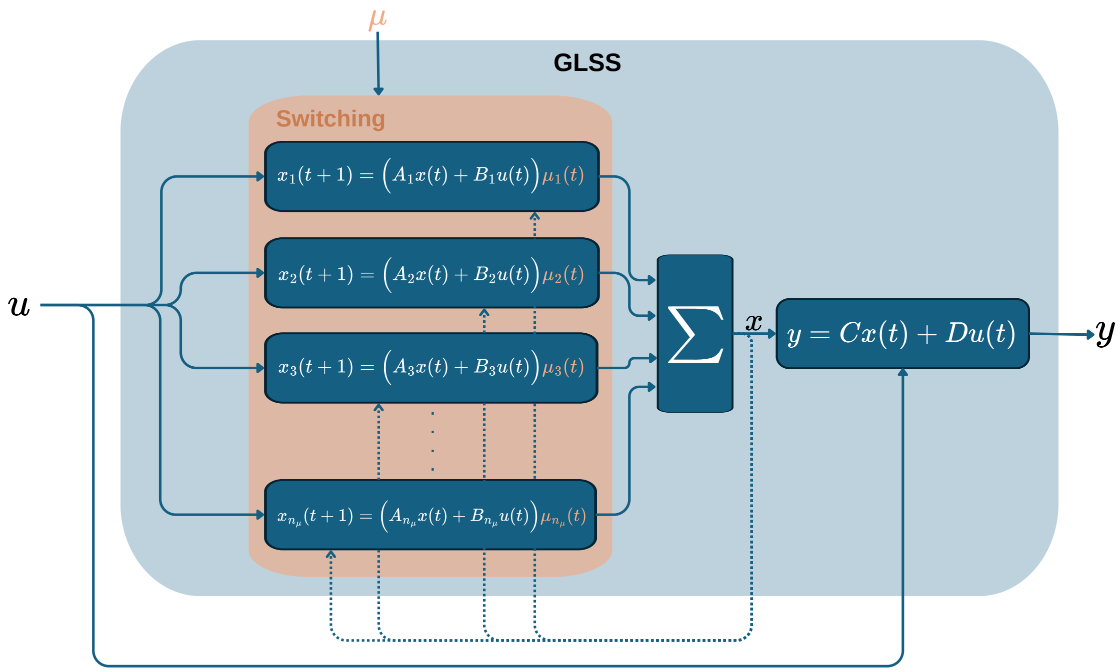

Figure 1.1: The GLSS framework shows the relation between the input signal u and the output signal y, dictated by the switching signal μ. The system is described by the state signal x which is a function of μ and all n_μ state signals of the LTI models, where the integer $n_\mu > 1$.

The main motivation for regrouping all these classes into one class is that they share certain relevant mathematical properties. The results developed in this thesis can be applied to many classes, therefore, having a class that include as many subclasses as possible can be very convenient for the different research communities.

The attentive reader might have noticed that the definition of GLSS resembles closely to the class of the so called hybrid systems. Recall that, generally, a hybrid system is a combination of continuous time (real-time) dynamics and discrete time dynamics (discrete events). However, many difficulties arise when using the latter term.

The main difficulty with using the term "hybrid systems" is that the latter is not an exclusive term, meaning that it can include a wide classes of dynamical systems to which some technical assumption has to be introduced in order to restrict the term "hybrid". Also, several communities from different fields (computer science, modeling, machine learning, systems and control, etc.) contribute to the area, each using their own definition.

In addition, the contributions of this thesis can be applied to the LPV class and while some subclasses of the LPV class are considered to be hybrid systems, others are not. Therefore, this is another difficulty that discouraged us from using the term "hybrid systems". For more details on hybrid systems and some examples, see [117, 25].

In short, the term GLSS is more restricted and exclusive than the "hybrid systems" term. Figure 1.2 shows the relationship between all the classes mentioned above.

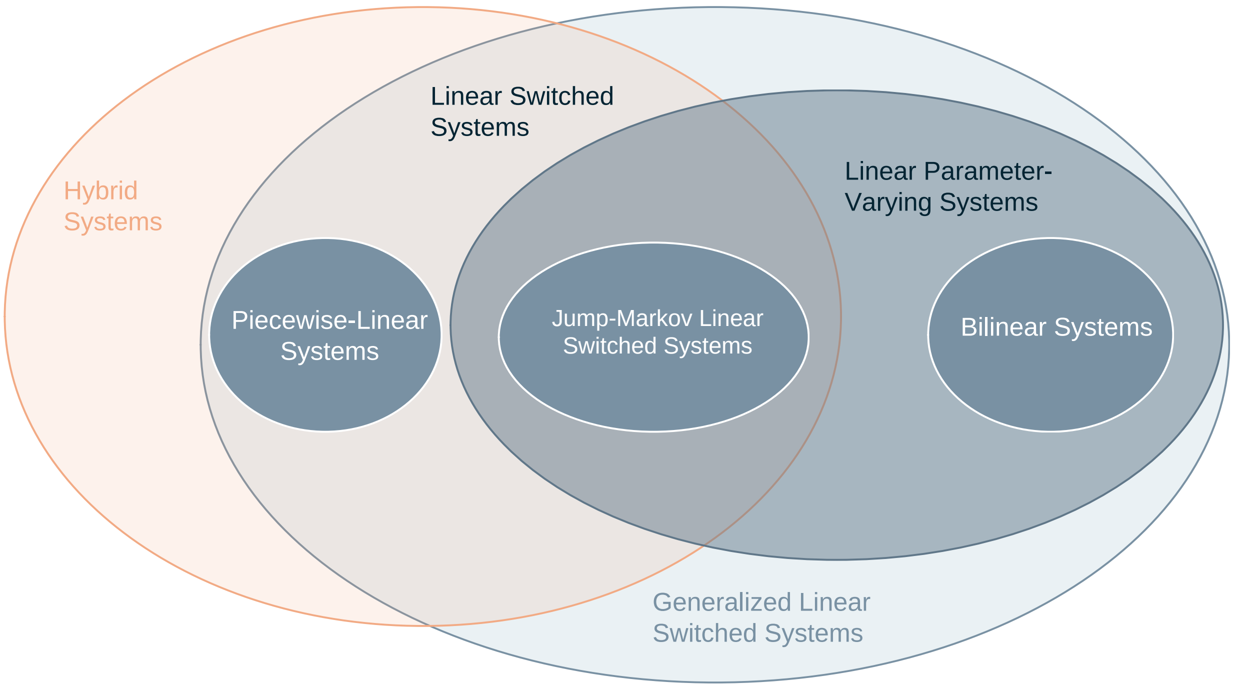

Figure 1.2: Relationship between the system classes.

The main interest of the GLSS model class is that they can represent a large variety of models which arise by approximating nonlinear models by a (possibly infinite) collection of LTI models. There are many ways to construct such an approximation. A particularly popular method is to linearize the nonlinear model at several (possibly infinite number) of operating points. While mathematical nonlinear models are more accurate in describing real models compared to LTI models, it is preferable to work with the latter for their reliability, practicality, stability and high performance. Also, engineers have more experience with LTI models. Additionally, practice has shown that a lot of nonlinear systems can be approximately modelled by one or more LTI models. Therefore, in the last couple decades of the twentieth century, researchers and scientists has adopted the concept of switching between the LTI models (this concept is also known as the gain-scheduling concept, for more details see the thesis of J.S. Shamma [104] and some follow up results [103, 105, 106]). While preserving the nonlinearity and the time-varying aspect of the real system, this approach helps in describing the system with linear relations between the input signal and the output signal at several operating points. Those relations, however, are dictated by what is called the switching signal denoted as μ in this thesis (in the literature and many references, it is sometimes referred to as p).

We should mention that LTI theory and tools cannot, in general, be used for GLSS without some modifications and that additional theory and specialised algorithms are needed in order to combine controllers computed for LTI subsystems into a controller which achieves the desired control objectives for the collection of the LTI systems. For example, it is well-known that even if a controller stabilises all the LTI systems, it may fail to stabilise the resulting GLSS if the signal μ changes too fast. However, the necessary theory is relatively well understood, and suitable controllers can often be computed using LMI (Linear Matrix Inequalities) techniques. In the same manner, existing theory of system identification and realization theory for LTI systems cannot be used for GLSS models without non-trivial extensions. In fact, new theory for system identification and realization theory, specific for GLSS models, has to be developed. This theory is inspired by and extends the corresponding results for LTI models, but it does not use results from LTI theory.

In what follows, we will briefly present various popular subclass of GLSS models and mention some examples where they were applied.

Linear parameter-varying systems: In the literature of LPV, the switching signal is usually referred to as the scheduling signal. The LPV approach has gained popularity after the introduction of the concept of gain-scheduling. One may argue that GLSS are quite close to LPV systems. However, in contrast to LPV systems, in general we are agnostic about the role of μ in control, hence the use of the term "Generalized".

More precisely, in LPV systems, the scheduling is assumed to be an arbitrary, usually bounded signal, which can be used by the controller to compute a control action, but which cannot be changed by the controller. One of the famous applications of LPV is the aircraft control application [76, 69, 124]. We know that, at constant altitude, the system can be seen as a LTI [17, 109]. Then the altitude can be considered as the scheduling signal dictating the LTI to be used. Another intuitively appealing example of using LPV systems is vehicle control. For vehicle systems, the scheduling signal can represent factors such as damping, road conditions, deflection, or certain nonlinear components. With an appropriate choice of scheduling, the relationship among the remaining variables becomes linear. This allows the transformation of the nonlinear system into a LPV system [31]. For a comprehensive overview of LPV system applications, see [75].

Among LPV systems, a particularly popular class is the one with *affine and static dependency on* *the parameters*. The latter means that the coefficients defining the linear relationship between the state, input and output signals are affine function of the current value of the scheduling signals [113]. This is in contrast to models with a dynamic and nonlinear dependency on parameters, where these coefficients are nonlinear functions of not only current but also past or future values of the scheduling signals (in discrete time) or high-order derivatives of the scheduling signal at a given time (in continuous time).

Linear switched systems: LSSs are, by definition, multi-model in nature, and like GLSSs, the switching is considered to be arbitrary. However, contrary to a GLSS, the switching signal of a LSS can only take a finite number of values. By way of illustration, in the LSS case, the state signal in Figure 1.1 is only equal to one state of a LTI chosen by the switching signal. The boost converter is a famous example of using LSS in the electrical engineering field, where the switching can only take two values (on or off). The switching can be controlled by an external signal, for instance, using a pre-determined pulse-width modulator is a common way to do so. For more details on the model design see [22]. The authors of [140] provide a good overview on the LSS models and examples. In addition, contrary to LPV systems, the switching signal can be a control input.

Piecewise-linear systems: In the literature, piecewise-linear systems are considered to be LSSs. However, in contrast to a LSS, the switching of a piecewise-linear system is not arbitrary, but state-dependent. In control theory for example, applying a saturated feedback control on a system is very common. This can make the system piecewise linear, because the system's current linear behavior depends on the state spaces induced by the actuator feedback saturation. For more details, see [50].

Jump-Markov linear switched systems: A JMLS is a LSS where its sub-models (modes) switch or *jump* according to a Markov chain, meaning that the switching is not arbitrary but a Markov process. Note that, if the probabilities of the jumps are all equal, we say that the JMLS model is a regular stochastic LSS model. For instance, in [112] it was desired to control the water flow through a boiler dependently on the weather, where its changes are dictated by a Markov process. To this end, a JMLS model was studied. JMLS are also used in economic theory in order to study the behavior of the system in question due to some changes in the economic periods of the country. For example, choosing the optimal policy for the national income can be determined by studying a JMLS model [51] were the switching between the modes is based on historical data of the nation's income. For an overview on JMLS models and their application see [18].

1.2 System Identification

Having a high-precision, low-complexity mathematical models is of paramount importance for the control. Laws of physics, chemistry, etc., - or what we call first principles - can be used to identify the model in question for synthesizing feedback controllers. However, such methods are time consuming or they are not practical because the equations are often very complex or they contain unknown parameter. System identification, in contrast, involves building mathematical models of dynamical systems based on measured data. Such data-driven methods have proven to be useful in many applications.

The procedure to identify models from the observed data is well understood and it is known as the system identification loop.

Procedure (System identification loop [66])**.**

Step 1) ***Experiment design and data processing:*** *the experiment design should focus on obtaining informative data sets. This is done by choosing persistently exciting inputs*[1]*. Such inputs result in having necessary information in the output signals in order to describe the dynamical relations between the signals. For instance, white noise inputs, for LTI systems, represent the simplest example of persistently exciting inputs.*

Step 2) ***Model structure selection:*** *this is considered to be a crucial step in the system identification loop. In this step, we determine the model structure in terms of representation form, parametrization, and the type of the noise modeling. In addition, the size of the model should be determined (number of parameters or the order of the model). Some model predictions can be constrained, with regard to the size of the model, due to some requirements of the to-be-identified system.*

Step 3) ***Choice of the identification criterion:*** *in this step, we choose the identification criteria which formulates, mathematically, the performance measure of the model estimate. There has been a lot of criteria for the identification problem, but the least-squares are still considered to be a very reliable choice.*

Step 4) ***Estimation of the model:*** *the model estimation step is the result of the model structure choice, the processed data and the criteria selected. The resulting model is obtained in this step.*

Step 5) ***Validation/invalidation of the estimated model:*** *In order to determine whether the estimate is a good model or not, measured data and prior knowledge should be used. The measured data is compared to the validation data set in order to verify (or refuse) the model estimate. Prior knowledge and past measured data are used to analyse the performance of the model estimate as a predictor of future outputs of the system. If the model estimate fails, we go back to previous steps until the validation is accomplished.*

In this thesis we are interested in the case when, in *Step 2)*, we choose a subclass of GLSS. System identification methods can be divided into two categories: parametric and subspace methods. Below, we present a brief overview of both approaches for GLSS models.

[1]Notions on persistently exciting inputs for linear switched systems can be found in [82].

1.3 Parametric identification methods for GLSS

A parametric identification method for state-space representation can fall in one of these two classes:

- Estimating so called Input/output (I/O) models from data and then transforming the estimated model to state-space forms.
- Direct parametric estimation method for the state-space identification.

In the following, we will talk about each class by presenting its contribution in the literature for the GLSS subclasses.

1.3.1 I/O to state-space

In general, contrary to state-space models that describe the I/O behavior with a state-space representation, I/O models describe linear/nonlinear **direct** relationship between the inputs and the outputs. One of the popular input-output model structure choice, for LTI systems, is the Auto Regression with eXogenous inputs (ARX) model. It is known for its simplicity and straightforwardness in describing the behavior of the system with a linear equation relating the inputs to the outputs. Later, the AutoRegression Moving Average with eXogenous inputs (ARMAX) model extended the ARX model. Contrary to ARX, the ARMAX incorporates a moving average (MA) component of the noise to account for system noise or disturbances more effectively. These linear I/O models are popular is system identification, as they can be estimated using methods which rely on extensions to the classical techniques for linear regression. In particular, there is a rich literature on algorithms and their analysis for the identification of such models. Motivated by the usefulness of the ARX models for LTI systems, the ARX models were extended to LPV and switched cases. See for example, [130, 68, 67, 33] for switched and piecewise linear systems, and [9, 36, 133, 132] for the LPV subclass, where they introduced an equivalence of ARX and ARMAX models for GLSS systems. Informally, an LPV-ARX (respectively switched ARX) is a classical ARX model, where its parameters depend on the scheduling (respectively the switching) signal.

While I/O models are useful for system identification, it is found that, in some applications, the state-space representation of the model is more suitable for control - depending on the control objectives. Therefore, in order to make I/O models useful for control, a transformation from I/O models to state-space models is needed. For linear systems, this transition from the I/O framework to the state-space framework can be effectively done using standard algorithms [66]. In particular, the thus obtained models cover all I/O behaviors which can be represented by LTI state-space representation.

However, unlike the linear case, the transformation from I/O representation to state-space representation for the GLSS class is difficult. For example, suppose we would like to obtain LPV state-space models using I/O LPV models. LPV-ARX models with static and affine dependencies with respect to (w.r.t.) the scheduling signal can easily be used for this purpose [20]: there exists an algorithm for converting LPV-ARX model with affine and static dependencies w.r.t. the scheduling signal to LPV state-space models with affine and static dependencies w.r.t. the scheduling signal. Unfortunately from [93], it follows that not all LPV state-space with affine and static dependencies w.r.t. the scheduling signal admit a LPV-ARX model with affine and static dependencies w.r.t. the scheduling signal. In fact, [93] shows that, in general, LPV state-space models with affine and static dependencies w.r.t. the scheduling signal correspond to LPV-ARX models with polynomial and dynamic dependencies w.r.t. the scheduling signal. That is, there exist I/O behaviors which can be described by LPV state-space models with affine and static dependencies w.r.t. the scheduling signal but not by LPV-ARX models with affine and static dependencies w.r.t. the scheduling signal.

We have a similar problem regarding the switched system case: switched ARX models correspond to LSS which are observable for any switching sequence, i.e., for which the initial state can be reconstructed from the I/O behavior for any switching [134]. However, there are I/O behaviors which can be represented by a LSS representation, but which cannot be represented by a LSS which is observable for any switching sequence. That is, not every LSS admits an equivalent representation as a switched ARX.

For these reasons, we concentrate, in this thesis, on state-space models for GLSS. That is, there is a need for methods which directly estimate the state-space models, in particular, GLSS state-space models.

1.3.2 Direct parametric estimation

For the parameters of a state-space framework, there are several parametric identification methods to estimate directly the state-space representation matrices. In these estimation methods, the desired state-space representation is assumed to belong to a family of models defined by parameters, and the task is to estimate the correct parameter values using data. The most common approach to estimate the parameters of the chosen model is the direct Prediction-Error Minimization or the direct PEM method: the model parameters are estimated by minimizing the squared prediction error based on input-output data. For state-space models, this process involves nonlinear optimization with a non-unique global minimum. This non-uniqueness can be found even in LTI systems [66].

Two common approaches for solving this optimization are Gradient-Based methods (GB) and Expectation-Maximization (EM) algorithms. GB methods use search strategies like Newton's method, Levenberg-Marquardt, or similar approaches. In addition, they may incorporate tools like regularization or backtracking for faster convergence. GB methods for LTI state-space identification can be found in [66, 56], where gradient descent methods are used for minimizing the criterion function. For EM methods, the reader can refer to [37, 35]. In a nutshell, the EM method, often used in signal processing, treats the state sequence as missing data. In the expectation step, the state sequence is reconstructed using a Kalman-like filtering approach, and in the maximization step, the model parameters are updated. These steps are repeated until convergence. The LTI EM method is more robust to poor initial estimates compared to GB PEM [139, Table III], but it converges more slowly near the optimum [131]. Both the EM and GB methods can be computationally intensive, and their convergence is highly dependent on a good initial guess of the parameters. A challenge in GB and EM state-space identification is the existence of many isomorphic representations that produce the same input-output behavior, due to the non-uniqueness of the state basis. In order to solve this problem, the Data-Driven Local Coordinate (DDLC) framework was developed [71, 63, 128, 139], which excludes isomorphic models from the search to prevent the optimization from drifting among equivalent solutions.

EM and GB methods were extended to the GLSS framework. For example, GB methods were used for LPV in [63, 64] and EM algorithms were used in [138, 49]. Usually, in LTI systems, subspace techniques are used to find a initialized model and direct PEM are then applied to refine the model. However, this approach is less reliable in the GLSS subclasses case due to the limited development of subspace-based identification methods for GLSS systems. In addition to EM and GB methods, other approaches for GLSS subclasses have been proposed for parametric identification which use machine or deep learning [130, 5, 57].

In this thesis, we don't propose new parametric methods. However, we explore the combination of parametric methods with subspace methods. In addition, the results of this thesis on uniqueness up to isomorphism of minimal GLSS is expected to be helpful for the theoretical justification of the DDLC algorithm for GLSS subclasses.

In the next section, we will explore the alternative to parametric estimation method, namely subspace identification methods, in particular, for the GLSS class.

1.4 Subspace identification methods

By contrast to direct parametric estimation methods, subspace identification methods do not require prior assumptions about the structure of the system. They are based on the success of realization theory, and emerged with it in the late 1980s and early 1990s. Subspace identification methods use the geometric properties of data matrices to extract state-space models. This makes them ideal for handling complex and multivariable systems more efficiently. In this way, they are more practical for real large-scale applications compared to traditional methods. In fact, as will be elaborated later on, the results of this thesis are relevant to subspace methods for GLSS.

1.4.1 Subspace identification methods for LTI models

For LTI systems, there are numerous papers for the identification of the system's representation using subspace identification methods. Among these methods, we can mention the Numerical Subspace State-Space System IDentification (N4SID) [119] and the Multivariable Output-Error State sPace (MOESP) [129]. These methods construct Hankel matrices from the measured data and then use realization theory techniques such as Singular Value Decomposition (SVD) or QR/LQ decomposition to extract the relevant subspaces that summarize the system's dynamics. Other subspace approaches include canonical variate analysis (CVA) [59, 61], and subspace splitting [48]. For a comprehensive overview of the foundations of subspace identification for both deterministic and stochastic linear systems, see [56].

1.4.2 Subspace identification for GLSS models

Later, a multitude of papers have attempted to extend various subspace identification methods for the subclasses of GLSS. For instance, [125, 26, 122, 114] address the LPV subspace identification. For a unified framework, see [21]. For other subclasses, [6, 7] focus on LSS, while [126] examines the piecewise linear systems. Let us briefly and informally present the common GLSS subspace approaches.

PO-MOESP: The LTI past-output multivariable output-error state-space (PO-MOESP) method has been adapted for the LPV case by [123], and for the LSS and piecewise linear cases by [12], assuming a periodic scheduling/switching signal to mitigate dimensionality issues. The PO-MOESP method uses realization theory to obtain the state and output matrices and projection for the remaining SS model matrices.

PBSID: The well-known Predictor Based Subspace IDentification (PBSID) method was applied to LPV subspace identification in [127]. PBSID$_{(opt)}$ was first used for the LPV case in [122], extending [15] from the LTI to the LPV case. In this extension, estimation bias was eliminated, and prediction was considered "optimal" (opt). This method was widely used for about half a decade in the LPV community. However, to our knowledge, there have been no applications of this method to other GLSS subclasses. Nonetheless, this method requires high computational complexity, making full regressors impractical and limiting its use to low-dimensional systems.

CVA: Canonical Variate Analysis (CVA) is a widely used projection-based technique for system identification, particularly for state-space models identification.While extensively applied to LTI systems, it has seen limited direct use in LSS. Nevertheless, several approaches have been developed to enhance LPV subspace identification using CVA [60], in particular, for LPV with affine dependencies on the scheduling signal. This assumption improves computational efficiency, mitigates the curse of dimensionality, and simplifies the mathematical structure. However, the constraints of this assumption have not been thoroughly evaluated through extensive experimental research.

KCCA: In [98], it was demonstrated how to extend the CVA and its closely related concept, the Canonical Correlation Analysis (CCA), to estimate the state-sequence and matrix coefficient functions non-parametrically. This extension is known as Kernelized Canonical Correlation Analysis (KCCA). In KCCA, the projection stage employs kernel methods to handle complex, nonlinear dependencies and recover the state-sequence non-parametrically. The matrix coefficient functions are then estimated non-parametrically using the Least-Squares Support Vector Machine (LS-SVM) framework [97], which builds upon the full-state measurements techniques. Despite its theoretical promise, KCCA has not yet been extensively applied in practice to thoroughly evaluate its properties. In particular, comparisons between KCCA and other subspace identification methods remain unexplored.

Remaining gaps: As mentioned earlier, the aim of this thesis is to develop an efficient and theoretically sound system identification algorithm for stochastic nonlinear systems using the GLSS framework. However, despite the extensive literature reviewed above, several important gaps remain:

- No proof of consistency exists when noise gain is present.
- Another gap involves relaxing the usual observability assumption used in [125, Theorem 4] and [127], which requires at least one of the LTI subsystems constituting the GLSS to be observable.
- If the local modes are not observable, noise gain estimation cannot take place [19].
- Existing methods assume innovation form and uniqueness up to isomorphism. However, it is not clear when, if ever, these conditions are satisfied. In fact, the definition of innovation is unclear.

As a step toward addressing these challenges, we decided to develop and extend realization theory for stochastic GLSS.

Before delving deeper into the scientific challenges, we will first provide an informal, non-technical overview of realization theory in the literature and its role in subspace system identification.

1.5 Realization theory

Realization theory can be attributed to R. E. Kálmán, the founder of modern control theory, who presented the realization problem by considering the system in question as a "black box", and trying to *realize* its input/output map [53, 54]. Later on, the behavioral approach of Willems was introduced as a more formal framework [137, 135, 136]. The literature on realization theory can be divided into two groups: realization theory of linear systems and of nonlinear systems.

1.5.1 Realization theory of linear systems

In time, realization theory has gained popularity for the deterministic LTI case, leading to the development of significant literature. For an overview of the deterministic linear case, see [23]. In a nutshell, realization theory for LTI systems relies on the observation that the input-output maps of linear (but not necessarily LTI) systems can be characterized by a sequence of matrices known as Markov parameters. Markov parameters admit a simple formula in terms of the matrices of an LTI state-space representation of the given input-output map, provided such an LTI system exists. Conversely, one can construct a special infinite matrix from the Markov parameters, the so called Hankel matrix, which has a recurrent structure with the following property: If the input-output map can be represented by an LTI system, then the rank of the Hankel matrix is finite. Moreover, if the rank of the Hankel matrix is finite, then an algorithm can be computed to deliver a minimal linear system from a sub-matrix of the Hankel matrix. This algorithm is called the Kalman-Ho realization algorithm [52, 55]. Theory has also proven that a minimal LTI system is both controllable and observable, and the converse also holds. Furthermore, all minimal LTI systems realizing the same input-output map are isomorphic, meaning that they can be transformed via basis change.

Next, contributors extended the realization theory to the stochastic linear systems without inputs. This extension relies on the observation that the covariances of the output process of a stochastic LTI system are the Markov parameters (entries of the Hankel matrix) of a deterministic LTI system. Also, conversely, any minimal deterministic LTI system whose Markov parameters are the covariances gives rise to a stochastic LTI system whose output is the given output process. It turns out that the representation of the latter stochastic LTI system is in the so called innovation form. In other words, the noise of the system is an innovation process, i.e., the difference between the current output and its best possible linear prediction. Furthermore, it was shown that minimal representations of LTI systems in innovation form for the same output are isomorphic. For a more detailed discussion, see [14, 65]. Additionally, using the so called covariance realization algorithm, the matrices of the stochastic LTI system can be obtained from output covariances. This algorithm is considered a combination of the Kalman-Ho realization algorithm and the Riccati equation or LMI. Later on, the extension to stochastic LTI systems with inputs was presented. For instance, see [65, 56] and the references therein.

The two main applications of realization theory are model reduction and system identification. In control design, having a minimal LTI system that preserves the input-output behavior is considered more efficient, since minimal LTI systems are controllable and observable. In other words, controlling such systems is easier, and observability is ensured, which makes them easier to handle.

Again, realization theory proved its significance when it gave rise to subspace identification, specifically, by extending the Kalman-Ho realization algorithm [118, 56]. We also highlight its role in identifiability analysis [38, 40, 116], and recursive parametric system identification algorithms [40, 77, 70].

1.5.2 Realization theory of nonlinear systems

Motivated by the success of the realization theory in the LTI case, researchers attempted to extend it to more general classes of systems. First, it was shown that a wide class of non-linear systems could be transformed to a minimal dimensional one and minimal systems are reachable and observable. In addition, any two minimal realizations of the same I/O map are isomorphic [111, 45]. Later, rank conditions, similar to the condition of the finite rank of the Hankel matrix, were derived [46, 47, 43, 32, 34]. Hovewer, these general results were not computationally effective, as no minimization or realization algorithms were proposed. Later these results were adapted to more restricted classes of dynamical systems such as deterministic bilinear and LPV systems [108, 107, 42, 44]. In contrast to the general non-linear case, realization theory for bilinear systems did lead to minimization and Ho-Kalman-like realization algorithms [43, 44, 42, 108, 107]. In particular, it was shown that any bilinear system can be transformed to minimal one (equivalent to being observable and reachable). In turn, observability and reachability of bilinear systems can be characterized by the rank of suitable matrices, and the transformation can be formulated as an algorithm. For the more general noiseless LPV case, realization theory was investigated in [115] where the behavioral approach was used. Results on realization theory for discrete-time noiseless LPV systems can be found in [114]. The latter studies deal with affine dependency on the scheduling signal. This type of system was then studied in [85, 86].

Additionally, for noiseless LSS and piecewise-linear systems, realization theory was first introduced by [39] for general hybrid systems and was later further developed for the case of externally generated switching [79, 80, 88, 87, 89]. These results were mostly based on [78]. Moreover, preliminary results for realization theory of JMLS systems were formulated in [90].

However, despite the significant development of realization theory for the GLSS subclasses, many problems remain unsolved. For instance, realization theory for stochastic GLSSs with inputs remains partially developed. Solving these challenges could lead to more efficient realization algorithms and subspace identification techniques for the GLSS class.

1.6 Scientific challenges: problem formulation

Following a quick overview of the vast literature on system identification and realization theory, it is clear that the estimation of GLSS models has been the subject of much research in recent decades. Nevertheless, a number of significant obstacles still exist in spite of these developments. Specifically, when it comes to subspace system identification, there is a noticeable lack of theoretically sound system identification algorithms for stochastic GLSS models. The purpose of this thesis is to fill in these gaps.

In order to accomplish this, the following challenges need to be addressed:

Challenge 1 (Decomposition of stochastic GLSS models)**.** *If we want to follow the approach adopted in stochastic realization theory of LTI systems, then stochastic GLSS models must be broken down into stochastic and deterministic parts: the stochastic part is dependent only on the noise and the deterministic part is only dependent on the inputs. This would allow the deterministic component to be addressed by applying current findings from deterministic realization theory [85, 113, 86], and the stochastic part could be addressed using realization theory of generalized bilinear systems [91]. For LPV systems, such a decomposition was suggested in [72], but it was not generalized for GLSS models.*

Challenge 2 (Algebraic characterization for minimality and innovation form)**.** *It is generally known that in order to guarantee the uniqueness of the system's representation for LTI systems, minimal representation in innovation form is a prerequisite. The notion of minimal innovation representations for GLSS models has only been explored for certain subclasses of autonomous GLSS, such as General Bilinear Systems (GBS) and stochastic autonomous GLSS models [91]. However, there is still no coverage in the literature on how to guarantee minimality, uniqueness and innovation form for GLSS models without inputs using only conditions on the matrices of the* *system's state-space representation. Establishing such guarantees for autonomous GLSS (without inputs) is considered to be a first step toward guaranteeing minimality, uniqueness and innovation form for GLSS models with inputs. Uniqueness of state-space representation (up to isomorphism) is important for the following reasons:*

a) Identifiability: *Uniqueness (up to isomorphism) is crucial for identifiability of GLSS models. In general, parametric system identification methods rely on estimating a fixed set of parameters to describe the system. If a model is not identifiable, then the parametric methods might not converge to the global minimum. Hence, the parameters might converge to different values, depending on the initial conditions or noise, leading to unreliable results.*

b) Control and Prediction: *Even for non-parametric methods, uniqueness (up to isomorphism) is crucial for defining consistency in a way that is useful for control. Intuitively, we want the estimated model to reproduce the behavior of the data-generating system for* ***all*** *inputs and switching signals. Standard notions of consistency ensure that the estimated model reproduces the output of the true system for the inputs and switching signals used in the identification experiment. Without uniqueness, multiple different models could fit the data used for the system identification experiment, making it impossible to determine which one correctly represents the actual system behavior. However, if the state-space representation that produces the* ***same*** *output for these inputs and switching signals is unique, then classical consistency results automatically imply that a model estimated by a consistent algorithm will be isomorphic to the true system. Consequently, for* ***any*** *input and switching signals, the two systems will generate the same output, at least as the number of data points tends to infinity.*

Challenge 3 (Minimality of GLSS in innovation form with inputs: characterization, existence, uniqueness)**.** *It is known that the minimality of a stochastic system can be established by proving the minimality of its deterministic counterpart. More precisely, covariances of the output of stochastic systems are impulse responses of a deterministic systems, and the matrices of these two systems are directly related and can be computed by a transformation that preserves minimality, and can be realized by an algorithm. This was extended to certain subclasses of GLSS without inputs [91]. However, this has not been thoroughly explored for GLSS models with inputs. Extending this principle requires formalizing the relationship between the deterministic and stochastic GLSS models. As argued above in Challenge 2, this is very important for control.*

Challenge 4 (Development of a covariance realization algorithm)**.** *Following the approach used in subspace identification for LTI systems, it is desirable to formulate a covariance realization algorithm that generates a minimal GLSS model in innovation form. This algorithm is essential for developing a subspace method that yields a minimal GLSS in innovation form. There is a substantial gap in the literature for GLSS models since, although a realization approach was presented in [72] for LPV systems, it does not ensure minimality.*

Challenge 5 (Consistency of realization-based system identification algorithms)**.** *For GLSS models, the identification algorithms which are based on realization theory, needs to be statistically* *consistent. In other words, the identified model should converge to a realization of the outputs as the number of training data points approaches infinity. To our knowledge, there are no consistency results for most of the existing realization-based system identification algorithms.*

Challenge 6 (Computational efficiency and observation assumptions)**.** *Naive versions of subspace identification of subclasses of GLSS often require the use of Hankel matrices whose size grows fast with the state-space dimension of the underlying system. Working with the large Hankel matrix can be very challenging, since the computation of the latter can be memory-intensive, making it impossible to run the algorithm using usual tools. In order to remedy this situation, most subspace identification algorithms for subclasses of GLSS assume that there is a dominant discrete state by which the corresponding LTI system is observable, and the contribution of other discrete states to the output is small. This allows for the use of Hankel matrices similar to LTI ones, and whose dimension grows linearly with the dimension of the system. However, this assumption is very restrictive. In particular, it is easy to generate examples of input-output behaviors for which there is no GLSS that satisfies this assumption.*[2] *Another assumption that is imposed to deal with the curse of dimensionality, is the assumption that the effect of the switching is absent after a certain number of time-steps. While asymptotically this assumption may be true, it is unclear how many data points we need for it to hold and it may lead, potentially, to biased estimates. Moreover, it is not clear how to enforce this assumption in the estimated model. In turn, this makes the consistency analysis more difficult: the assumed and estimated models belong to different classes.*

This thesis seeks to overcome these obstacles by creating a system identification technique for GLSS models with inputs, based on a statistically consistent and computationally efficient realization theory algorithm.

1.7 Contributions and outline of the thesis

This thesis is divided into six chapters, including the introduction and the conclusion chapters, i.e., Chapter 1 and Chapter 6.

The basic fundamentals that lay the groundwork for the in-depth study and applications covered in later chapters are presented in Chapter 2. In particular, this chapter presents the general model class examined in this thesis, as well as its numerous subclasses, both mathematically and formally. The class of stationary GLSS, which is the class of systems studied in this thesis, is formally introduced at the end of Chapter 2, along with technical definitions and key characteristics of the signals involved. A summary of realization theory for autonomous stationary GLSS models is also presented in Chapter 2. Discussing autonomous GLSS is justified by the fact that understanding

[2] Indeed, it is sufficient to consider a GLSS which is observable and reachable, but none of its LTI subsystems is observable and reachable: such GLSS is then minimal and let us call it $\mathcal{G}_1$. Then, assume there was a GLSS equivalent to $\mathcal{G}_1$, that we will call it $\mathcal{G}_2$, and that $\mathcal{G}_2$ contains a minimal LTI system. Then $\mathcal{G}_2$ is observable and reachable and hence minimal. Thus, $\mathcal{G}_2$ would be isomorphic to the $\mathcal{G}_1$, and hence, $\mathcal{G}_1$ would have a minimal LTI subsystem. However, we assumed the opposite. In other words, we have a contradiction.

the autonomous case is the foundational step toward developing stochastic realization theory for GLSS models with inputs. Building on this foundation, each of the subsequent chapters discusses particular contributions to the realization theory of the systems under study.

Our initial theoretical contribution to the GLSS model class is presented in Chapter 3, which is based on [101]. We demonstrate that the output of GLSS systems can be divided into stochastic and deterministic components in order to solve **Challenge 1**. We define sufficient criteria for GLSS state-space representations to be minimal and unique, and we demonstrate their existence in innovation form using this decomposition. The foundation for creating effective system identification algorithms is laid by this effort. However, it is essential to first examine the stochastic autonomous case, particularly the stochastic component of the GLSS decomposition mentioned above, in order to completely accomplish this purpose. The latter is the central theme of Chapter 4.

In Chapter 4, we focus specifically on autonomous stochastic GLSS. We present a new algebraic characterization for such systems to be minimal and in innovation form, in order to overcome **Challenge 2**. This characterization depends solely on the system's representation matrices. This contribution is important for understanding realization theory of GLSS with inputs, and also for handling some problems encountered in the system identification process which will be further discussed in the respective chapter. Chapter 4 is based on [99].

Chapter 5 addresses the four remaining challenges. First, Chapter 5 addresses **Challenge 3** and **Challenge 4** by proposing a covariance realization algorithm and proving that it returns a GLSS which is minimal, unique (up to isomorphism) and in innovation form. In addition, the proposed covariance algorithm uses a reduced size of the Hankel matrix, by applying the technique of "nice selection". Second, in Chapter 5, we formulate a system identification algorithm based on the previously-mentioned covariance realization theory algorithm. This system identification algorithm *addresses* **Challenge 5** and **Challenge 6**, as it has the following properties:

- **Small Hankel matrix:** The algorithm uses reduced-size Hankel matrices, which helps avoid the curse of dimensionality and ensures computational efficiency.

- **No state-observation assumption:** There is no assumption regarding the observability of the underlying LTI systems.

- **No decaying effect of the switching:** All past and current switching values are treated with their respective relevance to the output. In other words, we do not assume that the effect of the switching on the state/output evolution decreases with time.

- **Uniqueness of minimality in innovation form:** The algorithm ensures that a minimal GLSS model in innovation form is obtained, as uniqueness (up to isomorphism) is guaranteed.

- **Consistency:** We can show that the algorithm is asymptotically consistent. That is, in the limit, the identified model will be isomorphic to the data generator if the latter is also minimal. Consequently, the model will produce the same response as the data generator to *any* input and switching signals.

Finally, we combine the system identification algorithm with a gradient-descent search in order to improve the algorithm's accuracy.[3] Chapter 5 is based on [100].

Chapter 6 wraps up the thesis by revisisting its key points again, combining the theoretical developments, and considering their wider ramifications. To highlight their influence on the field of realization theory and system identification, the main findings are reviewed briefly.

By addressing significant gaps in realization theory and providing a rigorous framework for system identification, this thesis aims at filling the gap in the theoratical foundations of system identification of GLSS. By doing this, the work contributes to theoretical research and has the potential to be useful for practical applications, offering knowledge and resources that could benefit the broader systems and control community.

[3]Note that the only drawback to this system identification algorithm is that it is restricted to using i.i.d. signals for the identification procedure, despite its capability of returning a valid model for any switching and inputs signal. This can be limiting for some applications. Nonetheless, we believe that such system identification algorithm is a first step toward developing more general algorithms.

Chapter 2

Preliminaries

2.1 Introduction

In this chapter, we will formally define the class of general mathematical model that we will be working with throughout the thesis. This class of models includes subclasses of Jump-Markov, Piecewise-linear and LPV systems. Therefore, this class of models is called "Generalized Linear Switched System" (GLSS), because it switches among potentially infinitely many linear systems. In addition, we will present some technical definitions handling the properties of the designated signals. The definitions of this chapter will be used in the following chapters of the thesis. Throughout the following chapters, we will use the notations and terminologies defined in this chapter. However, if needed, we will recall some of the definitions or invite the reader to revise this chapter.

The outline of the chapter will be as follows. In Section 2.2, we introduce informally the class of systems studied in this thesis. The definition in Section 2.2 is too general and too informal for our purposes. In order to refine it, we need to define some notation and introduce the necessary terminology. In particular, we need to restrict the class of switching signals. Furthermore, we should restrict attention to processes with suitable stationary properties. To this end, in Section 2.3, we define the necessary notations. In particular, we define the notation used in the thesis regarding probability theory and formal languages in Subsection 2.3.2. Section 2.4 introduces the necessary properties that the processes of GLSSs should have. Specifically, in Subsection 2.4.1, we define the class of admissible switching signals, which will be used to formally define the class of switched systems used in the thesis. In Subsection 2.4.2, we introduce an extension of the notion of stationarity, called Zero Mean Wide Sense Stationary with respect to the switching sequence (ZMWSSI). This extension is necessary in order to capture properly the interplay between the switching signals and the state, the output and the noise processes. Properties of the latter processes will also be defined, namely, the property Square Integrable with respect to Inputs (SII) and the property white noise with respect to the switching signal. Finally, in Section 2.5, we define the class of stationary GLSSs. Informally, the systems belonging to the latter class are such that they are quadratically stable, and their state, input, output and noise processes are stationary in the sense defined in Subsection 2.4.2, and their switching signals are admissible in the sense defined in Subsection 2.4.1.

2.2 Generalized Linear Switched System (GLSS)

In this section, we define the class of systems studied in this thesis. All the stochastic processes in this thesis are discrete-time ones defined over the time-axis $\mathbb{Z}$ of the set of integers: a stochastic process $\mathbf{r}$ is a collection of random variables $\{\mathbf{r}(t)\}_{t\in\mathbb{Z}}$ taking values in some set X. Note that we use bold letters to denote random variables and stochastic processes. A discrete-time stochastic *Generalized Linear Switched System (abbreviated as GLSS)* is a system of the form

$$\mathcal{S}\begin{cases}\mathbf{x}(t+1) = \sum_{i=1}^{n_\mu}(A_i\mathbf{x}(t) + B_i\mathbf{u}(t) + K_i\mathbf{v}(t))\boldsymbol{\mu}_i(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + F\mathbf{v}(t), \ \ t\in\mathbb{Z}\end{cases} \tag{2.1}$$

where $A_i \in \mathbb{R}^{n_x\times n_x}$, $B_i \in \mathbb{R}^{n_x\times n_u}$, $K_i \in \mathbb{R}^{n_x\times n_n}$, $i = 1,\dots,n_\mu$, $C \in \mathbb{R}^{n_y\times n_x}$ and $D \in \mathbb{R}^{n_y\times n_u}$, $F \in \mathbb{R}^{n_y\times n_n}$ are constant matrices. All the processes present in the model above, are stochastic. In particular, $\mathbf{x}$ is the state process, $\mathbf{u}$ is the input process, $\mathbf{y}$ is the output process, $\mathbf{v}$ is the noise process and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_{n_\mu})^T$ is the switching process, taking values respectively in $\mathbb{R}^{n_x}, \mathbb{R}^{n_u}, \mathbb{R}^{n_y}, \mathbb{R}^{n_n}, \mathbb{R}^{n_\mu}$. Intuitively, (2.1) is a generalization of linear switched systems to the case of infinitely many discrete modes.

GLSSs cover a variety of system classes. These classes are defined depending on the choice of the switching signal. This means that there exist several possible choice of the switching signal $\boldsymbol{\mu}$, and each choice correspond to a well known class of systems. For instance:

- Assume that $\boldsymbol{\mu}(t)$ takes values in the set of unit vectors, i.e., $\boldsymbol{\mu}_i(t) \in \{0,1\}$, $i = 1,\dots,n_\mu$,

  $$\sum_{i=1}^{n_\mu}\boldsymbol{\mu}_i(t) = 1.$$

  and define the process $\boldsymbol{\theta}(t)$ taking values in $\{1,\dots,n_\mu\}$ as $\boldsymbol{\theta}(t) = i \iff \boldsymbol{\mu}_i(t) = 1$. Then (2.1) is a *switched system [110] and can be written as follows:*

  $$\mathcal{S}\begin{cases}\mathbf{x}(t+1) = A_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + B_{\boldsymbol{\theta}(t)}\mathbf{u}(t) + K_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + F\mathbf{v}(t)\end{cases} \tag{2.2}$$

- If, in addition, the process $\theta(t)$ is a Markov chain, then (2.2) is a jump-Markov system [18].

- If $\boldsymbol{\mu}$ takes values from a possibly infinite set and $\boldsymbol{\mu}_1 = 1$, then (2.1) could be viewed as a subclass of LPV systems [113] with an *affine dependence* on scheduling, and $\tilde{\boldsymbol{\mu}}(t) =$

$(\boldsymbol{\mu}_2(t), \ldots, \boldsymbol{\mu}_{n_\mu}(t))^T$ corresponds to the scheduling signal. In other words, (2.1) can be expressed as an LPV system:

$$\mathcal{S}\begin{cases}\mathbf{x}(t+1) = A(\tilde{\boldsymbol{\mu}}(t))\mathbf{x}(t) + B(\tilde{\boldsymbol{\mu}}(t))\mathbf{u}(t) + K(\tilde{\boldsymbol{\mu}}(t))\mathbf{v}(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + F\mathbf{v}(t)\end{cases} \tag{2.3}$$

where the matrices of the state equation can be written as follows:

$$A(\tilde{\boldsymbol{\mu}}(t)) = A_1 + \sum_{i=2}^{n_\mu} A_i \boldsymbol{\mu}_i(t)$$
$$B(\tilde{\boldsymbol{\mu}}(t)) = B_1 + \sum_{i=2}^{n_\mu} B_i \boldsymbol{\mu}_i(t)$$
$$K(\tilde{\boldsymbol{\mu}}(t)) = K_1 + \sum_{i=2}^{n_\mu} K_i \boldsymbol{\mu}_i(t)$$

Remark 2.1. *Note that the references in what follows use the terminologies of LPV, LSS, GBS or piecewise-linear systems. The term GLSS is used in this thesis, in order to cover all those system classes.*

Example 2.1 (Hybrid Electric Vehicle)**.** *A practical application of switched systems is in Hybrid Electric Vehicles (HEVs), which operate by switching between different power sources—electric motor, combustion engine, or both—depending on driving conditions. These switching behaviors can be modeled using discrete-time* linear switched systems.

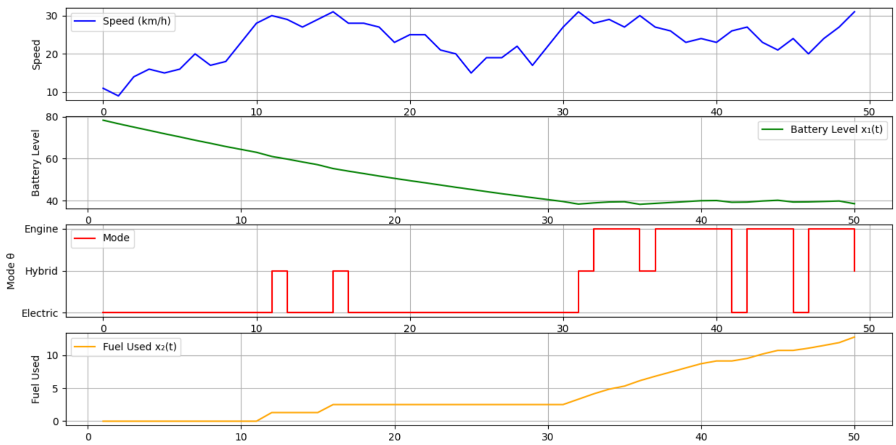

Figure 2.1: HEV switching simulation

System variables: The system evolves according to:

$$\mathbf{x}(t+1) = A_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + B_{\boldsymbol{\theta}(t)}\mathbf{u}(t) + K_{\boldsymbol{\theta}(t)}\mathbf{v}(t)$$

where:

- $\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} \in \mathbb{R}^2$ is the state vector where $\mathbf{x}_1(t)$ is the battery charge level (%), and $\mathbf{x}_2(t)$ is the cumulative fuel used (liters).
- $\mathbf{u}(t) \in \mathbb{R}$ is the control input (e.g., power request).
- $\mathbf{v}(t) \in \mathbb{R}$ is an external disturbance (e.g., terrain slope).
- $\boldsymbol{\theta}(t) \in \{1, 2, 3\}$ is the active mode at time t.

Figure 2.1 illustrates a switching simulation.
Modes and dynamics:
– Mode $\boldsymbol{\theta} = 1$: **Electric mode.**
Used at low speeds when the battery is sufficiently charged. The vehicle is powered entirely by the electric motor:

$$A_1 = \begin{bmatrix} 0.98 & 0.00 \\ 0.00 & 1.00 \end{bmatrix}, \quad B_1 = \begin{bmatrix} -0.10 \\ 0.00 \end{bmatrix}, \quad K_1 = \begin{bmatrix} -0.05 \\ 0.00 \end{bmatrix}.$$

Battery charge decreases; no fuel is consumed.
– Mode $\boldsymbol{\theta} = 2$: **Hybrid mode.**
Used at moderate speeds. Both electric motor and engine contribute:

$$A_2 = \begin{bmatrix} 0.97 & 0.00 \\ 0.02 & 1.00 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -0.08 \\ 0.05 \end{bmatrix}, \quad K_2 = \begin{bmatrix} -0.06 \\ 0.03 \end{bmatrix}.$$

Battery and fuel are both consumed.
– Mode $\boldsymbol{\theta} = 3$: **Engine mode.**
Used at high speeds or when battery charge is low:

$$A_3 = \begin{bmatrix} 0.995 & 0.00 \\ 0.01 & 0.98 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 0.5 \\ 0.32 \end{bmatrix}, \quad K_3 = \begin{bmatrix} 0.03 \\ 0.04 \end{bmatrix}.$$

Fuel usage increases, and the battery may slightly recharge.

Switching logic: The mode $\theta(t)$ is chosen according to the following rule:

$$\theta(t) = \begin{cases} 1, & \text{if speed} < 30 \text{ km/h and battery charge} \geq 40, \\ 2, & \text{if } 30 \leq \text{speed} \leq 70 \text{ km/h}, \\ 3, & \text{if speed} > 70 \text{ km/h or battery charge} < 40. \end{cases}$$

This logic prioritizes electric propulsion at low speeds when the battery is charged, hybrid power at intermediate speeds, and engine power at high speeds or low battery levels.

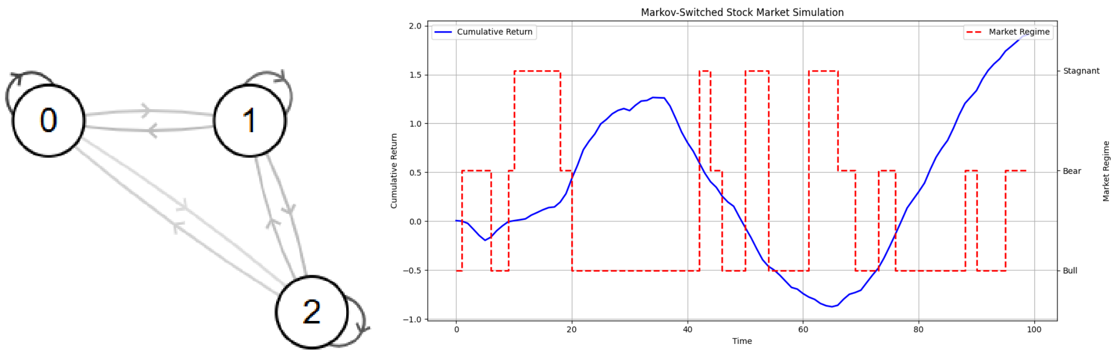

Figure 2.2: Stock market regime switching

Example 2.2 (Stock Market Regime Switching)**.** *A market regime is a state of the financial market characterized by distinct statistical properties (such as return trends and volatility). At each time step t, the system evolves according to:*

$$\mathbf{x}(t+1) = A_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + B_{\boldsymbol{\theta}(t)}\mathbf{u}(t) + K_{\boldsymbol{\theta}(t)}\mathbf{v}(t),$$

where $\mathbf{x}(t) \in \mathbb{R}$ is the latent market state (interpreted as log-return or price signal), $\mathbf{u}(t), \mathbf{v}(t) \sim \mathcal{N}(0,1)$ are independent Gaussian input and noise processes, A_i, B_i, and K_i are scalar parameters specific to each regime $i \in \{0,1,2\}$. The three regimes we modeled are:

- *Bull Market $(\boldsymbol{\theta} = 0)$: High returns, low volatility $(A_0 = 1.015, B_0 = 0.04, K_0 = 0.01)$.*

- *Bear Market $(\boldsymbol{\theta} = 1)$: Negative returns, high volatility $(A_1 = 0.995, B_1 = -0.01, K_1 = 0.03)$.*

- *Stagnant Market $(\boldsymbol{\theta} = 2)$: Low or near-zero returns, moderate volatility $(A_2 = 1.000, B_2 = 0.005, K_2 = 0.015)$.*

Let us consider that the regime-switching process $\boldsymbol{\theta}(t)$ follows a Markov chain, with transition probabilities matrix:

$$P = \begin{bmatrix} 0.85 & 0.1 & 0.05 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.15 & 0.65 \end{bmatrix}$$

where rows represent the current state $\boldsymbol{\theta}(t)$, columns represent the probability of switching to the next state $\boldsymbol{\theta}(t+1)$. For instance, if the market is currently in a bull phase, there's an 85% probability it remains bullish, a 10% probability it switches to a bear market, and a 5% probability it becomes stagnant. Figure 2.2 illustrates the relationship between the states of the market regime (represented as nodes) along with a numerical simulation.

Note that, for some applications, the input process of the GLSS may be absent. We call such systems *autonomous stochastic Generalized Linear Switched Systems (aGLSS)*. That is, an aGLSS is a system described by

$$\mathcal{S}_a \begin{cases} \mathbf{x}(t+1) = \sum_{i=1}^{n_\mu} (A_i \mathbf{x}(t) + K_i \mathbf{v}(t)) \boldsymbol{\mu}_i(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + F\mathbf{v}(t) \end{cases} \tag{2.4}$$

where $A_i \in \mathbb{R}^{n_x \times n_x}, K_i \in \mathbb{R}^{n_x \times n_v}$, $C \in \mathbb{R}^{n_y \times n_x}$, $F \in \mathbb{R}^{n_y \times n_v}$, and $\mathbf{x}$ is the state process, $\boldsymbol{\mu} = \left[\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{n_\mu}\right]^T$ is the switching process, $\mathbf{v}$ is the noise process and $\mathbf{y}$ is the output process.

Also note that all the involved processes of (2.1) and (2.4) are defined for both negative and positive time. This may create technical problems for the existence of a solution and the role of initial state. In this thesis, we will circumvent this problem by considering GLSS which are mean-square stable in a suitable sense and state process of which is stationary, i.e. stationary GLSS and autonomous stationary GLSS. These assumptions guarantee that a solution exist for both negative and positive times.

Remark 2.2 (Negative time instants)**.** *More precisely, the motivation behind considering both negative and positive times for the processes is to avoid unnecessary technical complications related to non-stationarity of the processes involved and the choice of the initial state. As it was already mentioned, we assume that the GLSSs involved are stable. Hence, after a sufficiently long time, the effect of the initial state effectively disappears, i.e., the system approaches its stationary regime. This means that, for system identification, and to some extent for control, only the stationary behavior is of interest. Our definition captures exactly the stationary regime. The assumption that the time is negative means that the system was started infinitely long time ago at some initial state, and the state/output trajectory represents the stationary behavior of the system. This assumption is standard in stochastic systems theory [65].*

In order to deal with this class of GLSSs, we introduce some terminology and notations in the following section.

2.3 Terminology and notations

In this section, we will present certain notions from linear algebra, probability theory and formal languages.

2.3.1 Linear algebra

We recall that a Schur matrix is a square matrix all eigenvalues of which are inside the complex unit disk. We denote by $\otimes$ the Kronecker product of matrices, and by I_n then $n \times n$ identity matrices.

2.3.2 Probability theory

Throughout the thesis, we use the standard terminology of probability theory [11], unless it is mentioned otherwise. All the random variables and stochastic processes are understood w.r.t. to a fixed probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-algebra over the sample space Ω, and P is the probability measure. In other words, $\mathcal{F}$ is a collection of subsets of Ω, that includes Ω itself. It is closed under complement, closed under countable unions and closed under countable intersections. For two σ-algebras $\mathcal{F}_i$, $i = 1, 2$, $\mathcal{F}_1 \vee \mathcal{F}_2$ denotes the smallest σ-algebra containing the σ-algebras $\mathcal{F}_1, \mathcal{F}_2$. The expected value of a random variable $\mathbf{r}$ is denoted by $E[\mathbf{r}]$ and conditional expectation w.r.t. σ-algebra $\mathcal{F}$ is denoted by $E\left[\mathbf{r} \mid \mathcal{F}\right]$. The stochastic processes in this thesis are discrete-time ones defined over the time-axis $\mathbb{Z}$.

After identifying the standard terminology of the probability theory, we are now ready to introduce the following notation.

Notation 2.1 (Orthogonal projection E_l)**.** *Recall that the set of square integrable random variables taking values in $\mathbb{R}$, forms a Hilbert-space with the scalar product defined as $< \mathbf{z}_1, \mathbf{z}_2 >= E[\mathbf{z}_1 \mathbf{z}_2]$. We denote this Hilbert-space by $\mathcal{H}_1$. Let $\mathbf{z}$ be a square integrable* vector-valued *random variable taking its values in $\mathbb{R}^k$. Let M be a closed subspace of $\mathcal{H}_1$. By the orthogonal projection of $\mathbf{z}$ onto the subspace M,* denoted by $E_l[\mathbf{z} \mid M]$*, we mean the vector-valued square-integrable random variable $\mathbf{z}^* = \left[\mathbf{z}_1^*, \ldots, \mathbf{z}_k^*\right]^T$ such that $\mathbf{z}_i^* \in M$ is the orthogonal projection of the ith coordinate $\mathbf{z}_i$ of $\mathbf{z}$ onto M, as it is usually defined for Hilbert spaces. Let $\mathfrak{S}$ be a subset of square integrable random variables in $\mathbb{R}^p$ for some integer p, and suppose that M is generated by the coordinates of the elements of $\mathfrak{S}$, i.e. M is the smallest (with respect to set inclusion) closed subspace of $\mathcal{H}_1$ which contains the set $\{\alpha^T s \mid s \in \mathfrak{S}, \alpha \in \mathbb{R}^p\}$. Then instead of $E_l[\mathbf{z} \mid M]$ we use $E_l[\mathbf{z} \mid \mathfrak{S}]$.*

Intuitively, $E_l[\mathbf{z} \mid \mathfrak{S}]$ is *the best (minimal variance) linear prediction* of $\mathbf{z}$ based on the elements of $\mathfrak{S}$. For example, consider the case where M is finite, i.e., $M = \{\mathbf{m}_1, \ldots, \mathbf{m}_n\}$. Let us define

$$T_M = E\left[\begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_1^T & \cdots & \mathbf{m}_n^T \end{bmatrix}\right] = \begin{bmatrix} E\left[\mathbf{m}_1 \mathbf{m}_1^T\right] & \cdots & E\left[\mathbf{m}_1 \mathbf{m}_n^T\right] \\ \vdots & & \vdots \\ E\left[\mathbf{m}_n \mathbf{m}_1^T\right] & \cdots & E\left[\mathbf{m}_n \mathbf{m}_n^T\right] \end{bmatrix},$$

$$\Lambda_{\mathbf{z}M} = \begin{bmatrix} E\left[\mathbf{z}\, \mathbf{m}_1^T\right] \\ \vdots \\ E\left[\mathbf{z}\, \mathbf{m}_n^T\right] \end{bmatrix},$$

then $E_l[\mathbf{z} \mid M] = \Lambda_{\mathbf{z}M} T_M^{-1} \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_n \end{bmatrix}$. Now we can write $\mathbf{z} = E_l[\mathbf{z} \mid M] + \mathbf{e}$ where $\mathbf{e}$ and M are orthogonal and $\mathbf{e}$ is the projection error. In other words:

$$E_l[\mathbf{z} \mid M] = \underset{\alpha \in \mathrm{Span}\{M\}}{\arg\min}\; E\left[||\mathbf{z} - \alpha||^2\right].$$

If $(\mathbf{z}, M)$ are jointly Gaussian, then $E_l[\mathbf{z} \mid M] = E_l[\mathbf{z} \mid \mathcal{F}_M]$, where $\mathcal{F}_M$ is the σ-algebra generated by M.

2.3.3 Finite sequences

Here, we recall a few standard notions from automata theory used in the rest of the thesis.

Notation 2.2 (Sequences over Σ)**.** *An alphabet is a finite set of symbols in automata. Alphabets are a set of symbols used to construct a language. In what follows,*

$$\Sigma = \{1, \ldots, n_\mu\}$$

denotes the finite alphabet, where $n_\mu > 0$. A non empty word *over Σ is a finite sequence of letters (elements) of Σ, i.e., $w = \sigma_1\sigma_2 \cdots \sigma_k$, for some $k \in \mathbb{N}$, $k > 0$, $\sigma_1, \sigma_2, \ldots, \sigma_k \in \Sigma$. The length of the word w is denoted by $|w| = k$. The set of* all *nonempty words is denoted by Σ^+. We denote the* empty word *by ϵ and by convention $|\epsilon| = 0$. Let $\Sigma^* = \{\epsilon\} \cup \Sigma^+$. The concatenation of two nonempty words $v = a_1a_2 \cdots a_m$ and $w = b_1b_2 \cdots b_n$ is defined as $vw = a_1 \cdots a_m b_1 \cdots b_n$ for some $m, n > 0$. By convention $v\epsilon = \epsilon v = v$ for all $v \in \Sigma^*$.*

Here is an example to illustrated the previous notation introduced from automata theory.

Example 2.3. *For $n_\mu = 2$, $\Sigma = \{1, 2\}$, $\Sigma^* = \{\epsilon, 1, 2, 11, 12, 21, 22, 111, \ldots\}$, for the word $w = 111 \in \Sigma^*$, $|w| = 3$ and $w\epsilon = 111\epsilon = 111$. If $v = 12211 \in \Sigma^*$, then $wv = 11112211$ and $|wv| = 8$.*

Notation 2.3 (Notation for matrices)**.** *We denote by I_n the $n \times n$ identity matrix. Consider $n \times n$ square matrices $\{A_\sigma\}_{\sigma \in \Sigma}$. For any $w = \sigma_1\sigma_2 \cdots \sigma_k \in \Sigma^+$, $k > 0$ and $\sigma_1, \ldots, \sigma_k \in \Sigma$, we define $A_w = A_{\sigma_k} A_{\sigma_{k-1}} \cdots A_{\sigma_1}$. For an empty word ϵ, set $A_\epsilon = I_n$.*

Example 2.4. *For instance, let us consider the following three matrices:*

$$A_1 = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

In this case, we will consider the sequence $\Sigma = \{1, 2, 3\}$ where $n_\mu = 3$. For a word $w = 123 \in \Sigma^* = \{\epsilon, 1, 11, 12, 13, 111, 121, 121, \ldots\}$ we will have

$$A_w = A_3A_2A_1 = \begin{bmatrix} 7 & 12 \\ 5 & 9 \end{bmatrix}.$$

We can also realize from Notation 2.2 and Notation 2.3 that, for example, for the word $v = 12 \in \Sigma^*$, $A_wA_v = A_3A_2A_1A_2A_1 = A_{vw}$.

2.4 Process properties

This section recalls from [91] the concept of admissible processes and the concept of zero mean wides sense stationary processes with respect to the switching. These definitions of properties concerning the processes are used in the following chapters. But before, let us recall from the literature on probability theory the following definition.

Definition 2.1 (Wide-Sense Stationary (WSS)). *A random process is called weak-sense stationary or wide-sense stationary (WSS) if its mean and its autocovariances do not change if time is shifted. More precisely, if* $\mathbf{r}$ *is a WSS process, then for all* t_1, $t_2 \in \mathbb{Z}$ *and for all* $\delta \in \mathbb{Z}$,

$$E\left[\mathbf{r}(t_1)\right] = E\left[\mathbf{r}(t_2)\right],$$

$$E\left[\mathbf{r}(t_1)\mathbf{r}(t_2)\right] = E\left[\mathbf{r}(t_1+\delta)\mathbf{r}(t_2+\delta)\right].$$

This is one of the most common forms of stationarity that is widely used in practice and literature. One can easily see that the autocorrelation of the process at two different instant depends only on the time difference (the time shift).

In contrast, a Strict Stationarity (SS) is a stronger condition where all the joint distributions of the process are invariant under time shifts. For strict stationarity, the entire distribution of the random process does not change over time. So, a process can be wide-sense stationary without being strictly stationary if it satisfies the conditions on the mean, variance, and autocorrelation, but does not exhibit time-invariant higher-order moments or joint distributions.

2.4.1 Admissible process

With the notations above, we can identify the process $\boldsymbol{\mu}$ with the collection $\{\boldsymbol{\mu}_\sigma\}_{\sigma\in\Sigma}$ of its components, in order to introduce the definition of an admissible process. To this end, let us consider that for every word $w \in \Sigma^+$ where $w = \sigma_1\sigma_2\cdots\sigma_k$, $k \geq 1$, $\sigma_1,\ldots,\sigma_k \in \Sigma$, we define the process $\boldsymbol{\mu}_w$ as follows:

$$\boldsymbol{\mu}_w(t) = \boldsymbol{\mu}_{\sigma_1}(t-k+1)\boldsymbol{\mu}_{\sigma_2}(t-k+2)\cdots\boldsymbol{\mu}_{\sigma_k}(t). \tag{2.5}$$

For an empty word $w = \epsilon$, we set $\boldsymbol{\mu}_\epsilon(t) = 1$. Recall that the processes are defined over the time-axis $\mathbb{Z}$. Let us present the following example to illustrate what we just introduced:

Example 2.5. *We will consider the alphabet finite set* $\Sigma = \{1,2\}$, *where* $n_\mu = 2$. *Then for a word* $w = 121 \in \Sigma^*$, *where* $|w| = k = 3$, *we will get, for an instance* $t \in \mathbb{Z}$,

$$\boldsymbol{\mu}_w(t) = \boldsymbol{\mu}_{12}(t-1)\boldsymbol{\mu}_1(t) = \boldsymbol{\mu}_1(t-2)\boldsymbol{\mu}_2(t-1)\boldsymbol{\mu}_1(t).$$

Notice that, for a word $w \in \Sigma^*$, the product $\boldsymbol{\mu}_w(t)$ at an instance $t \in \mathbb{Z}$ depends on some past values of the switching. Now we are ready to define what we mean by an admissible swtching.

Definition 2.2 (Admissible process [91]). *The switching process $\boldsymbol{\mu}$ is called* admissible, *if the following holds:*

1. *There exists a set $\mathcal{E} \subseteq \Sigma \times \Sigma$ of pairs of letter from Σ, such that:*

– *for every letter $\sigma \in \Sigma$, there exists a letter $\sigma' \in \Sigma$ such that $(\sigma, \sigma') \in \mathcal{E}$.*

– *Define the set of admissible words L as the set of all words $w \in \Sigma^+$ such that*

$$w = \sigma_1 \cdots \sigma_k \in \Sigma^+,$$

for $\sigma_1, \dots, \sigma_k \in \Sigma$, $k > 0$ and $(\sigma_i, \sigma_{i+1}) \in \mathcal{E}$ for all $i = 1, \dots, k-1$. Then for all $w \in \Sigma^+$, such that $w \notin L$, it holds that $\boldsymbol{\mu}_w = 0$.

2. *Denote by $\mathscr{F}_t^{\boldsymbol{\mu},-}$ the σ-algebra generated by the random variables $\{\boldsymbol{\mu}(k)\}_{k<t}$. There exists positive numbers $\{p_\sigma\}_{\sigma\in\Sigma}$ such that for any $w, v \in \Sigma^+, \sigma, \sigma^{'} \in \Sigma$, $t \in \mathbb{Z}$:*

$$E[\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{v\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}] = \begin{cases} p_\sigma \boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1) & \text{if } \sigma = \sigma^{'} \text{and } w\sigma, v\sigma \in L \\ 0 & \text{otherwise} \end{cases}$$

$$E[\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}] = \begin{cases} p_\sigma \boldsymbol{\mu}_w(t-1) & \text{if } \sigma = \sigma^{'} \text{ and } w\sigma \in L \\ 0 & \text{otherwise} \end{cases}$$

$$E[\boldsymbol{\mu}_{\sigma}(t)\boldsymbol{\mu}_{\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}] = 0 \quad \text{if } \sigma \neq \sigma^{'}$$

3. *There exist real numbers $\{\alpha_\sigma\}_{\sigma\in\Sigma}$ such that $\sum_{\sigma\in\Sigma} \alpha_\sigma \boldsymbol{\mu}_\sigma(t) = 1$ for all $t \in \mathbb{Z}$.*

4. *For each $w, v \in \Sigma^+$, the process $\left[\boldsymbol{\mu}_w, \boldsymbol{\mu}_v\right]^T$ is wide-sense stationary.*

Let us now explain the intuition for each condition:

a) The first condition states that the product $\boldsymbol{\mu}_w$ is zero for words which are not in L. Intuitively, this means that the product $\boldsymbol{\mu}_w$ is of interest only for sequences w which belongs to L. In addition, it is assured that the elements of L are generated by a binary relation $\mathcal{E}$ that determines which letters are allowed to succeed the other, and for words outside L, $\boldsymbol{\mu}_w$ is zero.

Example 2.6. *To illustrate the first condition, consider an example where $\Sigma = \{1, 2, 3\}$ and the relation $\mathcal{E}$ is defined as follows:*

$$\mathcal{E} = \{(1, 2), (2, 3), (3, 1)\}$$

*$w = 123$ is an admissible **word** because:*

$$(1, 2) \in \mathcal{E}, \quad (2, 3) \in \mathcal{E}.$$

$v = 13$ is not admissible because:

$$(1,3) \notin \mathcal{E}.$$

In this case, the word v is not allowed, and if the switching process $\boldsymbol{\mu}$ tries to follow this sequence (word), it would result in $\boldsymbol{\mu}_v = 0$.

b) The second condition states that future products of $\boldsymbol{\mu}$ depend on past products of $\boldsymbol{\mu}$, and that $\boldsymbol{\mu}_\sigma(t)$ and $\boldsymbol{\mu}_{\sigma'}(t)$ are conditionally uncorrelated (independent) with respect to the past values of $\boldsymbol{\mu}$ if $\sigma \neq \sigma'$. This condition allows us to later prove a recurrence relation on the covariances of the outputs of the GLSS.

c) Intuitively, we want to ensure that past values of the outputs are products of past outputs and past switching. This will allow us to argue that innovation noises are uncorrelated. More precisely, the condition

$$\sum_{\sigma=1}^{k} \alpha_\sigma \boldsymbol{\mu}_\sigma(t) = 1,$$

for some coefficients $\{\alpha_\sigma\}_{\sigma=1}^{k}$, allows us to conclude that

$$\mathbf{y}(t-k) = \mathbf{y}(t-k)\boldsymbol{\mu}_1(t-k)\alpha_1 + \cdots + \mathbf{y}(t-k)\boldsymbol{\mu}_k(t-k)\alpha_k,$$

i.e., $\mathbf{y}(t-k)$ belongs to the linear spam generated by $\mathbf{y}(t-k), \boldsymbol{\mu}_{\sigma_1}(t-k), \ldots, \boldsymbol{\mu}_{\sigma_k}(t-k)$.

d) The last condition requires wide-sense stationarity of $\boldsymbol{\mu}_w$, i.e., that relevant moments of $\boldsymbol{\mu}_w$ are independent of time.

Note that Definition 2.2 is a reformulation of [91, Definition 1]. Notice, if $\boldsymbol{\mu}$ is admissible, then $\boldsymbol{\mu}_w, \boldsymbol{\mu}_v$ are jointly wide-sense stationary. Moreover, $\boldsymbol{\mu}_w(t)$ and $\boldsymbol{\mu}_v(t)$ are uncorrelated, if the last letters of w and v are different, and the real positive numbers $\{p_\sigma\}_{\sigma\in\Sigma}$ from part **2.** of Definition 2.2 satisfy

$$E[\boldsymbol{\mu}_{w\sigma}^2(t)] = p_\sigma E[\boldsymbol{\mu}_w^2(t)] \tag{2.6}$$

for any sequence $w \in \Sigma^+$, and letter $\sigma \in \Sigma$.

Below we recall from [91] a number of examples of admissible processes.

Example 2.7 (White noise). *If $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{n_\mu}]^T$ is an i.i.d. process such that $\boldsymbol{\mu}_1 = 1$, for all $i, j = 2, \ldots, n_\mu$, $t \in \mathbb{Z}$, $\boldsymbol{\mu}_i(t), \boldsymbol{\mu}_j(t)$ are independent and $\boldsymbol{\mu}_i(t)$ is zero mean, then $\boldsymbol{\mu}$ is admissible with $\mathcal{E} = \Sigma \times \Sigma$ and $p_\sigma = E[\boldsymbol{\mu}_\sigma^2(t)]$.*

Example 2.8 (Discrete valued i.i.d process). *Let $\boldsymbol{\theta}$ be an i.i.d process with values in $\Sigma = \{1, \ldots, n_\mu\}$. Let $\boldsymbol{\mu}_\sigma(t) = \chi(\boldsymbol{\theta}(t) = \sigma)$ for all $\sigma \in \Sigma$, $t \in \mathbb{Z}$, where χ is the indicator function, i.e.,*

$$\chi(\boldsymbol{\theta}(t) = \sigma) = \begin{cases} 1, & \text{if } \boldsymbol{\theta}(t) = \sigma \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathcal{E} = \Sigma \times \Sigma$, and $p_\sigma = P(\boldsymbol{\theta}(t) = \sigma)$, $\alpha_\sigma = 1$ for all $\sigma \in \Sigma$. Then $\boldsymbol{\mu}$ is admissible.

Example 2.9 (Markov chain)**.** *Assume that $\boldsymbol{\theta}$ is a stationary and ergodic Markov process whose state space is the finite set Θ. Let us take $\Sigma = \Theta \times \Theta$, and*

$$\mathcal{E} = \{(\sigma_1, \sigma_2) \in \Sigma \times \Sigma \mid \sigma_1 = (q_2, q_1), \sigma_2 = (q_3, q_2), q_1, q_2, q_3 \in \Theta\}.$$

Define

$$p_{(q_2,q_1)} := P(\boldsymbol{\theta}(t) = q_2 \mid \boldsymbol{\theta}(t-1) = q_1)$$

for $q_1, q_2 \in \Theta$, and note that $p_{(q_2,q_1)} > 0$. Note the switching from state q_1 to q_2 as:

$$\boldsymbol{\mu}_{(q_2,q_1)}(t) = \chi(\boldsymbol{\theta}(t+1) = q_2, \boldsymbol{\theta}(t) = q_1)$$

for all $q_1, q_2 \in \Theta$ and $t \in \mathbb{Z}$. Define $\alpha_\sigma = 1$ for all $\sigma \in \Sigma$. Let us identify Σ with the set $\{1, \dots, n_\mu\}$, where n_μ is the square of cardinality of Θ. Then, $\boldsymbol{\mu}$ is an admissible switching process.

Sketch of the proof that $\boldsymbol{\mu}$ is admissible. For the switching process $\boldsymbol{\mu}$ to be admissible, the admissibility conditions must hold:

1. **Existence of set $\mathcal{E}$:** In this example the first condition is satisfied, as for each pair (q_2, q_1), there exists a corresponding (q_3, q_2), such that (q_2, q_1), (q_3, q_2) are in the set $\mathcal{E}$, and where $q_1, q_2, q_3 \in \Theta$. Note that the set of admissible words L consists of words

$$w = (\theta_2, \theta_1)(\theta_3, \theta_2) \dots (\theta_k, \theta_{k-1}) \in \Sigma^+,$$

   where $\theta_1, \theta_2, \theta_3, \dots, \theta_k \in \Theta$ and $\big((\theta_i, \theta_{i-1}), (\theta_{i+1}, \theta_i)\big) \in \mathcal{E}$, for $i = 2, \dots, k-1$.

2. **Conditional expectations:** Consider four letters in Σ for $\{q_i\}_{i=1}^4, \{q_i'\}_{i=1}^4 \in \Theta$:

$$\sigma = (q_2, q_1) \text{ and } \sigma' = (q_2', q_1')$$

$$s = (q_4, q_3) \text{ and } s' = (q_4', q_3')$$

   Consider the following two words: $w = w's$, $v = v's' \in \Sigma^*$ where $s, s' \in \Sigma$. Then

$$w\sigma = w'(q_4, q_3)(q_2, q_1) \text{ and } v\sigma' = v'(q_4', q_3')(q_2', q_1').$$

   We know that for some $k_1, k_2 \in \Theta$ and $k = (k_1, k_2) \in \Sigma$:

$$\boldsymbol{\mu}_k(t) = \boldsymbol{\mu}_k^2(t) = \boldsymbol{\mu}_{(k_2,k_1)}(t) = \chi(\theta(t+1) = k_2, \theta(t) = k_1).$$

   We can then write:

$$\begin{aligned}\boldsymbol{\mu}_{w\sigma}(t) &= \boldsymbol{\mu}_{w'}(t-2)\boldsymbol{\mu}_s(t-1)\boldsymbol{\mu}_\sigma(t)\\ &= \boldsymbol{\mu}_{w'}(t-2)\chi(\theta(t) = q_4, \theta(t-1) = q_3)\chi(\theta(t+1) = q_2, \theta(t) = q_1),\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}_{v\sigma'}(t) &= \boldsymbol{\mu}_{v'}(t-2)\boldsymbol{\mu}_{s'}(t-1)\boldsymbol{\mu}_{\sigma'}(t)\\ &= \boldsymbol{\mu}_{v'}(t-2)\chi(\theta(t)=q_4',\theta(t-1)=q_3')\chi(\theta(t+1)=q_2',\theta(t)=q_1').\end{aligned}$$

From the definition of $\mathcal{E}$, we know that $w\sigma = w'(q_4,q_3)(q_2,q_1)$ is an admissible word, if $w's$ is an admissible word and $(s,\sigma) = ((q_4,q_3)(q_2,q_1)) \in \mathcal{E}$, i.e., if $q_4 = q_1$. Similarly, $v\sigma' = v'(q_4',q_3')(q_2',q_1')$ is admissible if $q_4' = q_1'$ and $v's'$ is an admissible word. In other words, we can say:

$$\begin{aligned}\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{v\sigma'}(t) &= \begin{cases}\boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)\boldsymbol{\mu}_\sigma^2(t), & \text{if } \sigma=\sigma'\\ \boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)\boldsymbol{\mu}_\sigma(t)\boldsymbol{\mu}_{\sigma'}(t), & \text{otherwise}\end{cases}\\ &= \begin{cases}\boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)\chi(\theta(t+1)=q_2,\theta(t)=q_1), & \text{if } \sigma=\sigma'\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

In the equation above, we used the fact that

$$\boldsymbol{\mu}_\sigma(t)\boldsymbol{\mu}_{\sigma'}(t) = \chi(\theta(t+1)=q_2,\theta(t)=q_1)\chi(\theta(t+1)=q_2',\theta(t)=q_1'),$$

which is equal to zero when $\sigma \neq \sigma'$, because $\sigma \neq \sigma'$ implies that at least one of the indicator functions is equal to zero. We can now start computing the conditional expectation – that we will call E_c – for the case where $\sigma = \sigma'$ and (s,σ) and (s',σ') are both admissible sequences ($E_c = 0$ otherwise):

$$\begin{aligned}E_c &= E\left[\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{v\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}\right]\\ &= E\left[\boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)\chi(\theta(t+1)=q_2,\theta(t)=q_1) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}\right].\end{aligned}$$

We know that:

$$\begin{aligned}\mathscr{F}_t^{\boldsymbol{\mu},-} &= \{\boldsymbol{\mu}_k(s) \mid s<t\}\\ &= \{\chi(\theta(s+1)=k_2,\ \theta(s)=k_2) \mid s<t \text{ and } k_1,k_2\in\Theta\}.\end{aligned}$$

Since $\boldsymbol{\mu}_w(t-1)$ and $\boldsymbol{\mu}_w(t-1)$ are $\mathscr{F}_t^{\boldsymbol{\mu},-}$ measurable, as they are products of the random variables which are among the generators of $\mathscr{F}_t^{\boldsymbol{\mu},-}$, it follows that:

$$E_c = \boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)E\left[\chi(\theta(t+1)=q_2,\theta(t)=q_1) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}\right].$$

It is clear that

$$\chi(\theta(t)=q) = \sum_{q_1\in\Theta}\boldsymbol{\mu}_{(q,q_1)}(t-1),$$

since at any time instant t, the process $\theta(t)$ must be in exactly one state q, the sum above counts all possible transitions from any previous state q_1 to the state q. This means that

$$\chi(\theta(t)=q) \in \mathscr{F}_t^{\boldsymbol{\mu},-}.$$

Consequently, the σ-algebra $\sigma\big(\chi(\theta(s)=q) \mid q \in \Theta, s \leq t\big)$ generated by $\{\chi(\theta(s)=q) \mid q \in \Theta, s \leq t\}$ and $\mathscr{F}_t^{\boldsymbol{\mu},-}$ are equal. It follows that:
$E_c = \boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)E\left[\chi(\theta(t+1)=q_2,\theta(t)=q_1) \mid \sigma\big(\chi(\theta(s)=q) \mid q \in \Theta, s \leq t\big)\}\right].$
This means that E_c represents the conditional probability that the system transitions from q_1 to q_2 given all past states. By definition of the Markov chain properties, we can replace the conditional expectation by the transition probability:

$$E\left[\chi(\theta(t+1)=q_2,\theta(t)=q_1) \mid \sigma\big(\chi(\theta(s)=q) \mid q \in \Theta, s \leq t\big)\right] = p_{(q_1,q_2)}\chi(\theta(t)=q_1).$$

Hence,

$$E_c = \boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)p_{(q_1,q_2)}\chi(\theta(t)=q_1),$$

where $p_{(q_1,q_2)} = P(\theta(t+1)=q_2,\theta(t)=q_1)$ is the transition probability from the state q_1 to the state q_2. Notice that:

$$\chi(\theta(t+1)=q_2,\theta(t)=q_1)\chi(\theta(t)=q_1) = \chi(\theta(t+1)=q_2,\theta(t)=q_1).$$

Then:

$$\begin{aligned}\boldsymbol{\mu}_w(t-1)\chi(\theta(t)=q_1) &= \boldsymbol{\mu}_{w'}(t-2)\chi(\theta(t)=q_3,\theta(t-1)=q_4)\chi(\theta(t)=q_1)\\ &= \boldsymbol{\mu}_{w'}(t-2)\chi(\theta(t)=q_3,\theta(t-1)=q_4)\\ &= \boldsymbol{\mu}_w(t-1)\end{aligned}$$

since $q_4 = q_1$ because we are studying the case where (s,σ) is admissible. Finally, we can then write:

$$E_c = E\left[\boldsymbol{\mu}_{w\sigma}(t)\boldsymbol{\mu}_{v\sigma'}(t) \mid \mathscr{F}_t^{\boldsymbol{\mu},-}\right] = \boldsymbol{\mu}_w(t-1)\boldsymbol{\mu}_v(t-1)p_{(q_1,q_2)}$$

3. **Normalization condition:** with $\alpha_\sigma = 1$, for $\sigma \in \Sigma$, we get

$$\sum_{\sigma\in\Sigma}\alpha_\sigma\boldsymbol{\mu}_\sigma(t) = \sum_{\sigma\in\Sigma}\boldsymbol{\mu}_\sigma(t) = 1$$

because only one $\boldsymbol{\mu}_\sigma(t)$ can be equal to 1 at a time.

4. **Wide Stationarity condition:** Since θ is a stationary Markov process, the distribution of $\boldsymbol{\mu}_w(t)$ does not depend on t, i.e., $\boldsymbol{\mu}_w(t)$ is stationary and hence $\left[\boldsymbol{\mu}_w, \boldsymbol{\mu}_v\right]^T$ is wide-sense stationary for all $w, v \in \Sigma^+$.

□

Notice that, $\boldsymbol{\mu}$ uniquely determines a collection of numbers $\{p_\sigma\}_{\sigma\in\Sigma}$. Indeed, if $(\sigma_1,\sigma) \in \mathcal{E}$ and $\boldsymbol{\mu}_{\sigma_1}(t) \neq 0$, then

$$p_\sigma = \frac{E\left[\boldsymbol{\mu}^2_{\sigma_1\sigma}(t)\right]}{E\left[\boldsymbol{\mu}^2_{\sigma_1}(t)\right]}.$$

This latter collection will play an important role in the sequel. For instance, for $\boldsymbol{\mu}$ from Example 2.7, p_i is the variance of $\boldsymbol{\mu}_i$, and for $\boldsymbol{\mu}$ from Example 2.8, p_i is the probability $P(\theta(t)=i)$, for all $i \in \Sigma$.

2.4.2 Zero Mean Wide Sense Stationary process with respect to Inputs (ZMWSSI), Square Integrable process with respect to Inputs (SII) and white noise process with respect to switching

In this subsection, we recall the concept of Zero Mean Wide Sense Stationary process with respect to Inputs (ZMWSSI) from [91], in addition to the Square Integrable with respect to Inputs (SII) and white noise with respect to switching properties. Note that, by analogy with [91], the inputs are the switching in our case. However, in order to stay consistent with the terminology of the literature, we decided to stick with the abbreviations ZMWSSI and SII, despite that, in our case, the process is zero mean wide sense stationary and Square Integrable with respect to the switching.

Intuitively, these concepts concerning the processes are needed for defining the class of GLSS studied in the following chapters. To this end, we need to introduce some terminology which will be used later on in the thesis.

In this section, we assume that $\boldsymbol{\mu}$ is admissible switching, i.e., it satisfies Definition 2.2.

Let $\{p_\sigma\}_{\sigma\in\Sigma}$ be the constants determined by $\boldsymbol{\mu}$ as explained in **part (2.)** of Definition 2.2, and define the products

$$p_w = p_{\sigma_1} p_{\sigma_2} \cdots p_{\sigma_k}. \tag{2.7}$$

for any $w = \sigma_1\sigma_2\cdots\sigma_k \in \Sigma^+$, where $\sigma_1, \sigma_2, \dots, \sigma_k \in \Sigma$. For an empty word $w = \epsilon$, we set $p_\epsilon = 1$. For a stochastic process $\mathbf{r} \in \mathbb{R}^m$ and for each $w \in \Sigma^*$ we define the stochastic process $\mathbf{z}^{\mathbf{r}}_w$ as

$$\mathbf{z}^{\mathbf{r}}_w(t) = \mathbf{r}(t - |w|)\boldsymbol{\mu}_w(t-1)\frac{1}{\sqrt{p_w}}, \tag{2.8}$$

where $\boldsymbol{\mu}_w$ and p_w are as in (2.5) and (2.7). For $w = \epsilon$, let $\mathbf{z}^{\mathbf{r}}_w(t) = \mathbf{r}(t)$. The process $\mathbf{z}^{\mathbf{r}}_w$ in (2.8) is interpreted as the product of the *past* of $\mathbf{r}$ and $\boldsymbol{\mu}$. The process $\mathbf{z}^{\mathbf{r}}_w$ will be used as predictors for future values of $\mathbf{r}$ for various choices of $\mathbf{r}$.

After getting familiar with the necessary terminology, we introduce the following definition:

Definition 2.3 (ZMWSSI w.r.t. $\boldsymbol{\mu}$). *A process* $\mathbf{r}$ *is* Zero Mean Wide Sense Stationary w.r.t. admissible switching $\boldsymbol{\mu}$ *(ZMWSSI) if the following holds:*

(1) *For $t \in \mathbb{Z}$, let us consider*

- *the σ-algebra generated by the variables $\{\mathbf{r}(k)\}_{k\leq t}$ denoted by $\mathcal{F}^{\mathbf{r}}_t$,*
- *the σ-algebra generated by the variables $\{\boldsymbol{\mu}_\sigma(k)\}_{k<t,\sigma\in\Sigma}$ denoted by $\mathcal{F}^{\boldsymbol{\mu},-}_t$,*
- *and the σ-algebra generated by the variables $\{\boldsymbol{\mu}_\sigma(k)\}_{k\geq t,\sigma\in\Sigma}$, denoted by $\mathcal{F}^{\boldsymbol{\mu},+}_t$,*

Then $\mathcal{F}^{\mathbf{r}}_t$ and $\mathcal{F}^{\boldsymbol{\mu},+}_t$ are conditionally independent w.r.t. $\mathcal{F}^{\boldsymbol{\mu},-}_t$, [121, Definition 2.9.1].

(2) The processes $\{\mathbf{r}, \{\mathbf{z}^{\mathbf{r}}_w\}_{w\in\Sigma^+}\}$ *are zero mean, square integrable and are jointly wide sense stationary. i.e.,* $\forall t, s, k \in \mathbb{Z}, w, v \in \Sigma^+$,

$$
\begin{aligned}
&E\left[\mathbf{r}(t)\right] = 0,\\
&E\left[\mathbf{z}^{\mathbf{r}}_w(t)\right] = 0,\\
&E\left[\mathbf{r}(t+k)(\mathbf{z}^{\mathbf{r}}_w(s+k))^T\right] = E\left[\mathbf{r}(t)(\mathbf{z}^{\mathbf{r}}_w(s))^T\right],\\
&E\left[\mathbf{r}(t+k)(\mathbf{r}(s+k))^T\right] = E\left[\mathbf{r}(t)(\mathbf{r}(s))^T\right],\\
&E\left[\mathbf{z}^{\mathbf{r}}_w(t+k)(\mathbf{z}^{\mathbf{r}}_v(s+k))^T\right] = E\left[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{r}}_v(s))^T\right].
\end{aligned}
$$

The first condition means that future values of $\boldsymbol{\mu}$ do not depend on past or current values of $\mathbf{r}$, except through their shared dependency on past values of $\boldsymbol{\mu}$. In other words, there is no feedback from $\mathbf{r}$ to $\boldsymbol{\mu}$: the process $\mathbf{r}$ does not influence future values of $\boldsymbol{\mu}$ beyond what is already encoded in past values of $\boldsymbol{\mu}$. The only way information from $\mathbf{r}$ can influence future values of $\boldsymbol{\mu}$ is indirectly, via the effect that past values of $\boldsymbol{\mu}$ had on both $\mathbf{r}$ and $\boldsymbol{\mu}$. This type of assumption is common in stochastic processes when modeling systems where the observed process $\mathbf{r}$ is influenced by an external source $\boldsymbol{\mu}$, but does not itself alter the dynamics of $\boldsymbol{\mu}$.

The second condition of the ZMWSSI definition means that the processes involved are wide sense stationanry.

The first and second conditions imply, together with [91, Lemma 7], that the relationship covariances of $\mathbf{r}$ have the recursive structure described below.

Notation 2.4. *For any two ZMWSSI processes* $\mathbf{r}$ *and* $\mathbf{b}$ *and for sequences* $w, v \in \Sigma^*$, *define the covariances:*

$$
\Lambda^{\mathbf{r},\mathbf{b}}_w = E[\mathbf{r}(t)(\mathbf{z}^{\mathbf{b}}_w(t))^T], \quad T^{\mathbf{r},\mathbf{b}}_{w,v} = E[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{b}}_v(t))^T] \tag{2.9}
$$

In particular, if $\mathbf{r}$ is ZMWSSI w.r.t. $\boldsymbol{\mu}$, then, for $w, v, s \in \Sigma^*$ and $\sigma \in \Sigma$, we have:

$$
T^{\mathbf{r},\mathbf{r}}_{w\sigma,v\sigma'} = \begin{cases} T^{\mathbf{r},\mathbf{r}}_{w,v} & \text{if } \sigma = \sigma' \text{ and } w\sigma \in L, v\sigma' \in L \\ 0 & \text{if } \sigma \neq \sigma' \text{ or } w\sigma \notin L \text{ or } v\sigma' \notin L \end{cases}
$$

In particular:

$$
T^{\mathbf{r},\mathbf{r}}_{w,v} = \begin{cases} (\Lambda^{\mathbf{r},\mathbf{r}}_s)^T & \text{if } w = sv \text{ and } v, w \in L, |s| > 0 \\ \Lambda^{\mathbf{r},\mathbf{r}}_s & \text{if } v = sw \text{ and } v, w \in L, |s| > 0 \\ 0 & \text{otherwise} \end{cases}
$$

and if $|w| = |v|$ then the following holds

$$
T^{\mathbf{r},\mathbf{r}}_{w,v} = \begin{cases} T^{\mathbf{r},\mathbf{r}}_{\sigma,\sigma} & \text{if } v = w = \sigma s, w \in L \\ 0 & \text{if } w \neq v \text{ or } w \notin L \text{ or } v \notin L \end{cases}
$$

This means that ZMWSSI is an extension of wide-sense stationarity, where Σ^+ is viewed as time axis: ZMWSSI implies that the covariances $E\left[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{r}}_v(t))^T\right]$ do not depend on t, and they depend on the difference between v and w.

In order to understand better the intuition behind the properties of a ZMWSSI process, let us present the following covariance matrix: $T^{\mathbf{r},\mathbf{r}} = (T^{r,r}_{w,v})_{w,v \in L} = (E\left[\mathbf{z}^{\mathbf{r}}_w(t)(\mathbf{z}^{\mathbf{r}}_v(t))^T\right])_{w,v \in l}$. This matrix cannot be easily illustrated. Therefore, we will consider, for a simple illustration, the case where $\Sigma = \{1,2\}$ and where the set of the admissible sequences $L = \Sigma^+ = \{1,2,11,12,21,22,111,112,121,122,\dots\}$. In this case, we will get:

$$T^{\mathbf{r},\mathbf{r}} = \begin{bmatrix}
T^{\mathbf{r},\mathbf{r}}_{1,1} & T^{\mathbf{r},\mathbf{r}}_{1,2} & T^{\mathbf{r},\mathbf{r}}_{1,11} & T^{\mathbf{r},\mathbf{r}}_{1,12} & T^{\mathbf{r},\mathbf{r}}_{1,21} & T^{\mathbf{r},\mathbf{r}}_{1,22} & T^{\mathbf{r},\mathbf{r}}_{1,111} & T^{\mathbf{r},\mathbf{r}}_{1,112} & T^{\mathbf{r},\mathbf{r}}_{1,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{2,1} & T^{\mathbf{r},\mathbf{r}}_{2,2} & T^{\mathbf{r},\mathbf{r}}_{2,11} & T^{\mathbf{r},\mathbf{r}}_{2,12} & T^{\mathbf{r},\mathbf{r}}_{2,21} & T^{\mathbf{r},\mathbf{r}}_{2,22} & T^{\mathbf{r},\mathbf{r}}_{2,111} & T^{\mathbf{r},\mathbf{r}}_{2,112} & T^{\mathbf{r},\mathbf{r}}_{2,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{11,1} & T^{\mathbf{r},\mathbf{r}}_{11,2} & T^{\mathbf{r},\mathbf{r}}_{11,11} & T^{\mathbf{r},\mathbf{r}}_{11,12} & T^{\mathbf{r},\mathbf{r}}_{11,21} & T^{\mathbf{r},\mathbf{r}}_{11,22} & T^{\mathbf{r},\mathbf{r}}_{11,111} & T^{\mathbf{r},\mathbf{r}}_{11,112} & T^{\mathbf{r},\mathbf{r}}_{11,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{12,1} & T^{\mathbf{r},\mathbf{r}}_{12,2} & T^{\mathbf{r},\mathbf{r}}_{12,11} & T^{\mathbf{r},\mathbf{r}}_{12,12} & T^{\mathbf{r},\mathbf{r}}_{12,21} & T^{\mathbf{r},\mathbf{r}}_{12,22} & T^{\mathbf{r},\mathbf{r}}_{12,111} & T^{\mathbf{r},\mathbf{r}}_{12,112} & T^{\mathbf{r},\mathbf{r}}_{12,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{21,1} & T^{\mathbf{r},\mathbf{r}}_{21,2} & T^{\mathbf{r},\mathbf{r}}_{21,11} & T^{\mathbf{r},\mathbf{r}}_{21,12} & T^{\mathbf{r},\mathbf{r}}_{21,21} & T^{\mathbf{r},\mathbf{r}}_{21,22} & T^{\mathbf{r},\mathbf{r}}_{21,111} & T^{\mathbf{r},\mathbf{r}}_{21,112} & T^{\mathbf{r},\mathbf{r}}_{21,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{22,1} & T^{\mathbf{r},\mathbf{r}}_{22,2} & T^{\mathbf{r},\mathbf{r}}_{22,11} & T^{\mathbf{r},\mathbf{r}}_{22,12} & T^{\mathbf{r},\mathbf{r}}_{22,21} & T^{\mathbf{r},\mathbf{r}}_{22,22} & T^{\mathbf{r},\mathbf{r}}_{22,111} & T^{\mathbf{r},\mathbf{r}}_{22,112} & T^{\mathbf{r},\mathbf{r}}_{22,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{111,1} & T^{\mathbf{r},\mathbf{r}}_{111,2} & T^{\mathbf{r},\mathbf{r}}_{111,11} & T^{\mathbf{r},\mathbf{r}}_{111,12} & T^{\mathbf{r},\mathbf{r}}_{111,21} & T^{\mathbf{r},\mathbf{r}}_{111,22} & T^{\mathbf{r},\mathbf{r}}_{111,111} & T^{\mathbf{r},\mathbf{r}}_{111,112} & T^{\mathbf{r},\mathbf{r}}_{111,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{112,1} & T^{\mathbf{r},\mathbf{r}}_{112,2} & T^{\mathbf{r},\mathbf{r}}_{112,11} & T^{\mathbf{r},\mathbf{r}}_{112,12} & T^{\mathbf{r},\mathbf{r}}_{112,21} & T^{\mathbf{r},\mathbf{r}}_{112,22} & T^{\mathbf{r},\mathbf{r}}_{112,111} & T^{\mathbf{r},\mathbf{r}}_{112,112} & T^{\mathbf{r},\mathbf{r}}_{112,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{121,1} & T^{\mathbf{r},\mathbf{r}}_{121,2} & T^{\mathbf{r},\mathbf{r}}_{121,11} & T^{\mathbf{r},\mathbf{r}}_{121,12} & T^{\mathbf{r},\mathbf{r}}_{121,21} & T^{\mathbf{r},\mathbf{r}}_{121,22} & T^{\mathbf{r},\mathbf{r}}_{121,111} & T^{\mathbf{r},\mathbf{r}}_{121,112} & T^{\mathbf{r},\mathbf{r}}_{121,121} & \cdots \\
T^{\mathbf{r},\mathbf{r}}_{122,1} & T^{\mathbf{r},\mathbf{r}}_{122,2} & T^{\mathbf{r},\mathbf{r}}_{122,11} & T^{\mathbf{r},\mathbf{r}}_{122,12} & T^{\mathbf{r},\mathbf{r}}_{122,21} & T^{\mathbf{r},\mathbf{r}}_{122,22} & T^{\mathbf{r},\mathbf{r}}_{122,111} & T^{\mathbf{r},\mathbf{r}}_{122,112} & T^{\mathbf{r},\mathbf{r}}_{122,121} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.$$

Because $\mathbf{r}$ is a ZMWSSI process, then the value of $T^{\mathbf{r},\mathbf{r}}_{w,v}$ is determined by the values of $T^{\mathbf{r},\mathbf{r}}_{s,s'}$ where $s, s' \in \Sigma$ are prefixes of w and v. Moreover, if the last letters of v and w are not the same, or that the words does not belong to the set of admissible sequences, then the covariances $T^{\mathbf{r},\mathbf{r}}_{w,v}$ are zero. It follows that the matrix above can be written as follows:

$$T^{\mathbf{r},\mathbf{r}} = \begin{bmatrix}
T^{\mathbf{r},\mathbf{r}}_{1,1} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{1} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{2} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{11} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{12} & \cdots \\
0 & T^{\mathbf{r},\mathbf{r}}_{2,2} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{1} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{2} & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{11} & 0 & \cdots \\
(\Lambda^{\mathbf{r},\mathbf{r}}_{1})^T & 0 & T^{\mathbf{r},\mathbf{r}}_{1,1} & 0 & 0 & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{1} & 0 & 0 & \cdots \\
0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{1})^T & 0 & T^{\mathbf{r},\mathbf{r}}_{1,1} & 0 & 0 & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{1} & 0 & \cdots \\
(\Lambda^{\mathbf{r},\mathbf{r}}_{2})^T & 0 & 0 & 0 & T^{\mathbf{r},\mathbf{r}}_{2,2} & 0 & 0 & 0 & \Lambda^{\mathbf{r},\mathbf{r}}_{1} & \cdots \\
0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{2})^T & 0 & 0 & 0 & T^{\mathbf{r},\mathbf{r}}_{2,2} & 0 & 0 & 0 & \cdots \\
(\Lambda^{\mathbf{r},\mathbf{r}}_{11})^T & 0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{1})^T & 0 & 0 & 0 & T^{\mathbf{r},\mathbf{r}}_{1,1} & 0 & 0 & \cdots \\
0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{11})^T & 0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{1})^T & 0 & 0 & 0 & T^{\mathbf{r},\mathbf{r}}_{1,1} & 0 & \cdots \\
(\Lambda^{\mathbf{r},\mathbf{r}}_{12})^T & 0 & 0 & 0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{1})^T & 0 & 0 & 0 & T^{\mathbf{r},\mathbf{r}}_{1,1} & \cdots \\
0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{12})^T & 0 & 0 & 0 & (\Lambda^{\mathbf{r},\mathbf{r}}_{1})^T & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.$$

We can conclude that the covariance $T^{\mathbf{r},\mathbf{r}}$ has a certain ***recursive*** structure in the length of w and v. One can easily notice that the whole matrix can be computed if we know the first two rows due to the repetition of their values. For the more general case, where $n_\mu > 2$, one needs the values of the first n_μ rows to compute $T^{\mathbf{r},\mathbf{r}}$.

Furthermore, let us illustrate the more restricted linear case [56] where $\Sigma = 1$, i.e., $\boldsymbol{\mu}(t) = 1$

for all $t \in \mathbb{Z}$. To this end, let us define the following matrices:

$$T(k,l) = T_{w,v}^{\mathbf{r},\mathbf{r}} = E\left[\mathbf{r}(t+k)(\mathbf{r}(t+l))^T\right]$$
$$\Lambda(l) = \Lambda_v^{\mathbf{r},\mathbf{r}} = E\left[\mathbf{r}(t+l)(\mathbf{r}(t))^T\right],$$

where k and l are the lengths of w and v respectively. Since $\Sigma = 1$, in this case, w and v are sequences of 1 repeated k and l times respectively. This means that $T(k,l) = \Lambda(k-l) = E\left[\mathbf{r}(k-l)(\mathbf{r}(0))^T\right]$ because $\mathbf{r} \in \mathbb{R}^{n_y}$ is ZMWSSI. Now, let us consider the stacked infinite dimensional vectors of the past value given by

$$p(t) = \begin{bmatrix} \mathbf{r}(t-1) \\ \mathbf{r}(t-2) \\ \vdots \end{bmatrix}$$

then, the covariance matrix of the auto-covariance matrices of the past is given by:

$$T_- = E\left[p(t)p^T(t)\right] = \begin{bmatrix} \Lambda(0) & \Lambda(1) & \Lambda(2) & \cdots \\ (\Lambda(1))^T & \Lambda(0) & \Lambda(1) & \cdots \\ (\Lambda(2))^T & (\Lambda(1))^T & \Lambda(0) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} \Lambda(0) & \Lambda(1) & \Lambda(2) & \cdots \\ (\Lambda(1))^T & & & \\ (\Lambda(2))^T & & T_- & \\ \vdots & & & \end{bmatrix}.$$

Consider now the auto-covariance matrices of the past where they are finite Toeplitz matrices:

$$T_-^N = \begin{bmatrix} \Lambda(0) & \Lambda(1) & \Lambda(2) & \cdots & \Lambda(N-1) \\ (\Lambda(1)))^T & \Lambda(0) & \Lambda(1) & \cdots & \Lambda(N-2) \\ \vdots & \vdots & \vdots & & \vdots \\ (\Lambda(N-1))^T & (\Lambda(N-2))^T & (\Lambda(N-3))^T & \cdots & \Lambda(0) \end{bmatrix},$$

$$T_-^N = \begin{bmatrix} \Lambda(0) & \Lambda(1) & \Lambda(2) & \cdots & \Lambda(N-1) \\ (\Lambda(1))^T & & & & \\ (\Lambda(2))^T & & T_-^{N-1} & & \\ \vdots & & & & \\ (\Lambda(N-1))^T & & & & \end{bmatrix}.$$

We can see the same ***recursive*** relationship as mentioned before in the general case. This is important, because it means that the projection $\hat{\mathbf{r}}_N(t+k)$ of the future output $\mathbf{r}(t+k)$ onto the past outputs $\{\mathbf{r}(t-1), \ldots, \mathbf{r}(t-N)\}$, defined by

$$\hat{\mathbf{r}}_N(t+k) = \begin{bmatrix} \Lambda(k) & \cdots & \Lambda(k-N) \end{bmatrix} (T_-^N)^{-1} \begin{bmatrix} \mathbf{r}(t-1) \\ \vdots \\ \mathbf{r}(t-N) \end{bmatrix}$$

will have a ***recursive*** expression. This recursive structure of T_{-}^{N} helps us to derive formulas for best linear prediction of $\mathbf{r}$ based on past values of $\mathbf{r}$.

Note that, like the general case, we need the values of the first row ($n_\mu = 1$) in order to compute T_{-} (which is equal to $T^{\mathbf{r},\mathbf{r}}$ for the general case). However, unlike the general case, T_{-} is a Toeplitz matrix due to the absence of the zeroes.

The argument above can be extended to the general case. In particular, we can formulate the same recursive relationship between the estimator of the future value $\hat{\mathbf{r}}_N(t+k)$ and the past values of $\mathbf{r}$ by selecting the necessary rows and columns from $T^{\mathbf{r},\mathbf{r}}$.

In addition to stationarity, we will need another technical condition for outputs of the GLSSs of interest. More precisely, we will need to consider projections of products of $\{\boldsymbol{\mu}_\sigma\}_{\sigma\in\Sigma}$ and future outputs to the space generated by products of past outputs and past values of $\{\boldsymbol{\mu}_\sigma\}_{\sigma\in\Sigma}$. To this end, we need to ensure that the former product are in $\mathcal{H}_1$, which will be ensured by the following definition.

Definition 2.4 (SII process w.r.t. $\boldsymbol{\mu}$, [91])**.** *A process* $\mathbf{r}$ *is* Square Integrable w.r.t. the switching $\boldsymbol{\mu}$ (SII)*, if for all $w \in \Sigma^*$, $t \in \mathbb{Z}$, the random variable* $\mathbf{r}(t+|w|)\boldsymbol{\mu}_w(t+|w|-1)$ *is square integrable.*

Remark 2.3. *As it was mentioned in [91, Section III, Remark 2], if* $\mathbf{r}$ *is square integrable and* $\boldsymbol{\mu}_\sigma$ *is essentially bounded for all $\sigma \in \Sigma$, i.e, there exists a constant $\mathscr{K} > 0$ such that $|\boldsymbol{\mu}_\sigma(t)| \leq \mathscr{K}$ almost everywhere for all $\sigma \in \Sigma$, $t \in \mathbb{Z}$, then* $\mathbf{r}$ *is SII. In particular, if the switching is bounded, and* $\mathbf{r}$ *is ZMWSSI, then* $\mathbf{r}$ *is SII*

Among ZMWSSI and SII processes, we will need another subclass of processes which is an extension of the notion of white noise process. Therefore we present the following definition:

Definition 2.5 (White noise w.r.t. $\boldsymbol{\mu}$)**.** *A ZMWSSI process* $\mathbf{r}$ *is a white noise w.r.t.* $\boldsymbol{\mu}$*, if for all $w \in \Sigma^+$, $v \in \Sigma^*$, $\sigma \in \Sigma$,*

$$T_{w,\sigma v}^{\mathbf{r},\mathbf{r}} = E[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_{\sigma v}^{\mathbf{r}}(t))^T] = \begin{cases} 0 & \textit{if } w \neq \sigma v \textit{ or } w \notin L \textit{ or } \sigma v \notin L \\ T_{\sigma,\sigma}^{\mathbf{r},\mathbf{r}} & \textit{if } w = \sigma v \textit{ and } w \in L \end{cases}$$

and $E[\mathbf{z}_\sigma^{\mathbf{r}}(t)(\mathbf{z}_\sigma^{\mathbf{r}}(t))^T]$ is nonsingular for all $\sigma \in \Sigma$.

Intuitively, if $\mathbf{r}$ is a white noise process w.r.t. $\boldsymbol{\mu}$, then $\{\mathbf{z}_w^{\mathbf{r}}(t)\}_{w\in\Sigma^+}$ is a sequence of uncorrelated random variables. Due to **part (3.)** of Definition 2.2, the product $\mathbf{r}(t-k)\mathbf{r}(t)$ is a linear combination of $\{\mathbf{z}_w^{\mathbf{r}}(t)\}_{w\in\Sigma^+}$, hence a white noise process w.r.t. $\boldsymbol{\mu}$ is also a white noise process in the classical sense. Conversely, if $\mathbf{r}$ is a white noise and it is independent of $\{\boldsymbol{\mu}(s)\}_{s\in\mathbb{Z}}$, then it is a white noise process w.r.t. $\boldsymbol{\mu}$.

With the definitions of these properties, we are now ready to present the assumption of stationarity discussed in Section 2.2.

2.5 Stationary GLSS

In this section, after introducing the necessary definitions and characterizations, we present the definition of stationary GLSS and autonomous stationary GLSS.

Definition 2.6 (Stationary GLSS). *A* stationary GLSS *(abbreviated sGLSS) realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *(sGLSS of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *for short) is a system in the form of* (2.1)*, such that*

1. $\mathbf{w} = \begin{bmatrix} \mathbf{v}^T, & \mathbf{u}^T \end{bmatrix}^T$ *is a white noise process w.r.t.* $\boldsymbol{\mu}$*, and the following orthogonality condition holds:* $E[\mathbf{v}(t)\mathbf{u}^T(t)\boldsymbol{\mu}_\sigma^2(t)] = 0$*, for all* $\sigma \in \Sigma$*.*
2. *The process* $\begin{bmatrix} \mathbf{x}^T, & \mathbf{w}^T \end{bmatrix}^T$ *is a ZMWSSI, and the following orthogonality conditions hold:* $E[\mathbf{z}_\sigma^{\mathbf{x}}(t)(\mathbf{z}_\sigma^{\mathbf{w}}(t))^T] = 0$ *and* $E[\mathbf{x}(t)(\mathbf{z}_w^{\mathbf{w}}(t))^T] = 0$*, for all* $\sigma \in \Sigma$*, and* $w \in \Sigma^+$*.*
3. *The eigenvalues of the matrix* $\sum_{\sigma \in \Sigma} p_\sigma A_\sigma \otimes A_\sigma$ *are inside the open unit disk, where* $\otimes$ *is the Kronecker product.*
4. *For all letters* $\sigma_1, \sigma_2 \in \Sigma$*, if* $(\sigma_1, \sigma_2) \notin \mathcal{E}$*, i.e.,* $\sigma_1\sigma_2$ *is not an admissible sequence, then* $A_{\sigma_2}A_{\sigma_1} = 0$ *and* $A_{\sigma_2} \begin{bmatrix} B_{\sigma_1} & K_{\sigma_1} \end{bmatrix} E[\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t))^T] = 0$*.*

If $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *is clear from the context, then we refer to a sGLSS realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *as sGLSS.*

Some remarks are in order: First, we defined sGLSS as a realization of tuple of processes $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. That is, in order to define a sGLSS, one should not only specify the system matrices and the noise $\mathbf{v}$, but also the input, switching and output processes. In particular, a certain collection of matrices $\{A_i, B_i, K_i\}_{i=1}^{n_\mu}, C$ and D may cease to give rise to a valid sGLSS, if $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ is changed. This is reflected in the terminology, as we speak of sGLSS realizations of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. As it was mentioned in the definition, *for the sake of brevity we will refer to a sGLSS realization of* $(\mathbf{y}, \mathbf{y}, \boldsymbol{\mu})$ *as a sGLSS of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. We will use the two forms interchangeably in the rest of the thesis.

Second, this definition is coherent with the usual definition of stochastic dynamical system [121]. More precisely, we view a GLSS as encoded version of conditional probability distribution of future states and current outputs based on past outputs, inputs and state. The noise, intuitively, is the difference between the next state and current output and their best prediction based on past inputs, states and outputs.

The definition above includes several technical conditions that ensure stability, orthogonality, and well-behaved stochastic properties. Let's break down each condition intuitively:

- **Condition 1** ensures that the input and the noise are uncorrelated. Also, in ensures that they are both independent from the switching signal. These are standard assumptions in stochastic systems theory. Condition 1 also ensures that the input $\mathbf{u}$ is white noise. This ensures persistence of excitation of the input, and simplifies the underlying system identification problem.

- **Condition 2** ensures that the mean and auto-covariance of the processes do not change over time, which is essential for stationarity. In addition, it ensures that the state is uncorrelated with the current values of the noise and the switching. This helps in designing optimal state estimation algorithms (Kalman-like filters) since the state updates will not be contaminated by correlated noise.

- **Condition 3** ensures that the system is stable in the sense of mean-square stability. Since the Kronecker product represents how the system propagates uncertainty over time, then the eigenvalues being inside the unit disk means that the system does not diverge over time, which is necessary for long-term predictions and control.

- **Condition 4** imposes constraints on transitions between states. It ensures that the product of system matrices which correspond to inadmissible words are zero. These product capture the effect of products of switching signals $\boldsymbol{\mu}_w$ for $w \notin L$. Since $\boldsymbol{\mu}_w = 0$ for $w \notin L$, these matrix products never show up in the dynamics, hence they could take any values without influencing the observed behavior of the system. By requiring them to be zero we eliminate this source of ambiguity. This is particularly important for GLSS, in particular for switched or Jump-Markov systems, where the transition structure can be constrained by physical rules or external conditions.

Note that **Condition 2.** implies that $\mathbf{w}$ is a white noise, i.e., $E[\mathbf{w}(t)] = 0$ and, for $t \neq s$,

$$E[\mathbf{w}(t)\mathbf{w}^T(s)] = 0.$$

In this thesis we will also be interested in autonomous stationary GLSSs, which are sGLSSs with no inputs. Their formal definition goes as follows.

In this thesis we will also be interested in autonomous stationary GLSSs, which are sGLSSs with no inputs. Their formal definition goes as follows.

Definition 2.7 (Autonomous Stationary GLSS)**.** *An* autonomous stationary stochastic GLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ *(asGLSS of* $(\mathbf{y}, \boldsymbol{\mu})$ *for short) is a system of the form* (2.4)*, such that the same conditions of Definition 2.6 apply, but* $\mathbf{w} = \mathbf{v}$ *and the input process* $\mathbf{u}$ *does not exist and* $B_i = 0$*, for* $i \in \Sigma$*, and* $D = 0$*. If* $(\mathbf{y}, \boldsymbol{\mu})$ *is clear from the context, then we refer to an asGLSS realization of* $(\mathbf{y}, \boldsymbol{\mu})$ *as asGLSS*

The state of an sGLSS is uniquely determined by its matrices and noise process. From [91, Lemma 2] it follows that

$$\mathbf{x}(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma, \sigma v \in L \\ v \in \Sigma^*, |v| = N}} \sqrt{p_{\sigma v}} A_v K_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma v}(t) + \sqrt{p_{\sigma v}} A_v B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma v}(t) \tag{2.10}$$

where the infinite sum on the right-hand side is convergent in the mean square sense, see Appendix 2.9 for more details on the convergence of the right-hand side of (2.10). In fact, even the order of summation in the right-hand side of (2.10) can be changed, see Appendix 2.9 for more details.

Equation (2.10) is a direct generalization of the well-known formula for the state of a stochastic LTI system. More precisely, consider the following LTI system:

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{v}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t) + F\mathbf{v}(t) \end{cases}$$

where A is Schur[1] to ensure the stability of the system. Then, by [56],

$$\mathbf{x}(t) = \sum_{k=1}^{\infty} \Big(A^{k-1}K\mathbf{v}(t-k) + A^{k-1}B\mathbf{u}(t-k) \Big).$$

Since the state of a sGLSS is uniquely determined by its noise, input process and state matrices - as shown by the structure of (2.10) - we now introduce the following notations:

Notation 2.5. *We identify the sGLSS $\mathcal{S}$ of the form* (2.1) *with the tuple*

$$\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v}, \mathbf{u}, \boldsymbol{\mu}).$$

When $\mathbf{u}$ and $\boldsymbol{\mu}$ are clear from the context, we use the notation

$$\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v}).$$

We identify the asGLSS $\mathcal{S}_a$ of the form (2.4) *with the tuple*

$$\mathcal{S}_a = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v}, \mathbf{u}, \boldsymbol{\mu}).$$

When $\mathbf{u}$ and $\boldsymbol{\mu}$ are clear from the context, we use the notation

$$\mathcal{S}_a = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v}).$$

In the terminology of [91], a sGLSS and an asGLSS correspond both to a stationary **GBS** w.r.t. inputs $\{\boldsymbol{\mu}_\sigma\}_{\sigma\in\Sigma}$ with the noise process $\mathbf{w} = \begin{bmatrix}\mathbf{v}^T & \mathbf{u}^T\end{bmatrix}$ and $\mathbf{v}$ respectively. Note that the processes $\mathbf{x}$ and $\mathbf{y}$ are ZMWSSI, in particular, they are wide-sense stationary, and that $\mathbf{x}$ is orthogonal to the future values of the noise process $\mathbf{v}$. We focus on wide-sense stationary processes, because it is difficult to estimate the distribution of non-stationary processes. Moreover, wide-sense stationary processes solve the problem of the initial state conditions.

In the sequel, we will be interested in GLSS as models representing a certain input-output behavior. More precisely, we formalize the input-output behavior of a sGLSS as triples $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. This prompts us the following definition:

[1] all its eigenvalues are inside the unit disk

Definition 2.8 (Minimal realization)**.** *If $\mathcal{S}$ is of the form* (2.1)*, then we call the state-space dimension n_x the* dimension *of $\mathcal{S}$ and we denote it by* $\dim \mathcal{S}$*. We say that the sGLSS $\mathcal{S}$ is a* minimal *realization of $(\tilde{\mathbf{y}}, \tilde{\mathbf{u}}, \tilde{\boldsymbol{\mu}})$, if for any sGLSS realization $\mathcal{S}^{'}$ of $(\tilde{\mathbf{y}}, \tilde{\mathbf{u}}, \tilde{\boldsymbol{\mu}})$, the dimension of $\mathcal{S}^{'}$ is not smaller than* $\dim \mathcal{S}$*. This is also applicable for the class of asGLSS with no input* ($\mathbf{u} = 0$)*: we say that the asGLSS $\mathcal{S}$ is a* minimal *realization of $(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\mu}})$, if for any asGLSS realization $\mathcal{S}^{'}$ of $(\tilde{\mathbf{y}}, \tilde{\boldsymbol{\mu}})$,* $\dim(\mathcal{S}^{'}) \geq \dim \mathcal{S}$*.*

We can formulate the realization and identification problems for GLSS as follows.

Problem 1 (Realization problem)**.** *For the process $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, assuming that it has a sGLSS realization $\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v}, \boldsymbol{\mu})$, find, based on $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, a minimal sGLSS realization $\mathcal{S}^m = (\{A_\sigma^m, B_\sigma^m, K_\sigma^m\}_{\sigma=1}^{n_\mu}, C^m, D^m, F^m, \mathbf{v}, \boldsymbol{\mu})$ of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.*

Problem 2 (Identification problem)**.** *Assume that y, u, μ are sample paths of $\mathbf{y}, \mathbf{u}, \boldsymbol{\mu}$ respectively, and assume that there exists a sGLSS $\mathcal{S}$ of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. Compute from the samples $\mathcal{D} = \{y(t), u(t), \mu(t)\}_{t=1}^{N}$ the estimates $\{\{\hat{A}_i^N, \hat{B}_i^N, \hat{K}_i^N, \hat{Q}_i^N\}_{i=1}^{n_\mu}, \hat{C}^N, \hat{D}^N, \hat{F}^N\}$ of $\hat{\mathcal{S}}$, such that as $N \to \infty$, these estimated matrices converge to matrices $\{\{A_i, B_i, K_i, Q_i\}_{i=1}^{n_\mu}, C, D, F\}$ of the true system $\mathcal{S}$ respectively.*

Note that the realization problem is essentially an idealized version of the identification problem, where we assume that we know the stochastic processes involved, rather than only a sample path of each of them. The motivation for considering the realization problem is that its solution helps to analyze the solution of the identification problem. The realization theory helps to solve the identification problem above on several levels.

2.6 Motivation for stochastic realization theory of GLSS

The goals of realization theory can be listed as follows:

- Find a condition for the existence of a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.
- Find an algorithm for computing a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ for $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.
- Define a suitable notion of innovation form for sGLSS. By a suitable notion we mean that **(a)** the sGLSS in innovation form can be interpreted as an optimal predictor of current output based on past outputs, inputs and switching signals; **(b)** any sGLSS can be transformed to a sGLSS in innovation form.
- Propose a covariance realization algorithm which uses covariances of outputs and inputs to compute a sGLSS realization. Then, show that the covariance realization algorithm returns a minimal dimensional sGLSS in innovation form.
- Show that minimal sGLSS in innovation form (with the same output) are isomorphic.

The motivations for these goals are the following headlines:

M1 - Stochastic systems as predictors of current output based on past outputs, inputs and switching: In order to find a GLSS $\hat{\mathcal{S}}$ realizing $(\mathbf{y}, \boldsymbol{\mu}, \mathbf{u})$, a widespread approach is to interpret it as parametrization of a predictor, which maps past inputs and outputs to current output. Then, finding $\hat{\mathcal{S}}$ boils down to finding the best predictor. However, while the parameters of a GLSS can be used to compute such a predictor, apriori it is not clear that this correspondence is one-to-one and that the resulting predictor is a finite dimensional dynamical system of GLSS type.
Indeed, a naive way to construct such a predictor would be to construct a Kalman-filter based on GLSS. More precisely, for deterministic switching, we could view GLSS as a time-varying linear system and apply classical Kalman-filtering. However, from [20], we know that the resulting system would have matrices which depend on all the past switching, even if the switching is deterministic.
For stochastic switching, applying a Kalman-filtering is complicated. However, if we show that any GLSS can be converted to a minimal GLSS in innovation form, then it can naturally be interpreted as an optimal predictor. Moreover, there is a one-to-one (up to isomorphism) correspondence matrices of such a GLSS and parameters of the predictor.

M2 - Consistent estimator (predictor) means good model: A consistent estimator in system identification ensures that, as more data is collected, the estimated GLSS model $\hat{\mathcal{S}}$ converges to the true GLSS $\mathcal{S}$ for the specific input and switching signals used during the identification experiment. However, consistency alone does not guarantee that $\hat{\mathcal{S}}$ will behave like $\mathcal{S}$ for all possible inputs and switching sequences. This creates a blind spot: models may appear statistically valid while failing to capture the true system's behavior outside the scope of the identification data.
This limitation arises from the nature of stochastic identification, where one attempts to reconstruct a deterministic system's behavior from noisy and finite observations. In fact, two GLSS models may produce the same outputs for the inputs and switching signals used during identification but diverge significantly for others. This issue will be addressed in Chapter 4 and illustrated in the corresponding Numerical Example section.
The proposed solution relies on stochastic realization theory, which provides a rigorous framework for ensuring that the identified model reproduces the true system's complete input-output behavior—not just its behavior on previously observed data.
In the LTI case, this issue was resolved by assuming that both the data generator in (2.1) and the stochastic system corresponding to $\hat{\mathcal{S}}$ are minimal and in *innovation form*. In this setting, two minimal systems in innovation form with the same input-output behavior are *isomorphic* [65]. This means that the identified model faithfully reproduces the system's behavior for any input, not just those used during identification.
Realization theory allows us to avoid this blind spot by using minimal realizations in innovation form. Suppose that estimated model is a minimal sGLSS in innovation form, and suppose that it is a realization of $(\mathbf{y}, \boldsymbol{\mu}, \mathbf{u})$, i.e., $\hat{\mathcal{S}}$ generates the same output as the true system $\mathcal{S}$ for the designated input and switching used for the identification experiment. Assume that $\hat{\mathcal{S}}$ is also minimal and in innovation form. Then the matrices of $\mathcal{S}$ and $\hat{\mathcal{S}}$ are the same after a basis transformation. In

other words, when the estimated model perfectly matches the true system's outputs on training data, we achieve structural equivalence up to state-space transformations - meaning the model will generalize perfectly. In particular, $\mathcal{S}$ and $\hat{\mathcal{S}}$ will generate the same output for any input and switching, and not only for the ones used in the identification experiment.
However, in practice, $\hat{\mathcal{S}}$ only approximates the true output $\mathbf{y}$. In this case the argument above does not hold but can still be extended. More precisely, the system matrices of $\hat{\mathcal{S}}$ are continuous functions of the output covariances, due to the structure of the covariance realization algorithm. Hence, intuitively, if $\hat{\mathcal{S}}$ generates an output close to the true one, then its matrices are close to the matrices of the true system $\mathcal{S}$ after a suitable state-space transformation. In particular, the output of $\hat{\mathcal{S}}$ to any input and switching (not only the ones used in the identification experiment) will be close to that of the true system $\mathcal{S}$.

M3 - Identifiability: In system identification, a fundamental question is whether a given model structure can uniquely determine the system parameters from input-output data. We know from [3, 2], that a parametrization is globally identifiable if no two different parameter vectors produce the same input-output map. Non-identifiable models lead to ill-posed identification problems, where multiple parameter sets produce the same input-output behavior, complicating parameter estimation and control design. For GLSSs, this question is complicated because traditional LTI identifiability tools fail when applied to GLSS models because frozen LTI models (obtained by fixing the switching signal) may not capture the full dynamics of the GLSS system. Moreover, in [3, 2], the authors state that for LPV systems with affine dependencies on the scheduling, which are a subclass of the GLSS class, parametrization is identifiable if:

(1) the parametrization is injective (no two parameters give the same model),

(2) no two models with different parameters are related by an isomorphism.

We expect that the same reasoning can be extended to GLSSs. To this end, more investigations in realization theory are needed to ensure that minimal GLSS, producing the same outputs, are isomorphic.

M4 - Covariance realization algorithm leads to a subspace system identification algorithm: As it was mentioned in Chapter 1, system identification methods can be divided into parametric and subspace methods. For LTI systems, subspace methods relied heavily on realization theory. First, many subspace methods were based on covariance realization algorithm. Second, realization theory was instrumental for analyzing consistency of subspace methods. Consistency here refers to the property that, as the amount of data grows, the estimated parameters converge to the true system parameters.
We expect to reap the same benefits from realization theory for GLSS systems. In particular, the formulation of a covariance realization algorithm will yield a simple subspace method, if it is applied to empirical covariances instead of the true ones. In addition, as the empirical covariances converge to the true ones , the GLSS obtained from applying covariance realization algorithm to empirical covariances is also expected to converge to the true system. Moreover, it will be minimal and in innovation form.

Motivated by the discussion above, in this thesis, we develop a realization theory for the GLSS class with inputs. Our approach is inspired by classical realization theory for LTI. In particular, a key step in the proof of existence and isomorphism of minimal LTI systems in innovation form was the decomposition $\mathbf{y}(t) = \mathbf{y}^d(t) + \mathbf{y}^s(t)$ of the output into two components. The component $\mathbf{y}^d$ is determined by the input and it is the output of a noiseless LTI system. The component $\mathbf{y}^s$ is the output of an autonomous LTI system and it depends only on the noise. Then the results on minimal LTI systems in innovation form follow by applying realization theory [65] to the autonomous LTI system generating $\mathbf{y}^s$. The existence of this decomposition will be proven for the GLSS case in Chapter 3, in order to prove the existence of minimality and uniqueness of the GLSS realization. These results are necessary to introduce the realization algorithm and the identification algorithm in Chapter 5.

Nonetheless, in order to present all these contributions in the following chapters, we have to recall some prior results, concerning the autonomous case, that existed already in the literature.

2.7 Systematic overview: existence and minimality of asGLSS in innovation form

In this section, we review the principal results on existence and minimality of asGLSS in innovation form. To this end, in Subsection 2.7.1 we recall from [83, 81, 86] some results on realization theory of deterministic Linear Switched System (abbreviated as LSS). In Subsection 2.7.2 we present the definition of asGLSSs in innovation form, and in Subsection 2.7.3 we present results on existence and uniqueness of minimal asGLSSs. In Subsection 2.7.4 we present rank conditions for minimality of asGLSSs and an algorithm for converting any asGLSS to a minimal one in innovation form. The results presented in this section follow from [91], but they have never been formulated explicitly for GLSSs.

2.7.1 Deterministic realization theory of LSS

Let us consider a deterministic counterpart of sGLSS. We will consider only the case where $\boldsymbol{\mu}(t)$, for any time instance t, takes values in $\{0, 1\}$, as it is sufficient for our purpose. For this case, deterministic Linear Switched systems (dLSS) are the natural deterministic counterparts of GLSSs. A dLSS is a system described as follows:

$$\mathscr{S}\begin{cases} \mathrm{x}(t+1) = \mathcal{A}_{q(t)}\mathrm{x}(t) + \mathcal{B}_{q(t)}\mathrm{u}(t) \\ \mathrm{y}(t) = \mathcal{C}\mathrm{x}(t) + \mathcal{D}\mathrm{u}(t) \end{cases}, \tag{2.11}$$

where $\{\mathcal{A}_\sigma, \mathcal{B}_\sigma\}_{\sigma\in\Sigma}, \mathcal{C}, \mathcal{D}$ are matrices of suitable dimensions, $q : \mathbb{Z} \to \Sigma$ is the switching signal, $\mathrm{x} : \mathbb{Z} \to \mathbb{R}^{n_\mathrm{x}}$ is the state trajectory $\mathrm{u} : \mathbb{Z} \to \mathbb{R}^{n_u}$ is the input trajectory $\mathrm{y} : \mathbb{Z} \to \mathbb{R}^{n_\mathrm{y}}$ is the output trajectory with left bounded support[2]. We identify a dLSS of the form (2.11) with the tuple

$$\mathscr{S} = (\{\mathcal{A}_i, \mathcal{B}_i\}_{i=1}^{n_\mu}, \mathcal{C}, \mathcal{D}). \tag{2.12}$$

We refer to [83, 81] for a detailed definition input-output behavior of dLSSs, their realization theory, minimality, etc. In particular, let us call any function $M : \Sigma^* \to \mathbb{R}^{n_y \times n_u}$ a *Markov function*. The dLSS $\mathscr{S}$ *realizes* M, if for all $w \in \Sigma^*$,

$$M(w) = \begin{cases} \mathcal{C}\mathcal{A}_s\mathcal{B}_\sigma, & w = \sigma s,\ \sigma \in \Sigma,\ s \in \Sigma^* \\ \mathcal{D} & w = \epsilon \end{cases}. \tag{2.13}$$

If $\mathscr{S}$ is a realization of M, then M is referred to as the *Markov function of* $\mathscr{S}$ and it is denoted by $M_{\mathscr{S}}$. The values $\{M(w)\}_{w\in\Sigma^*}$ of $M_{\mathscr{S}}$ are the *Markov parameters* of $\mathscr{S}$. From [83, 81] it then follows that two dLSSs have the same input-output behavior, if and only if their Markov functions are equal. For the dLSS in (2.11) the integer n_x is called the *dimension* of $\mathscr{S}$, and we say that a dLSS is *minimal*, if there exists no other dLSS of smaller dimension which realizes the same Markov-function. From [83, 81], it follows that a dLSS is minimal if and only if it is reachable and observable, and the latter properties are equivalent to rank conditions of the extended reachability and observability matrices [81, Definition 23, Theorem 1, Theorem 2]. Furthermore, any dLSS can be transformed to a minimal one with the same Markov function, using a minimization algorithm [81, Procedure 3 (Minimization)][3]. For a more detailed discussion see [81]. Moreover, any two minimal dLSS which are input-output equivalent are isomorphic, i.e. they are related by a linear change of coordinates [81, Theorem 1].

We should note that, in [86], we can find these result extended to the class of LPV systems with affine dependence in parameters, which is another subclass of the GLSS class. In addition, because the dLSS class is a subclass of LPV, and which is equivalent to LPV in terms of realization theory, we will use the results of [86] to dLSS systems, in what follows in this subsection.

From [86], we notice that minimality of dLSSs can be characterized via rank conditions for suitable matrices. These matrices are counterpart of the well-known controllability and observability matrices for linear systems. More precisely, consider a dLSS $\mathscr{S}$ of the form (2.11) and recall from [86, Definition 1] the definition of the i-step extended reachability matrices $\mathcal{R}_{\mathscr{S},i}$ and the i-step observability matrices $\mathcal{O}_{\mathscr{S},i}$ of $\mathscr{S}$: $\mathcal{R}_{\mathscr{S},0} = \begin{bmatrix}\mathcal{B}_1 & \mathcal{B}_2 & \dots & \mathcal{B}_{n_\mu}\end{bmatrix}$, $\mathcal{O}_{\mathscr{S},0} = \mathcal{C}^T$, and for all $i > 1$

$$\begin{aligned} \mathcal{R}_{\mathscr{S},i} &= \begin{bmatrix}\mathcal{R}_{\mathscr{S},0} & \mathcal{A}_1\mathcal{R}_{\mathscr{S},i-1} & \dots & \mathcal{A}_{n_\mu}\mathcal{R}_{\mathscr{S},i-1}\end{bmatrix} \\ \mathcal{O}_{\mathscr{S},i} &= \begin{bmatrix}\mathcal{O}_{\mathscr{S},0}^T & \mathcal{A}_1^T\mathcal{O}_{\mathscr{S},i-1}^T & \dots & \mathcal{A}_{n_\mu}^T\mathcal{O}_{\mathscr{S},i-1}^T\end{bmatrix}^T \end{aligned} \tag{2.14}$$

Then the following holds

Theorem 2.1 (Rank Conditions [86]). *The dLSS $\mathscr{S}$ of dimension n_x is minimal, if and only if rank$\{\mathcal{R}_{\mathscr{S},n_x-1}\} = n_x$, and rank$\{\mathcal{O}_{\mathscr{S},n_x-1}\} = n_x$.*

[2] a function $g : \mathbb{Z} \longrightarrow \mathbb{R}^p$ has a left bounded support, then it $\exists\ t_0 \in \mathbb{Z}$, such that $\forall\ t < t_0$, $g(t) = 0$.

[3] [81, Corollary 1] should be applied with zero initial state

Remark 2.4. *We mention that, if a dLSS is reachable, this does not mean that his linear subsystems are individually reachable. In [79] the following counter example was presented:*

Example 2.10. *Let $n_\mu = 2$, $n_x = 3$, $n_u = 1$ and $n_y = 2$.*

$$A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$
$$A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \textit{ and } \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let us compute the reachability matrices $\mathcal{R}_{\mathscr{S},0}, \mathcal{R}_{\mathscr{S},1}$ and $\mathcal{R}_{\mathscr{S},n_x-1} = \mathcal{R}_{\mathscr{S},2}$ as follows:

$$\mathcal{R}_{\mathscr{S},0} = \begin{bmatrix} B_1 & B_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_1\mathcal{R}_{\mathscr{S},0} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2\mathcal{R}_{\mathscr{S},0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then:

$$\mathcal{R}_{\mathscr{S},1} = \begin{bmatrix} \mathcal{R}_{\mathscr{S},0} & A_1\mathcal{R}_{\mathscr{S},0} & A_2\mathcal{R}_{\mathscr{S},0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$A_1\mathcal{R}_{\mathscr{S},1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad A_2\mathcal{R}_{\mathscr{S},1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then rank$\{\mathcal{R}_{\mathscr{S},2}\}$ = rank$\{[\mathcal{R}_{\mathscr{S},1} \quad A_1\mathcal{R}_{\mathscr{S},1} \quad A_2R_{\mathscr{S},1}]\} = 3$. This means that the GLSS is reachable. However, note that neither of the individual subsystems is reachable on its own:

Subsystem 1: *The reachability matrix of (A_1, B_1) is*

$$\mathcal{R}_1 = \begin{bmatrix} B_1 & A_1B_1 & A_1^2B_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

with its rank equals $2 < n_x = 3$. Hence, subsystem 1 is not reachable.

Subsystem 2: *Since $B_2 = 0$, subsystem 2 is trivially not reachable.*

The discussion above suggests the following algorithm for checking if a dLSS is minimal.

Algorithm 1: Checking minimality,
Input : The dLSS $\mathscr{S} = (\{\mathcal{A}_i, \mathcal{B}_i\}_{i=1}^{n_\mu}, \mathcal{C}, \mathcal{D})$.
Output : Return 1 if $\mathscr{S}$ is minimal and 0 if it is not.

1: Compute the $(n_x - 1)$-step reachability matrix $\mathcal{R}_{\mathscr{S},n_x-1}$ using the recursion (2.14).
2: Compute the $(n_x - 1)$-step observability matrix $\mathcal{O}_{\mathscr{S},n_x-1}$ using the recursion (2.14).
3: Check the rank of $\mathcal{R}_{\mathscr{S},n_x-1}$ and $\mathcal{O}_{\mathscr{S},n_x-1}$. If both ranks are equal to n_x, then $\mathscr{S}$ is minimal, otherwise, $\mathscr{S}$ is not minimal.

Note that the number of columns (respectively rows) of the $(n_x - 1)$-step extended reachability matrix (resp. $n_x - 1$-step extended observability matrix) grows as $O(n_\mu^{n_x-1})$, where $O(\cdot)$ is the Big-O (Landau) notation. However, using the results of [10, Algorithm 1 - Algorithm 2], we can compute a matrix representation of the image of the $n_x - 1$ extended reachability and observability matrices in polynomial time and space. As Algorithm 1 requires only the knowledge of the image of the matrices $\mathcal{R}_{\mathscr{S},n_x-1}$ and $\mathcal{O}^T_{\mathscr{S},n_x-1}$ and not the knowledge of the matrices themselves, it can be modified to have polynomial complexity (see [10, Algorithm 1 - Algorithm 2]). Furthermore, any dLSS can be transformed to a minimal one with the same sub-Markov function, using Kalman decomposition [86, Corollary 1][4]. Subsequently, Algorithm 2 can be formulated.

Algorithm 2: Minimization algorithm,
Input: The dLSS $\mathscr{S} = (\{\mathcal{A}_i, \mathcal{B}_i\}_{i=1}^{n_\mu}, \mathcal{C}, \mathcal{D})$ representation matrices.
Output: Return minimal dLSS $\mathscr{S}_m$.

1: Choose a basis $\{b_i\}_{i=1}^{n_\mathrm{x}} \subset \mathbb{R}^{n_\mathrm{x}}$ such that $\mathrm{Span}\{b_1, \ldots, b_r\} = \mathrm{Im}\{\mathcal{R}_{n_\mathrm{x}-1}\}$ and $\mathrm{Span}\{b_{r_m+1}, \ldots, b_r\} = (\mathrm{Im}\{\mathcal{R}_{n_\mathrm{x}-1}\} \cap \ker\{\mathcal{O}_{n_\mathrm{x}-1}\})$ for some $r, r_m \geq 0$.
2: Define $\mathcal{T} = \begin{bmatrix} b_1 & b_2 & \ldots & b_{n_\mathrm{x}} \end{bmatrix}^{-1}$, and let $\hat{\mathcal{A}}_i = \mathcal{T}\mathcal{A}_i\mathcal{T}^{-1}$, $\hat{B}_i = \mathcal{T}\mathcal{B}_i, \hat{\mathcal{C}} = \mathcal{C}\mathcal{T}^{-1}$, $i = 1, \ldots, n_\mu$. Then

$$\hat{\mathcal{A}}_i = \begin{bmatrix} \mathcal{A}_i^{\mathrm{m}} & 0 & A_i'' \\ \mathcal{A}_i' & \hat{\mathcal{A}}' & \mathcal{A}_i''' \\ 0 & 0 & \mathcal{A}_i'''' \end{bmatrix}, \ \hat{\mathcal{B}}_i = \begin{bmatrix} \mathcal{B}_i^{\mathrm{m}} \\ \mathcal{B}_i' \\ 0 \end{bmatrix}, \ \hat{\mathcal{C}} = \begin{bmatrix} (\mathcal{C}^{\mathrm{m}})^T \\ 0 \\ (\mathcal{C}')^T \end{bmatrix}^T, \tag{2.15}$$

where $\mathcal{A}_i^{\mathrm{m}} \in \mathbb{R}^{r_m \times r_m}, \mathcal{B}_i^{\mathrm{m}} \in \mathbb{R}^{r_m \times n_\mathrm{u}}$, $\mathcal{A}_i''' \in \mathbb{R}^{(n-r)\times(n-r)}$, $\mathcal{B}_i' \in \mathbb{R}^{(r-r_m)\times n_\mathrm{u}}$ $i = 1, \ldots, n_\mu$, and $\mathcal{C}^{\mathrm{m}} \in \mathbb{R}^{n_\mathrm{y} \times r_m}$, , $\mathcal{C}' \in \mathbb{R}^{n_\mathrm{y} \times (n-r)}$.
3: Let $\mathscr{S}_m = \{\mathcal{A}_i^{\mathrm{m}}, \mathcal{B}_i^{\mathrm{m}}\}_{i=1}^{n_\mu}, \mathcal{C}^{\mathrm{m}}, \mathcal{D})$

For the convenience of the reader, we will present [86, Corollary 1] with our terminology:

Corollary 2.1 ([86])**.** *With the notation of* (2.15)*, the dLSS* $\mathscr{S}_m = \{\mathcal{A}_i^{\mathrm{m}}, \mathcal{B}_i^{\mathrm{m}}\}_{i=1}^{n_\mu}, \mathcal{C}^{\mathrm{m}}, \mathcal{D})$ *is minimal.*

[4] [86, Corollary 1] should be applied with zero initial state

Example 2.11. *To illustrate the algorithm above, consider the following dLSS $\mathscr{S}$ from [86], where $n_x = 3$, $n_\mu = 2$, $n_u = 1$ and $n_y = 2$:*

$$A_0 = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix}, \quad B_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$
$$A_1 = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -1 & -1 \\ 2 & -2 & -2 \end{bmatrix},$$

and some matrix $\mathcal{D} \in \mathbb{R}^{n_y \times n_u} = \mathbb{R}^{2\times 1}$. After calculating $\mathcal{R}_{\mathscr{S},2}$ and $\mathcal{O}_{\mathscr{S},2}$, we can see that rank$\{\mathcal{R}_{\mathscr{S},2}\} = 2$ and that rank$\{\mathcal{O}_{\mathscr{S},2}\} = 1$. Now, we can find a suitable base by taking

$$b_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad \textit{and} \quad b_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We can easily see that $\{b_1, b_2\}$ span $\mathrm{Im}\{\mathcal{R}_{\mathscr{S},2}\}$ and that b_2 spans $\mathrm{Im}\{\mathcal{R}_{\mathscr{S},2}\} \cap \ker\{\mathcal{O}_{\mathscr{S},2}\}$. After that, we can apply the transformation by taking

$$\mathcal{T} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

and $\hat{\mathcal{A}}_i = \mathcal{T}\mathcal{A}_i\mathcal{T}^{-1}$, $\hat{B}_i = \mathcal{T}\mathcal{B}_i, \hat{\mathcal{C}} = \mathcal{C}\mathcal{T}^{-1}$, $i = 1, 2$. These matrices are as of the form of (2.15). *We can then deduce that*

$$\mathcal{A}_1^{\mathrm{m}} = 2,\ \mathcal{A}_2^{\mathrm{m}} = 3,\ \mathcal{B}_1^{\mathrm{m}} = 1,\ \mathcal{B}_2^{\mathrm{m}} = -2,\ \mathcal{C}^{\mathrm{m}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Note that the matrix $\mathcal{D}$ stays the same, because it has no effect, neither on the observability nor on the reachability matrix. Also, it should keep the same dimension respecting the corresponding input and outputs. That said, by Corollary 2.1, $\mathscr{S}_m = \{\mathcal{A}_i^{\mathrm{m}}, \mathcal{B}_i^{\mathrm{m}}\}_{i=1}^2, \mathcal{C}^{\mathrm{m}}, \mathcal{D})$ is a minimal realization of the input-output behavior of the original dLSS $\mathscr{S}$.

2.7.2 Definition of asGLSS in innovation form

Next, we define what we mean by asGLSS in innovation form. To this end, we need Notation 2.1 for orthogonal projection from Section 2.3.2.

This said, considering an asGLSS of the form (2.4), we define the *innovation process* $\mathbf{e}$ *of* $\mathbf{y}$ *with respect to* $\boldsymbol{\mu}$ as follows:

$$\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}] \tag{2.16}$$

In other words, $\mathbf{e}(t)$ is the difference between the output and its projection on its past values w.r.t. the switching process $\boldsymbol{\mu}$, i.e., $\mathbf{e}$ is the best predictor of $\mathbf{y}$ using the product of the output and switching past values from (2.8).

Example 2.12. *Consider the following non-linear model, where $\Sigma = \{1, 2\}$:*

$$\mathbf{y}(t) = a_2\boldsymbol{\mu}_2(t-1)\mathbf{y}(t-1) + a_1\boldsymbol{\mu}_1(t-1)\mathbf{y}(t-1) + \mathbf{v}(t),$$
$$\boldsymbol{\mu}(t) = \begin{bmatrix} \boldsymbol{\mu}_1(t) \\ \boldsymbol{\mu}_2(t) \end{bmatrix},$$

where $\boldsymbol{\mu}$ is i.i.d. process and $\boldsymbol{\mu}(t)$ is independent of $\mathbf{y}(s)$ for $t > s \in \mathbb{Z}$. Also, $\mathbf{v}(t)$ is independent of $\boldsymbol{\mu}(s), \mathbf{y}(s)$ for $t > s \in \mathbb{Z}$, and $\mathbf{v}$ is zero mean i.i.d. process. Next, let us define the state process:

$$\mathbf{x}(t) := a_2\boldsymbol{\mu}_2(t-1)\mathbf{y}(t-1) + a_1\boldsymbol{\mu}_1(t-1)\mathbf{y}(t-1).$$

Then, we can write

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$$

and

$$\begin{aligned}\mathbf{x}(t+1) &= a_2\boldsymbol{\mu}_2(t)\mathbf{y}(t) + a_1\boldsymbol{\mu}_1(t)\mathbf{y}(t) \\ &= a_2\boldsymbol{\mu}_2(t)\mathbf{x}(t) + a_1\boldsymbol{\mu}_1(t)\mathbf{x}(t) + a_2\boldsymbol{\mu}_2(t)\mathbf{v}(t) + a_1\boldsymbol{\mu}_1(t)\mathbf{v}(t).\end{aligned}$$

Then, we can write

$$\begin{cases} \mathbf{x}(t+1) & = \sum_{\sigma=1}^{2} \Big(a_\sigma\mathbf{x}(t) + k_\sigma\mathbf{v}(t)\Big)\boldsymbol{\mu}_\sigma(t) \\ \mathbf{y}(t) & = \mathbf{x}(t) + \mathbf{v}(t) \end{cases}$$

where, in this example, $k_\sigma = a_\sigma$. This leads to

$$\mathbf{x}(t) = \sum_{\sigma=1}^{2} \Big((a_\sigma - k_\sigma)\mathbf{x}(t-1) + k_\sigma\mathbf{y}(t-1)\Big)\boldsymbol{\mu}_\sigma(t-1). \tag{2.17}$$

In what follows, for the ZMWSSI process $\mathbf{y}$, let $\mathcal{H}_t^{\mathbf{y}}$ be the closed subspace of $\mathcal{H}_1$ (see Notation 2.1) generated by the entries of $\{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}$. By construction, we know that $\mathbf{x}(t) \in \mathcal{H}_t^{\mathbf{y}}$, in particular:

$$\begin{aligned}\mathbf{x}(t) &= a_2\boldsymbol{\mu}_2(t-1)\mathbf{y}(t-1) + a_1\boldsymbol{\mu}_1(t-1)\mathbf{y}(t-1) \\ &= a_2\sqrt{p_2}\mathbf{z}_2^{\mathbf{y}}(t) + a_1\sqrt{p_1}\mathbf{z}_1^{\mathbf{y}}(t).\end{aligned}$$

Moreover, $\mathbf{v}(t)$ is orthogonal to $\mathcal{H}_t^{\mathbf{y}}$. Hence

$$\begin{aligned}E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}] &= E_l[\mathbf{y}(t) \mid \mathcal{H}_t^{\mathbf{y}}] = E_l[\mathbf{x}(t) \mid \mathcal{H}_t^{\mathbf{y}}] + E_l[\mathbf{v}(t) \mid \mathcal{H}_t^{\mathbf{y}}] \\ &= E_l[\mathbf{x}(t) \mid \mathcal{H}_t^{\mathbf{y}}] + 0 = \mathbf{x}(t)\end{aligned}$$

In other words, $\mathbf{x}(t)$ is the projection of $\mathbf{y}(t)$ onto the subspace $\{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}$. In this case, if we want to calculate $\mathbf{e}(t)$, we will get

$$\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}] = \mathbf{y}(t) - \mathbf{x}(t) = \mathbf{v}(t).$$

We can say that the noise process $\mathbf{v}$ is the innovation process.

Example 2.13. *A more straightforward case is the linear case, where $\Sigma = \{1\}$, and $\boldsymbol{\mu}_1(t) = 1$ for all $t \in \mathbb{Z}$:*

$$\mathbf{y}(t) = a\mathbf{y}(t-1) + (k-a)\mathbf{v}(t) + \mathbf{v}(t)$$

for some $a, k \in \mathbb{R}$, $|a| < 1, |a-k| < 1$. Next, we define:

$$\mathbf{x}(t) \coloneqq a\mathbf{y}(t-1) + (k-a)\mathbf{v}(t-1)$$

Then, one can easily see that

$$\begin{aligned}
\mathbf{y}(t) &= \mathbf{x}(t) + \mathbf{v}(t), \\
\mathbf{x}(t+1) &= a\mathbf{y}(t) + (k-a)\mathbf{v}(t) = a\mathbf{x}(t) + a\mathbf{v}(t) + (k-a)\mathbf{v}(t) = a\mathbf{x}(t) + k\mathbf{v}(t), \\
\mathbf{v}(t) &= \mathbf{y}(t) - \mathbf{x}(t).
\end{aligned}$$

We can then write:

$$\begin{aligned}
\mathbf{x}(t+1) &= a\mathbf{x}(t) + k\big(\mathbf{y}(t) - \mathbf{x}(t)\big), \\
&= (a-k)\mathbf{x}(t) + k\mathbf{y}(t) \\
&= (a-k)\big(a\mathbf{x}(t-1) + k\mathbf{v}(t-1)\big) + k\mathbf{y}(t) \\
&= (a-k)\big(a\mathbf{x}(t-1) + k\mathbf{y}(t-1) - k\mathbf{x}(t-21)\big) + k\mathbf{y}(t) \\
&= (a-k)^2\mathbf{x}(t-1) + (a-k)k\mathbf{y}(t-1) + k\mathbf{y}(t) \\
&= (a-k)^2\big(a\mathbf{x}(t-2) + k\mathbf{v}(t-2)\big) + (a-k)k\mathbf{y}(t-1) + k\mathbf{y}(t) \\
&= (a-k)^2\big(a\mathbf{x}(t-2) + k\mathbf{y}(t-2) - k\mathbf{x}(t-2)\big) + (a-k)k\mathbf{y}(t-1) + k\mathbf{y}(t) \\
&= (a-k)^3\mathbf{x}(t-2) + (a-k)^2k\mathbf{y}(t-2) + (a-k)k\mathbf{y}(t-1) + k\mathbf{y}(t) \\
&= \dots
\end{aligned}$$

We can then conclude that:

$$\begin{aligned}
\mathbf{x}(t) &= \sum_{l=1}^{\infty}(a-k)^{l-1}k\mathbf{y}(t-l), \\
\mathbf{y}(t) &= \sum_{l=1}^{\infty}(a-k)^{l-1}k\mathbf{y}(t-l) + \mathbf{v}(t).
\end{aligned}$$

In this case, when we calculate the innovation process $\mathbf{e}(t)$, we will get

$$\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{z}^{\mathbf{y}}_w(t)\}_{w\in\Sigma^+}] = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \{\mathbf{y}(t-1), \mathbf{y}(t-2), \dots\}].$$

One can easily see that the projection of $\mathbf{y}(t)$ onto the past values of $\mathbf{y}$ is $\mathbf{x}(t)$, i.e.,

$$\mathbf{x}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{y}(t-1), \mathbf{y}(t-2), \dots\}].$$

Then, $\mathbf{e}(t) = \mathbf{y}(t) - \mathbf{x}(t) = \mathbf{v}(t)$. We finally conclude that the noise process $\mathbf{v}$ is the innovation process.

Definition 2.9 (asGLSS in innovation form)**.** *An asGLSS of the form* (2.4) *is said to be in* innovation form*, if it is a realization of* $(\mathbf{y}, \boldsymbol{\mu})$*,* $F = I_{n_y}$ *and* $\mathbf{v}$ *is the innovation process of* $\mathbf{y}$*, i.e.,* $\mathbf{v} = \mathbf{e}$*.*

The representation of such system can be described as follows:

$$\mathcal{S}\begin{cases}\mathbf{x}(t+1) = \displaystyle\sum_{i=1}^{n_\mu}(A_i\mathbf{x}(t) + K_i\mathbf{e}(t))\boldsymbol{\mu}_i(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + I_{n_y}\mathbf{e}(t)\end{cases} \tag{2.18}$$

2.7.3 Existence and uniqueness of minimal asGLSS in innovation form

Let $\mathcal{S}$ be an asGLSS of the form (2.18) in innovation form. Let $\widetilde{\mathcal{S}} = (\{\widetilde{A}_\sigma, \widetilde{K}_\sigma, \}_{\sigma=1}^{n_\mu}, \widetilde{C}, I_{n_y}, \mathbf{e})$ be another asGLSS of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form. We say that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are *isomorphic*, if there exists a non-singular matrix T (referred to as *isomorphism from* $\mathcal{S}$ *to* $\widetilde{\mathcal{S}}$) such that

$$\forall\sigma \in \Sigma : \widetilde{A}_\sigma = TA_\sigma T^{-1},\ \widetilde{K}_\sigma = TK_\sigma,\ \text{and}\ \ \widetilde{C} = CT^{-1}.$$

Note that if T is an isomorphism from $\mathcal{S}$ and $\widetilde{S}$ and $\mathbf{x}$ and $\widetilde{\mathbf{x}}$ are the state processes of $\mathcal{S}$ and $\widetilde{\mathcal{S}}$, then $\widetilde{\mathbf{x}}(t) = T\mathbf{x}(t)$, i.e., T is a one-to-one linear map which preserves dynamics and which does not depend on the switching $\boldsymbol{\mu}$.

Definition 2.10 (Full rank)**.** *We will say that the process* $\mathbf{y}$ *is* full rank*, if for all* $i = 1, \dots, n_\mu$*,* $Q_i = E[\mathbf{e}(t)(\mathbf{e}(t))^T\boldsymbol{\mu}_i^2(t)]$ *is invertible.*

This is a direct extension of the classical notion of a full rank process [65].

Theorem 2.2 (Existence and uniqueness)**.** *Assume that* $(\mathbf{y}, \boldsymbol{\mu})$ *has an asGLSS realization and that* $\mathbf{y}$ *is full rank and SII w.r.t.* $\boldsymbol{\mu}$*. It follows that:*

1. $(\mathbf{y}, \boldsymbol{\mu})$ *has a minimal asGLSS realization in innovation form.*

2. *Any two minimal asGLSS realizations of* $(\mathbf{y}, \boldsymbol{\mu})$ *in innovation form are isomorphic.*

The proof of this theorem follows [91, Theorem 2] by using the correspondence between **GBS** and asGLSS.

Theorem 2.2 implies that asGLSSs in innovation form have the useful property that, when they are minimal and realize the same output, then they are unique up to isomorphism. Moreover, any asGLSS can be transformed to a minimal asGLSS in innovation form with the same output. Hence, in system identification of asGLSSs, in the absence of additional information on the true system, there is no loss of generality in assuming that the true system is a minimal asGLSS in innovation form. That is, for system identification it is preferable to consider parametrizing the elements of minimal asGLSSs in innovation form. This motivates finding conditions for an asGLSS to be minimal in innovation form and formulating algorithms for transforming an asGLSS to a minimal

one in innovation form.
To this end, in Subsection 2.7.4 we present a rank condition for minimality and a minimization algorithm based on the results of [91]. However, the results of Subsection 2.7.4 do not allow checking that an asGLSS is in innovation form. Moreover, the algebraic conditions for minimality are difficult to apply. This motivates us to present the more user-friendly characterization of minimality and being in innovation form in Chapter 4. However, in this chapter, we keep presenting what already exists in the literature.

2.7.4 Minimization algorithm

In order to present the rank conditions for minimality of asGLSS and the minimization algorithm, we need to define the *dLSS associated with asGLSS* as

$$\mathscr{S}_S = (\{\sqrt{p_i}A_i, G_i\}_{i=1}^{n_\mu}, C, I_{n_y}), \tag{2.19}$$

where

$$G_i = \frac{1}{\sqrt{p_i}}(A_i P_i C^T + K_i Q_i F^T), \tag{2.20}$$

and $Q_i = E[\mathbf{v}(t)\mathbf{v}^T(t)\boldsymbol{\mu}_i^2(t)]$ for $i = 1, \ldots, n_\mu$ is the noise covariance matrix, and $P_i = E[\mathbf{x}(t)\mathbf{x}^T(t)\boldsymbol{\mu}_i^2(t)]$, $i = 1, \ldots, n_\mu$ are the state covariance matrices, and they can be computed as follows:

$$P_i = \lim_{\bar{\mathcal{I}} \to \infty} P_i^{\bar{\mathcal{I}}},$$

and $\{P_i^{\bar{\mathcal{I}}}\}_{i=1}^{n_\mu}$ satisfy the following recursions:

$$P_i^{\bar{\mathcal{I}}+1} = p_i \sum_{j=1}^{n_\mu} (A_j P_j^{\bar{\mathcal{I}}} A_j^T + K_j Q_j K_j^T), \tag{2.21}$$

with $P_i^0 = \mathbf{O}_{n_x,n_x}$ where $\mathbf{O}_{n_x,n_x}$ denotes the $n_x \times n_x$ matrix with all zero entries. Note that the existence of the limit of $P_i^{\bar{\mathcal{I}}}$ when $\bar{\mathcal{I}}$ goes to infinity follows from [91, Lemma 5]. Note that the matrices G_σ can also be related to covariances of states and outputs as follows.

Remark 2.5 (Interpretation of G_σ)**.** *From equation (37) in [91], by applying it to $w = \epsilon$, it follows that for all $\sigma \in \Sigma$,*

$$G_\sigma = E\left[\mathbf{x}(t)\mathbf{z}_\sigma^{\mathbf{y}}(t)\right].$$

From [91] it also follows that the dLSS associated with an asGLSS $\mathscr{S}$ represents a realization of Markov-function $\Psi_{\mathbf{y}} : \Sigma^* \to \mathbb{R}^{n_y \times n_y}$,

$$\Psi_{\mathbf{y}}(w) = \begin{cases} E[\mathbf{y}(t)(\mathbf{z}_w^{\mathbf{y}}(t))^T] = \Lambda_w^{\mathbf{y},\mathbf{y}}, & \text{if} \quad w \in \Sigma^+, \\ I_{n_y}, & \text{if} \quad w = \epsilon. \end{cases}$$

computed from the covariances of $\mathbf{y}$. Then from [91] we can derive the following.

Theorem 2.3. *The dLSS $\mathscr{S}_{\mathcal{S}}$ associated with $\mathcal{S}$ is a realization of the Markov-function $\Psi_{\mathbf{y}}$. An asGLSS $\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$ is a minimal realization of $(\mathbf{y}, \boldsymbol{\mu})$, if and only if the associated dLSS $\mathscr{S}_{\mathcal{S}} = (\{\sqrt{p_i} A_i, G_i\}_{i=1}^{n_\mu}, C, I_{n_y})$ is a minimal realization of $\Psi_{\mathbf{y}}$.*

The proof follows from the proof of Theorem 2 of [91] by noticing that the recognizable representation associated with a GBS in the terminology of [91] corresponds to the dLSS associated with an asGLSS in the terminology of this chapter. From Theorem 2.3 it follows that minimality of $\mathcal{S}$ can be checked by applying Algorithm 1 to the dLSS $\mathscr{S}_{\mathcal{S}}$. Note, as mentioned in motivation **M2** of Section 2.6, that in contrast to dLSSs, minimal asGLSSs, realizing the same output, may not be isomorphic. In fact, this will be illustrated in the numerical section of Chapter 4.

Vice versa, with any dLSS realization of $\Psi_{\mathbf{y}}$ we can associate an asGLSS in innovation form. This latter relationship is useful for formulating a minimization algorithm. More precisely, consider a dLSS

$$\mathscr{S} = (\{\hat{A}_i, \hat{G}_i\}_{i=1}^{n_\mu}, \hat{C}, I_{n_y}),$$

which is a minimal realization of $\Psi_{\mathbf{y}}$. Define the *asGLSS $\mathcal{S}_{\mathscr{S}}$ associated with $\mathscr{S}$* as

$$\mathcal{S}_{\mathscr{S}} = (\{\frac{1}{\sqrt{p_i}} \hat{A}_i, \hat{K}_i\}_{i=1}^{n_\mu}, \hat{C}, I_{n_y}, \mathbf{e}),$$

where $\hat{K}_i = \lim_{\bar{\mathcal{I}} \to \infty} \hat{K}_i^{\bar{\mathcal{I}}}$, and $\{\hat{K}_\sigma^{\bar{\mathcal{I}}}\}_{\sigma \in \Sigma, \bar{\mathcal{I}} \in \mathbb{N}}$ satisfies the following recursion

$$\begin{aligned}
\hat{P}_\sigma^{\bar{\mathcal{I}}+1} &= \sum_{\substack{\sigma_1 \in \Sigma \\ \sigma\sigma_1 \in L}} p_\sigma \left(\frac{1}{p_{\sigma_1}} \hat{A}_{\sigma_1} \hat{P}_{\sigma_1}^{\bar{\mathcal{I}}} (\hat{A}_{\sigma_1})^T + \hat{K}_{\sigma_1}^{\bar{\mathcal{I}}} \hat{Q}_{\sigma_1}^{\bar{\mathcal{I}}} (\hat{K}_{\sigma_1}^{\bar{\mathcal{I}}})^T \right) \\
\hat{Q}_\sigma^{\bar{\mathcal{I}}} &= p_\sigma T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}} - \hat{C} \hat{P}_\sigma^{\bar{\mathcal{I}}} (\hat{C})^T \\
\hat{K}_\sigma^{\bar{\mathcal{I}}} &= \left(\hat{G}_\sigma \sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}} \hat{A}_\sigma \hat{P}_\sigma^{\bar{\mathcal{I}}} (\hat{C})^T \right) \left(\hat{Q}_\sigma^{\bar{\mathcal{I}}} \right)^{-1}
\end{aligned} \tag{2.22}$$

where $\hat{P}_\sigma^0 = \mathbf{O}_{n_x, n_x}$ and $T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}} = E\left[\mathbf{z}_\sigma^{\mathbf{y}}(t) (\mathbf{z}_\sigma^{\mathbf{y}}(t))^T\right]$. We can state the following reformulation of [91, Theorem 3]

Lemma 2.1. *Assume that $\mathbf{y}$ is full rank and $(\mathbf{y}, \boldsymbol{\mu})$ has a realization by an asGLSS and $\mathscr{S}$ is a minimal dLSS realization of $\Psi_{\mathbf{y}}$. Then the recusion (2.22) is well posed, i.e., the matrices $\{\hat{Q}_\sigma^{\mathcal{I}}\}_{\sigma \in \Sigma, \mathcal{I} \in \mathbb{N}}$ are invertable, the limits $\hat{K}_\sigma = \lim_{\bar{\mathcal{I}} \to \infty} \hat{K}_\sigma^{\bar{\mathcal{I}}}$, $\hat{Q}_\sigma = \lim_{\bar{\mathcal{I}} \to \infty} \hat{Q}_\sigma^{\bar{\mathcal{I}}}$, $\hat{P}_\sigma = \lim_{\bar{\mathcal{I}} \to \infty} \hat{P}_\sigma^{\bar{\mathcal{I}}}$ exist, and the associated asGLSS $\mathcal{S}_{\mathscr{S}}$ is a minimal asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form. Furthermore, $T_{\sigma,\sigma}^{\mathbf{e},\mathbf{e}} p_\sigma = \hat{Q}_\sigma$, and if $\mathbf{x}$ is the unique ZMWSSI state process of $\mathcal{S}_{\mathscr{S}}$, then $\hat{P}_\sigma = p_\sigma T_{\sigma,\sigma}^{\mathbf{x},\mathbf{x}}$.*

The discussion above suggests the following realization algorithm (Algorithm 3) for transforming an asGLSS $\mathcal{S}$ to a minimal one, which can be deduced from [91].

Algorithm 3: Minimization algorithm,
Input : The asGLSS $\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$ representation matrices.
Output : The asGLSS $\mathcal{S}_{\mathscr{S}_m}$.

1: Compute the dLSS $\mathscr{S}_S$ associated with $\mathcal{S}$ and compute $T^{\mathbf{y}}_{\sigma,\sigma} = \frac{1}{p_\sigma}(CP_\sigma C^T + FQ_\sigma F^T)$ and $Q_\sigma = E[\mathbf{v}(t)\mathbf{v}^T(t)\boldsymbol{\mu}_\sigma^2(t)]$
2: Transform the dLSS $\mathscr{S}_S$ to a minimal dLSS $\mathscr{S}_m$ using Algorithm 2.
3: Construct the asGLSS $\mathcal{S}_{\mathscr{S}_m}$ associated with $\mathscr{S}_m$.

From [91, Theorem 3], it follows that Algorithm 3 returns a minimal realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form, if $\mathcal{S}$ is an asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ and $\mathbf{y}$ is SII. Note that Algorithm 3 requires only the knowledge of the matrices of $\mathcal{S}$ and the noise covariance matrices $Q_\sigma = E[\mathbf{v}(t)\mathbf{v}^T(t)\boldsymbol{\mu}_\sigma^2(t)]$. Also note that all the steps above are computationally efficient. However, they require finding the limits of $P_\sigma^{\bar{\mathcal{I}}}$ and $\hat{K}_\sigma^{\bar{\mathcal{I}}}$ for $\bar{\mathcal{I}} \to \infty$ respectively. Also note that, in [74], there exists another minimization algorithm which uses covariances matrices.

Remark 2.6 (Challenges)**.** *The main disadvantage of verifying the rank condition of Theorem 2.3 or applying Algorithm 3 is the necessity of constructing a dLSS and the necessity to find the limit of the matrices in* (2.22)*. The latter represents an extension of algebraic Riccati equations [91, Remark 7] and even for the linear case requires attention. Moreover, the rank conditions of Theorem 2.3 are not easy to apply to parametrizations: even if the dependence of the matrices $\{A_i, K_i\}_{i=1}^{n_\mu}$ and C on a parameter θ are linear or polynomial, the dependence of the matrices of the associated dLSS need not remain linear or polynomial, due to the definition of G_i in* (2.20)*. For the same reason, it is difficult to analyze the result of applying Algorithm 3 to elements of a parametrization. Moreover, the conditions of Theorem 2.3 do not allow us to check if the elements of a parametrizations are in innovation form. These shortcomings motivate the contribution in Section 4.2 of Chapter 4.*

2.8 Conclusion

This chapter presented the classes of models we are dealing with throughout this thesis, the sGLSS and the asGLSS. To this end, it presents a number of definitions concerning the properties of the stochastic processes of the designated stochastic system. This was necessary in order to avoid repetitions in the following chapters, where we will be introducing the contributions of the thesis. In addition, we presented an overview of the realization theory for the subclass asGLSS. In turn, asGLSS will serve as a basis for stochastic realization theory of GLSS with inputs.

2.9 Appendix: further details on existence, uniqueness and series representation of state processes of sGLSSs.

In this section we will discuss in more detail the uniqueness of the state process of sGLSSs and the convergence of the infinite sum (2.10). We start with the following lemma, which is a consequence of [91, Lemma 3] and which will be used in the subsequent chapters.

Lemma 2.2 (Convergence infinite sums)**.** *Consider a $p \times n$ matrix C and matrices $n \times n_y$ $\{B_{i=1}\}_{i=1}^{n_\mu}$ and $n \times n$ matrices $\{F_i\}_{i=1}^{n_\mu}$ such that that $\sum_{i=1}^{n_y} p_i(F_i \otimes F_i)$ is stable. Let $\mathbf{v}$ be a white noise w.r.t. $\boldsymbol{\mu}$. Then the following holds:*

(a) *The infinite sum*

$$\mathbf{r}(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w|=N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} C F_w B_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t) \tag{2.23}$$

is convergent in the mean-square sense.

(b) *The process $\mathbf{r}(t)$ is ZMWSSI and it is the unique state process of the asGLSS $(\{F_i, B_i\}_{i=1}^{n_\mu}, C, I, \mathbf{v})$.*

(c) *The infinite sum*

$$\sum_{N=0}^{\infty} \sum_{\substack{|w|=N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} E[\|C F_w B_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t)\|_2^2] \tag{2.24}$$

is convergent.

By replacing $\mathbf{v}$ by $\mathbf{w} = \begin{bmatrix} \mathbf{v}^T & \mathbf{u}^T \end{bmatrix}^T$ in Lemma 2.2, it follows that (2.10) holds and that its right-hand side is mean-square convergent.

Before presenting the proof of Lemma 2.2, we need the following result

Lemma 2.3 (Orthogonal noises)**.** *Let $\mathbf{v}$ be a white noise w.r.t $\boldsymbol{\mu}$. Then $\mathbf{z}^{\mathbf{v}}_w(t), \mathbf{z}^{\mathbf{v}}_v(t)$ are uncorrelated for all $w, v \in \Sigma^+$, $v \neq w$.*

Proof of Lemma 2.3. Clearly $E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T]$ is either zero if w is not a suffix of v, or if v is not a suffix of w or if either $v \notin L$ or $w \notin L$, as it is ZMWSSI. If say $v = sw$, $s \neq \epsilon$, $w, v \in L$ then

$$E[\mathbf{z}^{\mathbf{v}}_w(t)(\mathbf{z}^{\mathbf{v}}_v(t))^T] = E[\mathbf{v}(t)\mathbf{z}^{\mathbf{v}}_s(t)] = 0,$$

due to the definition of a white noise w.r.t. $\boldsymbol{\mu}$. □

Proof of Lemma 2.2. The proof of Lemma 2.2 is divided into two stages: the first deals with propositions **(a)** and **(b)** of the Lemma and the second deals with the remainder of its propositions - **(c)**.
*Proof of **(a)** and **(b)**:*
From [91, Lemma 3] it follows that (2.23) is convergent in mean square sense and $\mathbf{r}(t)$ is ZMWSSI as it is the solution of the asGLSS $(\{F_i, B_i\}_{i=1}^{n_\mu}, C, I, \mathbf{v})$.
*Proof of **(c)**:*
From Lemma 2.3 and the assumption that $\mathbf{v}$ is a white nose w.r.t. $\boldsymbol{\mu}$ it follows that $\mathbf{z}_w^{\mathbf{v}}(t), \mathbf{z}_v^{\mathbf{v}}(t)$ for any $v \neq w$ are uncorrelated. Hence

$$E[\|\sum_{N=0}^{\infty} \sum_{\substack{|w|=N \\ w\in\Sigma^*}} \sum_{\substack{\sigma\in\Sigma \\ \sigma w\in L}} CF_w B_\sigma \mathbf{z}_{\sigma w}^{\mathbf{v}}(t)\|_2^2]$$

$$= trace\left(\sum_{N=0}^{\infty} \sum_{\substack{|w|=N \\ w\in\Sigma^*}} \sum_{\substack{\sigma\in\Sigma \\ \sigma w\in L}} \sum_{\substack{|v|=N \\ v\in\Sigma^*}} \sum_{\substack{\sigma_1\in\Sigma \\ \sigma_1 v\in L}} CA_w B_\sigma E[\mathbf{z}_{\sigma w}^{\mathbf{v}}(t)(\mathbf{z}_{\sigma_1 v}^{\mathbf{v}}(t))^T](CA_v B_{\sigma_1})^T\right)$$

$$= \sum_{N=0}^{\infty} \sum_{\substack{|w|=N \\ w\in\Sigma^*}} \sum_{\substack{\sigma\in\Sigma \\ \sigma w\in L}} E[\|CF_w B_\sigma \mathbf{z}_{\sigma w}^{\mathbf{v}}(t)\|_2^2]$$

and as the left-hand side of the latter equation is convergent by mean-square convergence of (2.23), it follows that (2.24) is convergent. □

In fact, the sum in (2.23) remains mean-square convergent if we change the order of the summands. To formalize this observation, we will need to recall from [41, Definition 3.2] the notion of unconditional convergence of a series of elements of a Hilbert-space: a series $\sum_{n=0}^{\infty} a_n$ is unconditionally convergent if it is convergent and for any bijection $\pi : \mathbb{N} \to \mathbb{N}$, the rearranged series $\sum_{n=0}^{\infty} a_{\pi(n)}$ is also convergent, and

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} a_{\pi(n)}.$$

Informally, this means that the elements of an unconditionally convergent series can be summed in any order without changing the resulting infinite sum.

In particular, let a_w, $w \in \Sigma^*$ be a collection of square-integrable random variables from the Hilbet-space $\mathcal{H}_1$ defined in Notation 2.1. Notice that as the set Σ^* is denumerable, we can always define a bijection $\phi : \mathbb{N} \to \mathbb{N}$ and identify the series $\sum_{w\in\Sigma^*} a_w$ with the series $\sum_{n=0}^{\infty} a_{\phi(n)}$. We will say that $\sum_{w\in\Sigma^*} a_w$ is *unconditionally convergent in the mean square sense*, if the series

$\sum_{n=0}^{\infty} a_{\phi(n)}$ is unconditionally convergent in the Hilbert-space $\mathcal{H}_1$. In particular, this means that for any bijection $\tau : \Sigma^* \to \Sigma^*$,

$$\sum_{n=0}^{\infty} a_{\phi(n)} = \sum_{n=0}^{\infty} a_{\tau(\phi(n))}.$$

Indeed, in this case $\tau(\phi(n)) = \phi(\pi(n))$, for the bijection $\pi : \mathbb{N} \to \mathbb{N}$, $\pi(n) = \phi^{-1}(\tau(\phi(n))$. In particular, in this case

$$\sum_{N=0}^{\infty} \sum_{w \in \Sigma^*, |w|=N} a_w = \sum_{n=0} a_{\phi(n)}.$$

The latter limit we will denote by $\sum_{w \in \Sigma^*} a_w$.

It is easy to see that unconditional convergence of $\sum_{w \in \Sigma^*} a_w$ does not depend on the choice of the bijection ϕ: if $\psi : \mathbb{N} \to \Sigma^*$ is another bijection, then for any bijection ϕ on $\mathbb{N}$, $\tilde{\pi} = \phi^{-1} \circ \psi \circ \pi$ is a bijection on $\mathbb{N}$ and

$$\psi(\pi(n)) = \phi(\tilde{\pi}(n)),$$

$n \in \mathbb{N}$, and hence unconditional convergence of $\sum_{n=0}^{\infty} a_{\phi(n)}$ implies that of $\sum_{n=0}^{\infty} a_{\psi(n)}$. That is, if $\sum_{w \in \Sigma^*} a_w$ is unconditionally convergent, then the infinite sum $\sum_{w \in \Sigma^*} a_w$ can be summed in any order without changing the limits.

Lemma 2.4. *With the assumptions and notation of Lemma 2.2, the infinite sum* (2.23) *is unconditionally convergent in the mean-square sense.*

Since the right-hand side of (2.10) is a particular case of (2.23), and for all $v \in \Sigma^+$, $v \notin L$, $\mathbf{z}_v^{\mathbf{v}} = 0$, $\mathbf{z}_v^{\mathbf{u}} = 0$, with convention above for unconditionally convergent sums, we can rewrite (2.10) as

$$\mathbf{x}(t) = \sum_{w \in \Sigma^*, \sigma \in \Sigma} \sqrt{p_{\sigma w}} \left(A_w B_\sigma \mathbf{z}_{\sigma w}^{\mathbf{u}}(t) + A_w K_\sigma \mathbf{z}_{\sigma w}^{\mathbf{v}}(t) \right)$$

Proof of Lemma 2.2. Since $\mathbf{z}_w^{\mathbf{v}}(t), \mathbf{z}_v^{\mathbf{v}}(t)$ are uncorrelated, i.e., their components are orthogonal elements of the Hilbert-space $\mathcal{H}_1$ from Notation 2.1 the statement of the lemma follows from [41, Exercise 3.1] and part **(c)** of Lemma 2.2. More precisely, (2.24) implies that

$$\sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \left(C_i F_w (B_\sigma)_j \right)^2$$

is convergent where $i = 1, 2, \ldots, n_y$ and $j = 1, 2, \ldots, n_u$. Indeed, if $\sigma w \in L$, then

$$\begin{aligned} E[\|CF_wB_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t)\|_2^2] &= trace\Big((CF_wB_\sigma)T^{\mathbf{v},\mathbf{v}}_{\sigma w,\sigma w}(CF_wB_\sigma)^T\Big) \\ &= trace\Big((CF_wB_\sigma)T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}(CF_wB_\sigma)^T\Big) \\ &= \sum_{r=1}^{n_y}(C_rF_wB_\sigma)T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}(C_rF_wB_\sigma)^T \\ &\geq (C_iF_wB_\sigma)T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}(C_iF_wB_\sigma)^T \end{aligned}$$

as for all $r = 1, \ldots, n_y$

$$(C_rF_wB_\sigma)T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}(C_rF_wB_\sigma)^T \geq 0.$$

However,

$$(C_iF_wB_\sigma)T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}(C_iF_wB_\sigma)^T \geq \lambda_{min}(T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma})\|C_iF_wB_\sigma\|_2^2,$$

where $\lambda_{min}(T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma})$ is the smallest eigenvalue of $T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}$. As $T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma} > 0$ is strictly positive definite, $\lambda_{min}(T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}) > 0$. Hence, for any $\sigma w \in L$:

$$(C_iF_w(B_\sigma)_j)^2 \leq \|C_iF_wB_\sigma\|_2^2 \leq E\left[\frac{\|CF_wB_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t)\|_2^2}{\lambda_{min}(T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma})}\right].$$

Define, for all $i = 1, \ldots, n_y$, $j = 1, \ldots, n_u$:

$$c^{i,j}_w = \begin{cases} C_iF_w(B_\sigma)_j, & \sigma w \in L \\ 0, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad e^j_w = \frac{\mathbf{z}^{\mathbf{v}}_{\sigma w}(t)}{(T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma})_{j,j}}.$$

Then

$$\sum_{N=0}^{\infty}\sum_{w\in\Sigma^*,|w|=N}(c^{i,j}_w)^2 < +\infty,$$

and $(e^j_w)_{w\in\Sigma^*}$ is orthogonal in $\mathcal{H}_1$. From [41, Exercise 3.1] it follows that $\sum_{w\in\Sigma^*} c^{i,j}_w e^j_w$ is unconditionally converged. Notice that the ith entry of $CF_wB_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t)$ equals $\sum_{w\in\Sigma^*} c^{i,j}_w e^j_w$ and hence

$$\sum_{N=0}^{\infty}\sum_{\substack{|w|=N \\ w\in\Sigma^*}}\sum_{\substack{\sigma\in\Sigma \\ \sigma w\in L}} C_iF_wB_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t)$$

is a sum of unconditionally convergent series, hence it is unconditionally convergent. Finally, as all entries of

$$\sum_{N=0}^{\infty}\sum_{\substack{|w|=N \\ w\in\Sigma^*}}\sum_{\substack{\sigma\in\Sigma \\ \sigma w\in L}} \sqrt{p_{\sigma w}}CF_wB_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t)$$

are unconditionally convergent, so is the entire series.

□

Chapter 3

Realization theory for GLSS with inputs: decomposition into stochastic and deterministic components

3.1 Introduction

In this chapter, we present the first theoretical contribution for the GLSS class, which, we recall, includes subclasses of jump-Markov, piecewise-linear and LPV systems. We prove that the output of such systems can be decomposed into deterministic and stochastic components. Using this decomposition, we show existence of state-space representation in innovation form, and we provide sufficient conditions for such representations to be minimal and unique up to isomorphism. These theoretical results represent a step towards stochastic realization theory, which, in turn, is important for the system identification algorithm to be presented in Chapter 5.

Contribution: In this chapter, we study GLSSs of the form (2.1) with white noise inputs, uncorrelated with the switching and with the noise. We show that the output of a GLSS can be written as $\mathbf{y}(t) = \mathbf{y}^d(t) + \mathbf{y}^s(t)$, where $\mathbf{y}^d$ is the output of a GLSS with no noise $\mathbf{v}$, and $\mathbf{y}^s$ is the output of a GLSS with no input $\mathbf{u}$. Furthermore, by using results on realization theory of GLSSs with no inputs [91], we use this decomposition to show existence of an innovation form for GLSSs with inputs. Moreover, we present sufficient conditions for minimality and uniqueness (up to isomorphism) of GLSSs in innovation form. Intuitively, an isomorphism between two GLSSs is a linear change of basis, independent of the discrete modes, which transforms the matrices of one GLSS to the corresponding matrices of the other GLSS. This means that minimal GLSSs in innovation form have the same useful properties for system identification as their LTI counterparts. In particular, if two minimal GLSSs in innovation form generate (approximately) the same output for the data used for identification, then they will generate (approximately) the same output for any input and switching signal, including those which do not satisfy the assumptions of the identification experiment.

Prior work related to this chapter: Stochastic realization theory of GLSSs with no inputs, i.e., jump-Markov systems with no inputs, bilinear systems and autonomous stochastic LPV systems were addressed in [91]. With respect to [91] the main difference is the presence of the control input $\mathbf{u}$.

Existence of a decomposition and existence of innovation form appeared in [72], but only for the case of LPV systems with zero mean i.i.d. scheduling,. With respect to [72], the main novelty is that we allow more general switching processes, including finite Markov chains, and that we address minimality and uniqueness of innovation representations.

The existence of innovation representation was studied for LPV systems in [19, 21]. In contrast to this chapter, in [19, 21] the noise gain of the innovation representation has a dynamical dependence on the scheduling signal, and there is no claim on minimality and uniqueness of such representations. In particular, it is unclear when the innovation representation from [19, 21] generates the same output as the original model for all scheduling signals. However, [19, 21] has the advantage that it is valid for any scheduling signal.

Outline of the chapter: In Section 3.2, we present the assumptions used in this chapter. Those assumptions concern the process involved in the GLSSs systems. After this, the main contributions of this chapter are presented in Section 3.3. Finally, the chapter is concluded in Section 3.4.

3.2 Preliminaries and assumptions

Below, in this section, we present the assumptions used in the chapter. To this end, the reader should recall the definitions of Chapter 2 concerning the properties of the processes (ZMWSSI, Whte noise, Admissible process, etc.). Also, we note that in this chapter we will handle the GLSS class of systems **with** inputs, i.e., the class of (2.1) defined in Section 2.2

This said, the first assumption applied in this chapter states that:

Assumption 3.1 (Stationarity)**.** *The GLSSs are considered to be stationary in the sense of Definition 2.6.*

Intuitively, sGLSSs are introduced in order to ensure that all the relevant stochastic processes are stationary in a suitable sense. The latter assumption is widespread in stochastic realization theory and system identification.

We recall from the previous chapter that, in the terminology of [91], a sGLSS (resp. asGLSS) corresponds to a stationary *Generalized Bilinear System* (**GBS**) with noise $\begin{bmatrix}\mathbf{v}^T & \mathbf{u}^T\end{bmatrix}^T$ (resp. $\mathbf{v}$), meaning that the results of [91] are applicable for sGLSSs. Then, from [91], it follows that the state and output process $\mathbf{x}$ and $\mathbf{y}$ are ZMWSSI, and hence Assumption 3.3 is satisfied for all $(\mathbf{u}, \boldsymbol{\mu}, \mathbf{y})$ generated by sGLSS.

As a result, Assumption 3.1 ***implicitly includes*** the following two assumptions, which we now state for clarity.

Assumption 3.2 (switching)**.** *The switching process $\boldsymbol{\mu}$ is admissible.*

This assumption imposes restrictions on the data used for system identification, but not necessarily for the model class which will be identified depending on the switching $\boldsymbol{\mu}$ as discussed in the preliminary chapter. It can be thought of as a persistence of excitation condition. In particular, binary and white noises, which are the simplest persistently exciting signals, satisfy our assumption. We believe that for developing realization theory for general switching signals, these simple cases must be understood first. Moreover, admissible switching signals cover the fairly general case of Markov chains.

Remark 3.1 (LPV systems)**.** *For LPV systems, our assumption implies that the scheduling signal used for identification is sampled from a stochastic process. In addition to this being a persistence of excitation condition, we argue that in certain cases this assumption is justified [95]: in the presence of measurement errors, or when the scheduling is externally generated, or is a function of the stochastic states/inputs.*

Assumption 3.3 (Inputs and outputs)**.** ***(1)*** $\mathbf{u}$ *is a white noise w.r.t.* $\boldsymbol{\mu}$, ***(2)*** *and the process* $\begin{bmatrix}\mathbf{y}^T, & \mathbf{u}^T\end{bmatrix}^T$ *is a ZMWSSI.*

The assumption that $\mathbf{u}$ is white noise was made for the sake of simplicity. We conjecture that the results of the chapter can be extended to more general inputs, e.g., inputs generated by autoregressive models driven by white noise.

Remark 3.2. *In stochastic realization theory of LTI systems [94], it is assumed that* $\mathbf{y}$ *does not Granger cause* $\mathbf{u}$*. If* $\mathbf{u}$ *is a white noise, this condition is satisfied, i.e. we are consistent with [94]. Note that part* ***(1)*** *of Definition 2.3 is similar to Granger causality, but it captures absence of feedback from* $\mathbf{y}$ *to* $\boldsymbol{\mu}$ *and it is different from Granger-causality used in [94].*

Remark 3.3. *Note that* $\mathbf{u}$ *and* $\mathbf{v}$ *may have different variances, as long as they satisfy condition* ***1.*** *of Definition 2.6.*

3.3 Main result: decomposition into deterministic and stochastic component

Using Notation 2.1, let us define the *deterministic component* $\mathbf{y}^d$ of $\mathbf{y}$ as follows

$$\mathbf{y}^d(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}]. \tag{3.1}$$

Also, define the *stochastic component* of $\mathbf{y}$ as

$$\mathbf{y}^s(t) = \mathbf{y}(t) - \mathbf{y}^d(t). \tag{3.2}$$

It then follows that $\mathbf{y}(t) = \mathbf{y}^s(t) + \mathbf{y}^d(t)$. Intuitively, $\mathbf{y}^d(t)$ represent the best prediction of $\mathbf{y}(t)$ which is linear in the present and past values $\{\mathbf{u}(s)\}_{s<t}$ of $\mathbf{u}$ and nonlinear in the past values $\{\boldsymbol{\mu}(s)\}_{s<t}$ of $\boldsymbol{\mu}$. In fact, $\mathbf{y}^d$ is the output of the asGLSS obtained from (2.1) by considering $\mathbf{v} = 0$ and viewing $\mathbf{u}$ as noise. That is, $\mathbf{y}^d$ is the output of the system with no noise, i.e. $\mathbf{y}^d(t)$ is a deterministic function of $\{\mathbf{u}(s), \boldsymbol{\mu}(s)\}_{s<t} \cup \{\mathbf{u}(t)\}$. This motivates calling $\mathbf{y}^d$ the deterministic component, despite it being in its mathematical nature a stochastic process. This terminology is consistent with the LTI literature [56, Definition 9.3]. In contrast, $\mathbf{y}^s$ is the output of the asGLSS obtained from (2.1) by taking $\mathbf{u} = 0$ and viewing $\mathbf{v}$ as noise. In particular, $\mathbf{y}^s$ depends only on the noise $\mathbf{v}$, and $\mathbf{y}^d$ does not depend on the noise but it depends only on the input $\mathbf{u}$. This can be seen from the following equations of the theorem.

Theorem 3.1. *For a sGLSS of the form* (2.1), $\mathcal{S}_d = (\{A_\sigma, B_\sigma\}_{\sigma=1}^{n_\mu}, C, D, \mathbf{u})$ *is an asGLSS of* $(\mathbf{y}^d, \boldsymbol{\mu})$ *and* $\mathcal{S}_s = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$ *is an asGLSS of* $(\mathbf{y}^s, \boldsymbol{\mu})$*. Define*

$$\mathbf{x}^d(t) = E_l[\mathbf{x}(t) \mid \{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}], \tag{3.3}$$

$$\mathbf{x}^s(t) = \mathbf{x}(t) - \mathbf{x}^d(t), \tag{3.4}$$

then,

$$\mathcal{S}_d \begin{cases} \mathbf{x}^d(t+1) = \sum_{i=1}^{n_\mu} (A_i \mathbf{x}^d(t) + B_i \mathbf{u}(t)) \boldsymbol{\mu}_i(t), \\ \mathbf{y}^d(t) = C\mathbf{x}^d(t) + D\mathbf{u}(t), \end{cases} \tag{3.5}$$

$$\mathcal{S}_s \begin{cases} \mathbf{x}^s(t+1) = \sum_{i=1}^{n_\mu} (A_i \mathbf{x}^s(t) + K_i \mathbf{v}(t)) \boldsymbol{\mu}_i(t), \\ \mathbf{y}^s(t) = C\mathbf{x}^s(t) + D\mathbf{v}(t), \end{cases} \tag{3.6}$$

and,

$$\mathbf{y}^d(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma\in\Sigma,\ w\in\Sigma^* \\ \sigma w \in L,\ |w|=N}} \left(\sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}_{\sigma w}^{\mathbf{u}}(t) \right) + D\mathbf{u}(t) \tag{3.7}$$

$$\mathbf{y}^s(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma\in\Sigma,\ w\in\Sigma^* \\ \sigma w \in L,\ |w|=N}} \left(\sqrt{p_{\sigma w}} A_w K_\sigma \mathbf{z}_{\sigma w}^{\mathbf{v}}(t) \right) + F\mathbf{v}(t) \tag{3.8}$$

The proof of Theorem 3.1 is presented in the Appendix 3.5.1. In fact, the converse of Theorem 3.1 also holds. To this end, recall from Chapter 2, Section 2.7.2, (2.16) (or, alternatively, from [91]) the definition of the *innovation process* of $\mathbf{y}^s$:

$$\mathbf{e}^s(t) = \mathbf{y}^s(t) - E_l[\mathbf{y}^s(t) \mid \{\mathbf{z}_w^{\mathbf{y}^s}(t)\}_{w\in\Sigma^+}] \tag{3.9}$$

Intuitively, $\mathbf{e}^s(t)$ is the difference between $\mathbf{y}(t)$ and the best linear prediction of $\mathbf{y}^s(t)$ based on its own past multiplied with past values of the switching process.

Theorem 3.2. *Assume that there exists a sGLSS of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *of the form* (2.1)*, and that the following holds:*

1. $\hat{\mathcal{S}}_d = (\{\hat{A}_i, \hat{B}_i\}_{i=1}^{n_\mu}, \hat{C}, \hat{D}, \mathbf{u})$ *is an asGLSS of* $(\mathbf{y}^d, \boldsymbol{\mu})$*.*
2. $\hat{\mathcal{S}}_s = (\{\hat{A}_i, \hat{K}_i\}_{i=1}^{n_\mu}, \hat{C}, I_{n_y}, \mathbf{v})$ *is an asGLSS of* $(\mathbf{y}^s, \boldsymbol{\mu})$ *in* innovation form*, i.e.* $\mathbf{v} = \mathbf{e}^s$*.*

Then $\hat{\mathcal{S}} = (\{\hat{A}_i, \hat{K}_i, \hat{B}_i\}_{i=1}^{n_\mu}, \hat{C}, \hat{D}, I, \mathbf{e}^s)$ *is a sGLSS of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$*, and the innovation process* $\mathbf{e}^s$ *satisfies* $\mathbf{e}^s(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t)$*, where*

$$\hat{\mathbf{y}}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t), \mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^+} \cup \{\mathbf{u}(t)\}]. \tag{3.10}$$

The proof of Theorem 3.2 is presented in Section 3.5.2.

Remark 3.4. *The condition that the matrices* $\{\hat{A}_i\}_{i=1}^{n_\mu}$ *and* $\hat{C}$ *of* $\hat{\mathcal{S}}_d$ *and* $\hat{\mathcal{S}}_s$ *are the same can be relaxed: if* $\bar{\mathcal{S}}_d = (\{\hat{A}_i^d, \hat{B}_i^d\}_{i=1}^{n_\mu}, \hat{C}^d, \hat{D}, \mathbf{u})$ *and* $\bar{\mathcal{S}}_s = (\{\hat{A}_i^s, \hat{B}_i^s\}_{i=1}^{n_\mu}, \hat{C}^s, I, \mathbf{e}^s)$ *are asGLSS of* $(\mathbf{y}^d, \boldsymbol{\mu})$ *and* $(\mathbf{y}^s, \boldsymbol{\mu})$ *respectively, then with*

$$\hat{A}_i = \begin{bmatrix} \hat{A}_i^d & 0 \\ 0 & \hat{A}_i^s \end{bmatrix}, \ \hat{B}_i = \begin{bmatrix} \hat{B}_i^d \\ 0 \end{bmatrix}, \ \hat{K}_i = \begin{bmatrix} 0 \\ \hat{K}_i^s \end{bmatrix}, \ \hat{C} = \begin{bmatrix} C^d & C^s \end{bmatrix},$$

$\hat{\mathcal{S}}_d$ *and* $\hat{\mathcal{S}}_s$ *from Theorem 3.2 are asGLSSs of* $(\mathbf{y}^d, \boldsymbol{\mu})$ *and* $(\mathbf{y}^s, \boldsymbol{\mu})$ *respectively and Theorem 3.2 applies.*

Thus, Theorem 3.1 – 3.2 means that finding sGLSSs of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ is equivalent to finding an asGLSS representations of the deterministic and stochastic components respectively.

Theorem 3.2 suggests that $\mathbf{e}^s(t)$ can be viewed as the innovation process of $\mathbf{y}$. Indeed, $\hat{\mathbf{y}}(t)$ from (3.10) is the best linear prediction of $\mathbf{y}(t)$ based on past values of $\mathbf{y}$ and past and current values of $\mathbf{u}$ multiplied by past values of the switching process. Then $\mathbf{e}^s(t)$ is the prediction error $\mathbf{y}(t) - \hat{\mathbf{y}}(t)$. This motivates the following definition.

Definition 3.1 (Innovation form)**.** *A sGLSS* $\mathcal{S}$ *is a* sGLSS in innovation form*, if* F *is the identity matrix and* $\mathbf{v} = \mathbf{e}^s$*.*

That is, $\mathcal{S}$ from 2.1 is in innovation form, then:

$$\mathcal{S} \begin{cases} \mathbf{x}(t+1) = \displaystyle\sum_{i=1}^{n_\mu} (A_i \mathbf{x}(t) + B_i \mathbf{u}(t) + K_i \mathbf{e}^s(t)) \boldsymbol{\mu}_i(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + I_{n_y} \mathbf{e}^s(t), \ \ t \in \mathbb{Z} \end{cases} \tag{3.11}$$

Similarly to the LTI case [65], a sGLSS in innovation form can be viewed as a recursive filter driven by $\mathbf{y}$, $\mathbf{u}$, $\boldsymbol{\mu}$, whose output is the optimal prediction $\hat{\mathbf{y}}(t)$ from (3.10). Indeed, from

$$\mathbf{e}^s(t) = \mathbf{y}(t) - C\mathbf{x}(t) - D\mathbf{u}(t)$$

it follows that $\mathbf{x}(t+1)$ is a function of $\mathbf{x}(t), \mathbf{u}(t), \mathbf{y}(t)$ and $\boldsymbol{\mu}(t)$, and

$$\hat{\mathbf{y}}(t) = C\mathbf{x}(t) + D\mathbf{u}(t).$$

In other words, the filter can be expressed as follows:

$$\begin{aligned}\mathbf{x}(t+1) &= \sum_{i=1}^{n_\mu}\Big((A_i - K_iC)\mathbf{x}(t) + (B_i - K_iD)\mathbf{u}(t) + K_i\mathbf{y}(t)\Big)\boldsymbol{\mu}_i(t)\\ \hat{\mathbf{y}}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t), \;\; \text{for } \; t\in\mathbb{Z}.\end{aligned} \tag{3.12}$$

In what follows we will make the mild assumption that $\mathbf{y}^s$ is SII. This assumption is always satisfied when $\boldsymbol{\mu}$ is bounded. This follows from [91, Remark 2] and from the fact that $\mathbf{y}^s$ is the output of an asGLSS, and thus, by [91], it is ZMWSSI.

Corollary 3.1 (Existence)**.** *Assume that $\mathbf{y}^s$ is SII. Then any sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ can be transformed to a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form.*

Proof. From Theorem 3.1 it follows that $\mathcal{S}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ and $\mathcal{S}_d$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$. From [91, Theorem 2] it follows that there exists a (minimal) asGLSS $\hat{\mathcal{S}}_s$ of $(\mathbf{y}^s, \boldsymbol{\mu})$ in innovation form and by [91, Theorem 3] it can be computed from $\mathcal{S}_s$ using [91, Algorithm 1] or the minimization algorithm presented in Chapter 2 (Algorithm 3). Then using Remark 3.4 and Theorem 3.2, it follows that $\hat{\mathcal{S}}$ defined in Theorem 3.2 is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form, where, in this case, $\hat{\mathcal{S}}_d$ from Theorem 3.2is equal to $\mathcal{S}_d$. □

Next, we formulate conditions for minimality of sGLSSs. To this end, let us consider a sGLSS $\mathcal{S}$ of the form of (2.1). Recall from Chapter 2 the dimension of $\mathcal{S}$, denoted by $\dim(\mathcal{S})$, as the dimension n_x of its state-space.

Corollary 3.2 (Minimality)**.** *Assume that $\mathcal{S}$ is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, $\mathbf{y}^s$ is SII, and assume that $\mathcal{S}_s$ from Theorem 3.1 is a minimal asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$. Then $\mathcal{S}$ is a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.*

Proof. Let $\hat{\mathcal{S}}$ be a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ such that $\dim(\hat{\mathcal{S}}) < \dim(\mathcal{S})$. Then by Theorem 3.1, $\hat{\mathcal{S}}_s$ is an asGLSS of $(\mathbf{y}, \boldsymbol{\mu})$ of the same dimension as $\hat{\mathcal{S}}$. However, from [91, Theorem 2], $\mathcal{S}_s$ is a minimal dimensional asGLSS of $(\mathbf{y}, \boldsymbol{\mu})$ and hence $\dim(\mathcal{S}_s) \leq \dim(\hat{\mathcal{S}}_s)$. However, $\dim(\mathcal{S}_s) = \dim(\mathcal{S})$ by construction and hence $\dim(\mathcal{S}) \leq \dim(\hat{\mathcal{S}})$, which is a contradiction. □

In order to verify that the asGLSS $\mathcal{S}_s = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$ is minimal, by Theorem 2.3, it is sufficient to verify that the dLSS $\mathscr{S}_{\mathcal{S}_s}$ of the form (2.19) associated with $\mathcal{S}_s$ is minimal, i.e., it is reachable and observable. This can be verified by the rank condition as follow:

1. Compute the the i-step extended reachability matrices $\mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},i}$ and the i-step observability matrices $\mathcal{O}_{\mathscr{S}_{\mathcal{S}_s},i}$ of $\mathscr{S}_{\mathcal{S}_s}$:

$$\begin{aligned} \mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},i} &= \begin{bmatrix} \mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},0} & \sqrt{p_1}A_1\mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},i-1} & \dots & \sqrt{p_{n_\mu}}A_{n_\mu}\mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},i-1} \end{bmatrix}, \\ \mathcal{O}_{\mathscr{S}_{\mathcal{S}_s},i} &= \begin{bmatrix} \mathcal{O}^T_{\mathscr{S}_{\mathcal{S}_s},0} & \sqrt{p_1}A_1^T\mathcal{O}^T_{\mathscr{S}_{\mathcal{S}_s},i-1} & \dots & \sqrt{p_{n_\mu}}A_{n_\mu}^T\mathcal{O}^T_{\mathscr{S}_{\mathcal{S}_s},i-1} \end{bmatrix}^T, \end{aligned} \tag{3.13}$$

where $\mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},0} = \begin{bmatrix} G_1 & G_2 & \dots & G_{n_\mu} \end{bmatrix}$, $\mathcal{O}_{\mathscr{S}_{\mathcal{S}_s},0} = C^T$, and for all $i > 1$:

2. Verify if $\operatorname{rank}\{\mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},n_x-1}\} = n_x$, and $\operatorname{rank}\{\mathcal{O}_{\mathscr{S}_{\mathcal{S}_s},n_x-1}\} = n_x$.

We also get the following sufficient condition for isomorphism of minimal sGLSSs in innovation form.

To this end, below we define the concept of isomoprhism between sGLSS as follows. Let $\mathcal{S}$ be an asGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form of the form (2.6) with $F = I_{n_y}$. and let $\hat{\mathcal{S}} = (\{\hat{A}_\sigma, \hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, I_{n_y}, \mathbf{e}^s)$ be another asGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form. We say that $\mathcal{S}$ and $\hat{\mathcal{S}}$ are *isomorphic*, if there exists a non-singular matrix T (referred to as *isomorphism from $\mathcal{S}$ to $\hat{\mathcal{S}}$*) such that

$$\left(\forall \sigma \in \Sigma : \hat{A}_\sigma = TA_\sigma T^{-1},\ \hat{K}_\sigma = TK_\sigma,\ \hat{B}_\sigma = TB_\sigma\right) \text{ and } \ \hat{C} = CT^{-1},\ \ \hat{D} = D. \tag{3.14}$$

Note that if T is an isomorphism from $\mathcal{S}$ and $\hat{\mathcal{S}}$ and $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the state processes of $\mathcal{S}$ and $\hat{\mathcal{S}}$, then $\hat{\mathbf{x}}(t) = T\mathbf{x}(t)$, i.e., T is a one-to-one linear map which preserves dynamics and which does not depend on the switching $\boldsymbol{\mu}$.

Corollary 3.3 (Isomorphism)**.** *Assume that $\mathcal{S}$ is of the form* (2.1) *and that $\hat{\mathcal{S}} = (\{\hat{A}_\sigma, \hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma\in\Sigma}, \hat{C}, \hat{D}, I, \mathbf{e}^s)$ and they are both sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form and $\mathcal{S}_s$ and $\hat{\mathcal{S}}_s$ are both minimal asGLSS realizations of $(\mathbf{y}^s, \boldsymbol{\mu})$. Assume that $\mathbf{y}^s$ is SII, and that the covariance matrix $E[\mathbf{e}^s(t)(\mathbf{e}^s(t))^T\boldsymbol{\mu}_\sigma^2(t)]$ is nonsingular and $\operatorname{Im}[B_\sigma^T, \hat{B}_\sigma^T]^T \subseteq \operatorname{Im}[K_\sigma^T, \hat{K}_\sigma^T]^T$, $\sigma \in \Sigma$ and $\hat{D} = D$. Then there exists an isomorphism from $\mathcal{S}$ to $\hat{\mathcal{S}}$.*

Proof. Since $\mathcal{S}_s$ and $\hat{\mathcal{S}}_s$ are both minimal asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ in innovation form, and by Theorem 2.2 they are isomorphic, i.e., there exists a non-singular matrix T such that

$$TA_\sigma T^{-1} = \hat{A}_\sigma, TK_\sigma = \hat{K}_\sigma, CT^{-1} = \hat{C}$$

holds. Since $\operatorname{Im}[B_\sigma^T, \hat{B}_\sigma^T]^T \subseteq \operatorname{Im}[K_\sigma^T, \hat{K}_\sigma^T]^T$, for some matrix Z_σ, $B_\sigma = K_\sigma Z_\sigma$, $\hat{B}_\sigma = \hat{K}_\sigma Z_\sigma$, from which (3.14) follows. □

We illustrate Corollary 3.3 with the following example.

Example 3.1. *Let $\Sigma = \{1, 2\}$, and define the sGLSS $\mathcal{S}$ of the form* (2.1) *with state dimension $n = 2$, $\boldsymbol{\mu}$ being an admissible switching process with $p_1 = p_2 = 1$ and $L = \Sigma^+$ such that $\boldsymbol{\mu}(t)$*

is essentially bounded (i.e. there exists $K > 0$ such that for all $t \in \mathbb{Z}$, with probability one $\|\boldsymbol{\mu}(t)\|_2 < K$) and with the following matrices:

$$
\begin{aligned}
A_1 &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.3 \end{bmatrix}, & A_2 &= \begin{bmatrix} 0.4 & 0 \\ 0 & 0.2 \end{bmatrix}, \\
B_1 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} = K_1, & B_2 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} = K_2, \\
C &= \begin{bmatrix} 1 & 1 \end{bmatrix}, & D &= 0,\ F = 1,
\end{aligned}
$$

and the noise $\mathbf{v}$ is such that $Q_\sigma = E[\mathbf{v}(t)\mathbf{v}^T(t)\boldsymbol{\mu}_\sigma^2(t)] = 1$ for $\sigma = 1, 2$. The dLSS $\mathscr{S}_{\mathcal{S}_s} = (\{A_\sigma, G_\sigma\}_{\sigma=1}^2, C, 1)$ associated with $\mathcal{S}_s$ is reachable and observable, and hence $\mathcal{S}_s$ is a minimal asGLSS. Indeed, from (2.20) *and* (2.21) *it follows that*

$$
P_1 = P_2 = \begin{bmatrix} \frac{1}{1-(0.5^2+0.4^2)} & 0 \\ 0 & \frac{1}{1-(0.3^2+0.2^2)} \end{bmatrix},
$$

$$
G_1 = \begin{bmatrix} \frac{0.5}{1-(0.5^2+0.4^2)} + 1 \\ \frac{0.3}{1-(0.3^2+0.2^2)} \end{bmatrix}, \quad \textit{and} \quad G_2 = \begin{bmatrix} \frac{0.4}{1-(0.5^2+0.4^2)} \\ \frac{0.2}{1-(0.3^2+0.2^2)} + 1 \end{bmatrix},
$$

and hence the columns of the extended reachability matrix $\mathcal{R}_{\mathscr{S}_{\mathcal{S}_s},0} = \begin{bmatrix} G_1 & G_2 \end{bmatrix}$ span the state space. Likewise, the extended observability matrix $\mathcal{O}_{\mathscr{S}_{\mathcal{S}_s},1} = \begin{bmatrix} C \\ CA_1 \\ CA_2 \end{bmatrix}$ has full column rank, and the output $\mathbf{y}$ depends on both state variables via C. The output process $\mathbf{y}^s$ is SII by Remark 2.3. Therefore, this sGLSS satisfies the conditions of Corollary 3.2 and is minimal.

Define another sGLSS $\hat{\mathcal{S}} = (\{\hat{A}_\sigma, \hat{K}_\sigma, \hat{B}_\sigma\}_{\sigma=1}^2, \hat{C}, \hat{D}, 1, \mathbf{v})$ with the same alphabet, switching process $\boldsymbol{\mu}$, noise process $\mathbf{v}$, and state dimension:

$$
\begin{aligned}
\hat{A}_1 &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.3 \end{bmatrix}, & \hat{A}_2 &= \begin{bmatrix} 0.4 & 0 \\ 0 & 0.2 \end{bmatrix}, \\
\hat{B}_1 &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} = K_1, & \hat{B}_2 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} = K_2, \\
\hat{C} &= \begin{bmatrix} 1 & 0 \end{bmatrix}, & \hat{D} &= 0.
\end{aligned}
$$

The dLSS $\mathscr{S}_{\hat{\mathcal{S}}_s} = (\{\hat{A}_\sigma, \hat{G}_\sigma\}_{\sigma=1}^2, \hat{C}, 1)$ associated with $\hat{\mathcal{S}}_s$ reachable but not observable*, as the second state variable is not seen in the output. Hence, $\hat{\mathcal{S}}_s$ is not minimal, and thus the system $\hat{\mathcal{S}}$ does* not *satisfy Corollary 3.2 due to lack of observability. More precisely, if $\hat{\mathbf{x}}^s$ is the unique state* *process of the stochastic component $\hat{\mathcal{S}}_s$ of $\hat{\mathcal{S}}$, then from* (2.20) *it follows that the state covariances $\hat{P}_i = E[\hat{\mathbf{x}}^s(t)(\hat{\mathbf{x}}^s)^T(t)\boldsymbol{\mu}_i^2(t)]$, $i = 1, 2$ satisfy*

$$\hat{P}_1 = \hat{P}_2 = \begin{bmatrix} \frac{1}{1-(0.5^2+0.4^2)} & 0 \\ 0 & \frac{1}{1-(0.3^2+0.2^2)} \end{bmatrix}$$

$$\hat{G}_1 = \begin{bmatrix} \frac{0.5}{1-(0.5^2+0.4^2)} + 1 \\ 0 \end{bmatrix}, \quad \textit{and} \quad \hat{G}_2 = \begin{bmatrix} \frac{0.4}{1-(0.5^2+0.4^2)} \\ 1 \end{bmatrix}.$$

Then the extended reachability matrix $\mathcal{R}_{\mathscr{S}_{\hat{\mathcal{S}}_s},0} = \begin{bmatrix} \hat{G}_1 & \hat{G}_2 \end{bmatrix}$ is full row rank, i.e., $\mathscr{S}_{\hat{\mathcal{S}}_s}$ is indeed reachable. However, it is easy to see that the second column of the 1*-step extended observability matrix $\mathcal{O}_{\mathscr{S}_{\hat{\mathcal{S}}_s},1} = \begin{bmatrix} \hat{C} \\ \hat{C}\hat{A}_1 \\ \hat{C}\hat{A}_2 \end{bmatrix}$ is always zero, i.e., $\mathscr{S}_{\hat{\mathcal{S}}_s}$ is not observable.*

Note that, in later chapters, a more general (necessary and sufficient) condition for minimality of GLSS will be presented. Furthermore, uniqueness of minimal GLSS will be shown for a more general case.

3.4 Conclusion

We have shown that outputs of stochastic generalized linear switched systems can be decomposed into two parts, deterministic and stochastic one, and we used it to derive existence of representation in innovation form and to formulate sufficient conditions for minimality and uniqueness of such representations up to isomorphism. Future work will be directed towards extending these results for a larger class of inputs and switching signals. In the following chapters, we use the results of this chapter to formulate a realization algorithm for GLSS systems with inputs. The latter algorithm could not be unfolded without the technical decomposition presented in this chapter. This also led to develop a subspace identification algorithm based on the realization theory.

3.5 Appendix: proofs of Theorems 3.1 and 3.2

In what follows, for a ZMWSSI process $\mathbf{r}$, we denote by $\mathcal{H}_{t,+}^{\mathbf{r}}$ the closed subspace of $\mathcal{H}_1$ (see Notation 2.1) generated by the entries of $\{\mathbf{z}_w^{\mathbf{r}}(t)\}_{w\in\Sigma^+} \cup \{\mathbf{r}(t)\}$ and by $\mathcal{H}_t^{\mathbf{r}}$ the closed subspace of $\mathcal{H}_1$ generated by the entries of $\{\mathbf{z}_w^{\mathbf{r}}(t)\}_{w\in\Sigma^+}$.

In addition, we will say that $\mathbf{r}$ is orthogonal to $\mathbf{v}$, if $E\left[\mathbf{r}(t)\mathbf{v}^T(s)\right] = 0$ for any $t, s \in \mathbb{Z}$, where $\mathbf{r}$ and $\mathbf{v}$ are square integrable random variables. Moreover, if $\mathcal{M}$ is a closed subspace $\mathcal{H}_1$, then we say that $\mathbf{r}$ is orthogonal to $\mathcal{M}$, if $\mathbf{r}$ is orthogonal to every element of $\mathcal{M}$.

3.5.1 Proof of Theorem 3.1

In order to prove Theorem 3.1, we need the following lemmas, namely, Lemma 3.1, Lemma 3.2, Lemma 3.3.

Lemma 3.1. *The entries of* $\mathbf{v}(t)$, $\{\mathbf{z}_w^{\mathbf{v}}(t)\}_{w\in\Sigma^+}$ *are orthogonal to* $\mathcal{H}_{t,+}^{\mathbf{u}}$,

Proof. By definition, $\mathbf{w}(t) = [\mathbf{v}^T(t), \mathbf{u}^T(t)]^T$ is a white noise process w.r.t. $\boldsymbol{\mu}$, and $\mathbf{v}(t)$ is the upper n_n block of $\mathbf{w}(t)$. Since $\mathbf{w}$ is a white noise w.r.t. $\boldsymbol{\mu}$, it holds that

$$E[\mathbf{w}(t)(\mathbf{z}_w^{\mathbf{w}}(t))^T] = 0,$$

and $E[\mathbf{v}(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T]$ is the upper-right block of that latter matrix, and hence it is also zero. From the definition of sGLSS (Definition 2.6), it follows that

$$E[\mathbf{v}(t)(\mathbf{u}(t))^T\boldsymbol{\mu}_i^2(t)] = 0 \text{ for } i \in \Sigma.$$

Since

$$E[\mathbf{v}(t)(\mathbf{u}(t))^T\boldsymbol{\mu}_i(t)\boldsymbol{\mu}_j(t)] = 0$$

for $i \neq j$ due to $\mathbf{w}$ being ZMWSSI (see [91, Lemma 7] or refer to the properties of a ZMWSSI process in Definition 2.3), and $\sum_{i=1}^{n_p}\alpha_i\boldsymbol{\mu}_i = 1$ for some $\{\alpha_i\}_{i=1}^{n_\mu}$, then

$$E[\mathbf{v}(t)(\mathbf{u}(t))^T] = \sum_{i,j=1}^{n_\mu}\alpha_i\alpha_j E[\mathbf{v}(t)(\mathbf{u}(t))^T\boldsymbol{\mu}_i(t)\boldsymbol{\mu}_j(t)] = 0.$$

That is, $\mathbf{v}(t)$ is orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$.
Since $\mathbf{w}(t)$ is a ZMWSSI, from [91, Lemma 7] it follows that

$$E[\mathbf{z}_w^{\mathbf{w}}(t)(\mathbf{z}_v^{\mathbf{w}}(t))^T] = 0$$

for all $v \in \Sigma^+$, $v \neq w$ or $v \notin L$ or $w \notin L$.
However, if $v = w \in L$ and σ is the first letter of w, then

$$E[\mathbf{z}_w^{\mathbf{w}}(t)(\mathbf{z}_w^{\mathbf{w}}(t))^T] = E[\mathbf{z}_\sigma^{\mathbf{w}}(t)(\mathbf{z}_\sigma^{\mathbf{w}}(t))^T].$$

Since $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_v^{\mathbf{u}}(t))^T]$ is the upper right block of $E[\mathbf{z}_w^{\mathbf{w}}(t)(\mathbf{z}_v^{\mathbf{w}}(t))^T]$, it follows that

1. $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_v^{\mathbf{u}}(t))^T] = 0$ if $v \neq w$
2. $E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T] = E[\mathbf{z}_\sigma^{\mathbf{v}}(t)(\mathbf{z}_\sigma^{\mathbf{u}}(t))^T] = \frac{1}{p_\sigma}E[\mathbf{u}(t-1)\mathbf{v}(t-1)\boldsymbol{\mu}_\sigma^2(t-1)]$ if $w \in L$,

where σ is the first letter of w, and from Definition 2.6, it follows that the latter expectation is zero. As $E[\mathbf{u}(t)\mathbf{v}(t)\boldsymbol{\mu}_\sigma^2(t)] = 0$, it follows that

$$E[\mathbf{z}_w^{\mathbf{v}}(t)(\mathbf{z}_v^{\mathbf{u}}(t))^T] = 0 \text{ for all } v \in \Sigma^+.$$

Since $\mathbf{w}(t)$ is a white noise w.r.t. $\boldsymbol{\mu}$, it follows that

$$E[\mathbf{z}^{\mathbf{w}}_{ws}(t)(\mathbf{z}^{\mathbf{w}}_{s}(t))^T] = 0,$$

for any $s \in \Sigma^+$. Furthermore, by [91, Lemma 7]

$$E[\mathbf{z}^{\mathbf{w}}_{w}(t)(\mathbf{w}(t))^T] = E[\mathbf{z}^{\mathbf{w}}_{ws}(t)(\mathbf{z}^{\mathbf{w}}_{s}(t))^T]$$

for any $s \in \Sigma^+$. Then

$$E[\mathbf{z}^{\mathbf{w}}_{w}(t)(\mathbf{w}(t))^T] = 0.$$

Since $E[\mathbf{z}^{\mathbf{v}}_{w}(t)(\mathbf{u}(t))^T]$ is the upper right block of $E[\mathbf{z}^{\mathbf{w}}_{w}(t)(\mathbf{w}(t))^T]$, then

$$E[\mathbf{z}^{\mathbf{v}}_{w}(t)(\mathbf{u}(t))^T] = 0.$$

Since $\mathbf{z}^{\mathbf{v}}_{w}(t)$ is orthogonal to the generators of $\mathcal{H}^{\mathbf{u}}_{t,+}$, the lemma follows. □

Lemma 3.2. *Define* $\mathbf{x}^d(t) = E_l[\mathbf{x}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}]$. *The entries of* $\mathbf{x}^d(t)$ *belong to* $\mathcal{H}^{\mathbf{u}}_{t}$ *and*

$$\mathbf{x}^d(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t), \tag{3.15}$$

where the convergence is in the mean square sense.

Proof. It is clear from the definition that the components of $\mathbf{x}^d(t)$ belong to $\mathcal{H}^{\mathbf{u}}_{t,+}$ because it is the orthogonal projection of $\mathbf{x}(t)$ onto the subspace $\mathcal{H}^{\mathbf{u}}_{t,+}$.
From Lemma 3.1 it follows that,

$$E_l[\mathbf{z}^{\mathbf{v}}_{\sigma w}(t) \mid H^{\mathbf{u}}_{t,+}] = 0$$

and since the components of $\mathbf{z}^{\mathbf{u}}_{\sigma w}(t)$ belong to $\mathcal{H}^{\mathbf{u}}_{t,+}$, it follows that

$$E_l[\mathbf{z}^{\mathbf{u}}_{\sigma w}(t) \mid H^{\mathbf{u}}_{t,+}] = \mathbf{z}^{\mathbf{u}}_{\sigma w}(t).$$

Recall (2.10) from Chapter 2 that

$$\mathbf{x}(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} A_w K_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t) + \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t).$$

Since (2.10) holds and the map $z \mapsto E_l[z \mid M]$ (where $z \in \mathcal{H}_1$) is a continuous linear operator for any closed subspace M, it follows that $\mathbf{x}^d(t)$ will have the following expression:

$$\begin{aligned}
\mathbf{x}^d(t) &= E_l[\mathbf{x}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}] \\
&= \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} A_w \left(K_\sigma E_l[\mathbf{z}^{\mathbf{v}}_{\sigma w}(t) \mid H^{\mathbf{u}}_{t,+}] + B_\sigma E_l[\mathbf{z}^{\mathbf{u}}_{\sigma w}(t) \mid H^{\mathbf{u}}_{t,+}] \right) \\
&= \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t)
\end{aligned}$$

i.e., (3.15) holds.
Since the components of $\mathbf{z}^{\mathbf{u}}_{\sigma w}(t)$ belong to $\mathcal{H}^{\mathbf{u}}_t$, the components of the right-hand side of (3.15) belong to $\mathcal{H}^{\mathbf{u}}_t$ and hence the components of $\mathbf{x}^d(t)$ belong to $\mathcal{H}^{\mathbf{u}}_t$.
The convergence of the right-hand side of (3.15) in the mean square sense follows from Lemma 2.2. $\square$

Lemma 3.3. *Define $\mathbf{x}^s(t) = \mathbf{x}(t) - \mathbf{x}^d(t)$. The entries of $\mathbf{x}^s(t)$ belong to $\mathcal{H}^{\mathbf{v}}_t$, they are orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$ and*

$$\mathbf{x}^s(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} A_w K_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t), \tag{3.16}$$

where the sum converges in the mean-square sense.

Proof. From (3.15) and (2.10), we can write:

$$\begin{aligned}\mathbf{x}^s(t) &= \mathbf{x}(t) - \mathbf{x}^d(t) \\ &= \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \Big[\sqrt{p_{\sigma w}} A_w K_\sigma \mathbf{z}^{\mathbf{v}}_{\sigma w}(t) + \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t) - \sqrt{p_{\sigma w}} A_w B_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t) \Big].\end{aligned}$$

It follows that (3.16) holds, and it is clear that $\mathbf{x}^s(t)$ belong to the subspace $\mathcal{H}^{\mathbf{v}}_t$.
By Lemma 3.1, $\{\mathbf{z}^{\mathbf{v}}_w(t)\}_{w \in \Sigma^+}$ are orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$, hence all the summands in the right-hand side of (3.16) are orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$. $\square$

Proof of Theorem 3.1. The proof is an extension of [73, proof of Lemma 1]. Since

$$\sum_{\sigma \in \Sigma} p_\sigma A_\sigma \otimes A_\sigma \text{ is Schur}$$

and $\mathbf{u}$ and $\mathbf{v}$ are white noise processes, from (3.15)-(3.16) and Lemma 2.2 it follows that $\mathbf{x}^d$ is the unique state process of $\mathcal{S}_d$ and $\mathbf{x}^s$ is the unique state process of $\mathcal{S}_s$. Notice that

$$\mathbf{y}^d(t) = C E_l[\mathbf{x}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}] + D E_l[\mathbf{u}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}] + F E_l[\mathbf{v}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}].$$

By Lemma 3.1, $\mathbf{v}(t)$ is orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$, $E_l[\mathbf{v}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}] = 0$ and as the components $\mathbf{u}(t)$ belong to $\mathcal{H}^{\mathbf{u}}_{t,+}$, $E_l[\mathbf{u}(t) \mid \mathcal{H}^{\mathbf{u}}_{t,+}] = \mathbf{u}(t)$. Hence,

$$\mathbf{y}^d(t) = C\mathbf{x}^d(t) + D\mathbf{u}(t)$$

and

$$\mathbf{y}^s(t) = \mathbf{y}(t) - \mathbf{y}^d(t) = C\mathbf{x}(t) + D\mathbf{u}(t) + F\mathbf{v}(t) - C\mathbf{x}^d(t) - D\mathbf{u}(t) = C\mathbf{x}^s(t) + F\mathbf{v}(t).$$

That is, $\mathcal{S}_d$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$ and $\mathcal{S}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ respectively. $\square$

3.5.2 Proof of Theorem 3.2

In order to prove Theorem 3.2, we need to present Lemma 3.4, Lemma 3.5 and Lemma 3.6. To this end, assume that $\mathcal{S}$ of the form (2.1) is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.

Lemma 3.4. *The entries of* $\mathbf{y}^s(t), \mathbf{e}^s(t)$ *and* $\{\mathbf{z}_v^{\mathbf{y}^s}(t), \mathbf{z}_v^{\mathbf{e}^s}(t)\}_{v\in\Sigma^+}$ *belong to* $\mathcal{H}_{t,+}^{\mathbf{v}}$.

Proof. Recall from the proof of Theorem 3.1 that

$$\mathbf{y}^s(t) = C\mathbf{x}^s(t) + \mathbf{v}(t).$$

Then by (3.16), the components of $\mathbf{y}^s(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{v}}$.
Then by [91, Lemma 11], the components of $\mathbf{z}_v^{\mathbf{y}^s}(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{v}}$ and hence, $\mathcal{H}_t^{\mathbf{y}^s} \subseteq \mathcal{H}_{t,+}^{\mathbf{v}}$. Since

$$\mathbf{e}^s(t) = \mathbf{y}^s(t) - E_l[\mathbf{y}^s(t) \mid \mathcal{H}_t^{\mathbf{y}^s}],$$

this then implies that the components of $\mathbf{e}^s(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{v}}$. Since

$$\mathbf{z}_v^{\mathbf{v}}(t) = \sum_{i=1}^{n_\mu} \alpha_i \mathbf{z}_{vi}^{\mathbf{v}}(t+1), \quad \mathbf{v}(t) = \sum_{i=1}^{n_\mu} \alpha_i \mathbf{z}_i^{\mathbf{v}}(t+1)$$

as $\sum_{i=1}^{n_\mu} \alpha_i \boldsymbol{\mu}_i = 1$, it follows that $\mathcal{H}_{t,+}^{\mathbf{v}} \subseteq \mathcal{H}_{t+1}^{\mathbf{v}}$ and from [91, Lemma 11] it follows that the components of $\mathbf{z}_v^{\mathbf{e}^s}(t)$ belong to $\mathcal{H}_{t-1,+}^{\mathbf{v}} \subseteq \mathcal{H}_t^{\mathbf{v}} \subseteq \mathcal{H}_{t,+}^{\mathbf{v}}$. □

Lemma 3.5. *The entries of* $\{\mathbf{z}_v^{\mathbf{y}^s}(t), \mathbf{z}_v^{\mathbf{e}^s}(t)\}_{v\in\Sigma^+}$, $\mathbf{y}^s(t)$ *and* $\mathbf{e}^s(t)$ *are orthogonal to* $\mathcal{H}_{t,+}^{\mathbf{u}}$.

Proof. The coordinates of $\mathbf{y}^s(t)$, $\mathbf{e}^s(t)$, $\{\mathbf{z}_v^{\mathbf{y}^s}(t), \mathbf{z}_v^{\mathbf{e}^s}(t)\}_{v\in\Sigma^+}$ belong to $\mathcal{H}_{t,+}^{\mathbf{v}}$, and by Lemma 3.1 the elements of $\mathcal{H}_{t,+}^{\mathbf{v}}$ are orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$. □

Lemma 3.6. $\mathbf{r} = \begin{bmatrix}(\mathbf{e}^s)^T & \mathbf{u}^T\end{bmatrix}^T$ *is a white noise process w.r.t.* $\boldsymbol{\mu}$ *and* $E[\mathbf{e}^s(t)\mathbf{u}^T(t)\boldsymbol{\mu}_\sigma^2(t)] = 0$ *for all* $\sigma \in \Sigma$.

Proof. Recall that a white noise process w.r.t. the switching is, first, a ZMWSSI process. Therefore, we first show that $\mathbf{r}$ is a ZMWSSI, by showing that $\mathbf{r}$ satisfies the conditions of Definition 2.3 one by one.
• **Part 2. of Definition 2.3.** By assumption, $\mathbf{u}$ is a ZMWSSI and white noise process w.r.t. $\boldsymbol{\mu}$.
From the fact that $\hat{\mathcal{S}}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$ it follows that as $\mathbf{e}^s$ is a noise process of the asGLSS $\hat{\mathcal{S}}_s$, it is a ZMWSII and a white noise w.r.t. $\boldsymbol{\mu}$.
Thus $\mathbf{e}^s(t), \mathbf{u}(t)$ and $\{\mathbf{z}_w^{\mathbf{e}^s}(t), \mathbf{z}_w^{\mathbf{u}}(t)\}_{w\in\Sigma^+}$ are zero mean, square integrable, and hence so are $\mathbf{r}(t)$ and $\mathbf{z}_w^{\mathbf{r}}(t)$.
Finally, we show that $\mathbf{r}(t)$ and $\mathbf{z}_w^{\mathbf{r}}(t)$, for all $w \in \Sigma^+$, are jointly wide-sense stationary, i.e., for all $s, t, \in \mathbb{Z}$, $E[\mathbf{h}_1(t)(\mathbf{h}_1(s)^T)]$ depends on $t-s$, where $\mathbf{h}_1(\tau), \mathbf{h}_2(\tau) \in \{\mathbf{r}(\tau)\} \cup \{\mathbf{z}_w^{\mathbf{r}}(\tau)\}_{w\in\Sigma^+}$.
Note that, we show only the case where

$$E\left[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(s))^T\right] = E\left[\mathbf{z}_w^{\mathbf{r}}(t-s)(\mathbf{z}_v^{\mathbf{r}}(0))^T\right].$$

The proof of the general case is similar.
By Lemma 3.5,

$$E\left[\mathbf{z}_w^{\mathbf{e}^s}(t)(\mathbf{z}_v^{\mathbf{u}}(s))^T\right] = 0,$$

and hence $E\left[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(s))^T\right]$ is a block-diagonal matrix:

$$\begin{aligned} E\left[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(s))^T\right] &= \begin{bmatrix} E\left[\mathbf{z}_w^{\mathbf{e}^s}(t)(\mathbf{z}_v^{\mathbf{e}^s}(s))^T\right] & E\left[\mathbf{z}_w^{\mathbf{e}^s}(t)(\mathbf{z}_v^{\mathbf{u}}(s))^T\right] \\ E\left[\mathbf{z}_w^{\mathbf{u}}(t)(\mathbf{z}_v^{\mathbf{e}^s}(s))^T\right] & E\left[\mathbf{z}_w^{\mathbf{u}}(t)(\mathbf{z}_v^{\mathbf{u}}(s))^T\right] \end{bmatrix}, \\ &= \begin{bmatrix} E\left[\mathbf{z}_w^{\mathbf{e}^s}(t)(\mathbf{z}_v^{\mathbf{e}^s}(s))^T\right] & 0 \\ 0 & E\left[\mathbf{z}_w^{\mathbf{u}}(t)(\mathbf{z}_v^{\mathbf{u}}(s))^T\right] \end{bmatrix}. \end{aligned}$$

As $\mathbf{u}$ and $\mathbf{e}^s$ are ZMWSSI, the diagonal blocks depend on $t-s$:

$$\begin{aligned} E\left[\mathbf{z}_w^{\mathbf{r}}(t)(\mathbf{z}_v^{\mathbf{r}}(s))^T\right] &= \begin{bmatrix} E\left[\mathbf{z}_w^{\mathbf{e}^s}(t-s)(\mathbf{z}_v^{\mathbf{e}^s}(0))^T\right] & 0 \\ 0 & E\left[\mathbf{z}_w^{\mathbf{u}}(t-s)(\mathbf{z}_v^{\mathbf{u}}(0))^T\right] \end{bmatrix} \\ &= E\left[\mathbf{z}_w^{\mathbf{r}}(t-s)(\mathbf{z}_v^{\mathbf{r}}(0))^T\right] \end{aligned}$$

• **Part 1. of Definition 2.3.** By Lemma 3.4, the entries $\mathbf{e}^s(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{v}}(t)$. Moreover, by definition of sGLSS, $\mathbf{w} = [\mathbf{v}^T, \mathbf{u}^T]^T$ is ZMWSSI. Hence, the σ-algebras $\mathcal{F}_t^{\mathbf{w}}$ and $\mathcal{F}_t^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}_t^{\boldsymbol{\mu},-}$. Since $\mathbf{e}^s(t)$ belongs to $\mathcal{H}_{t,+}^{\mathbf{v}}$, $\mathbf{e}^s(t)$ is measurable with respect to the σ-algebra generated by $\{\mathbf{v}(t)\} \cup \{\mathbf{z}_v^{\mathbf{v}}(t)\}_{v\in\Sigma^+}$ and the latter σ-algebra is a subset of $\mathcal{F}_t^{\mathbf{w}} \vee \mathcal{F}_t^{\boldsymbol{\mu},-}$, where $\mathcal{F}_1 \vee \mathcal{F}_2$ denotes the smallest σ-algebra generated by the union of sets $\mathcal{F}_1$ and $\mathcal{F}_2$. That is, $\mathbf{e}^s(t)$ is measurable w.r.t. the σ algebra $\mathcal{F}_t^{\mathbf{w}} \vee \mathcal{F}_t^{\boldsymbol{\mu},-}$. Hence, $\mathcal{F}_t^{\mathbf{r}} \subseteq \mathcal{F}_t^{\mathbf{w}} \vee \mathcal{F}_t^{\boldsymbol{\mu},-}$. Since $\mathcal{F}_t^{\mathbf{w}}$ and $\mathcal{F}^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}_t^{\boldsymbol{\mu},-}$, from [120, Proposition 2.4] it follows that $\mathcal{F}_t^{\mathbf{w}} \vee \mathcal{F}_t^{\boldsymbol{\mu},-}$ and $\mathcal{F}_t^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}_t^{\boldsymbol{\mu},-}$, and as $\mathcal{F}_t^{\mathbf{r}} \subseteq \mathcal{F}_t^{\mathbf{w}} \vee \mathcal{F}_t^{\boldsymbol{\mu},-}$, it follows that $\mathcal{F}_t^{\mathbf{r}}$ and $\mathcal{F}^{\boldsymbol{\mu},+}$ are conditionally independent w.r.t. $\mathcal{F}_t^{\boldsymbol{\mu},-}$.

So far, we have shown that $\mathbf{r}$ is a ZMWSSI process. Next, we show that $\mathbf{r}$ is a white noise process w.r.t. $\boldsymbol{\mu}$. To this end, by [91, Lemma 7], it is enough to show that

$$E[\mathbf{r}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T] = 0$$

for all $w \in \Sigma^+$. We know that:

$$E\left[\mathbf{r}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T\right] = \begin{bmatrix} E\left[\mathbf{e}^s(t)(\mathbf{z}_w^{\mathbf{e}^s}(t))^T\right] & E\left[\mathbf{e}^s(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T\right] \\ E\left[\mathbf{u}(t)(\mathbf{z}_w^{\mathbf{e}^s}(t))^T\right] & E\left[\mathbf{u}(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T\right] \end{bmatrix}.$$

Since $\mathbf{e}^s$ and $\mathbf{u}$ are white noise w.r.t $\boldsymbol{\mu}$, then:

$$E\left[\mathbf{r}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T\right] = \begin{bmatrix} 0 & E\left[\mathbf{e}^s(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T\right] \\ E\left[\mathbf{u}(t)(\mathbf{z}_w^{\mathbf{e}^s}(t))^T\right] & 0 \end{bmatrix},$$

and by Lemma 3.5, $\mathbf{e}^s(t)$ and $\mathbf{z}_w^{\mathbf{e}^s}(t)$ are orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$, in particular, we get

$$E\left[\mathbf{e}^s(t)(\mathbf{z}_w^{\mathbf{u}}(t))^T\right] = 0 \quad \text{and} \quad E\left[\mathbf{u}(t)(\mathbf{z}_w^{\mathbf{e}^s}(t))^T\right] = 0.$$

Then,

$$E\left[\mathbf{r}(t)(\mathbf{z}_w^{\mathbf{r}}(t))^T\right] = 0.$$

Finally, by Lemma 3.5 again, as $\mathbf{z}_w^{\mathbf{e}^s}(t)$ is orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$, and as the latter contains $\mathbf{z}_\sigma^{\mathbf{u}}(t)$, then

$$E\left[\mathbf{z}_\sigma^{\mathbf{e}^s}(t)\mathbf{z}_\sigma^{\mathbf{u}}(t)\right] = 0.$$

Because $\mathbf{r}$ is ZMWSSI, then, in particular:

$$E\left[\mathbf{z}_\sigma^{\mathbf{e}^s}(t+1)\mathbf{z}_\sigma^{\mathbf{u}}(t+1)\right] = E\left[\mathbf{z}_\sigma^{\mathbf{e}^s}(t)\mathbf{z}_\sigma^{\mathbf{u}}(t)\right] = 0.$$

It follows that, $E[\mathbf{e}^s(t)\mathbf{u}^T(t)\boldsymbol{\mu}_\sigma^2(t)] = 0$ for all $\sigma \in \Sigma$ and $t \in \mathbb{Z}$. □

Proof of Theorem 3.2. The proof of Theorem 3.2 is an extension of [73, proof of Lemma 2].
To understand this proof, recall the definition of a sGLSS (Definition 2.6).
From Lemma 3.6, it follows that $\mathbf{r} = \begin{bmatrix}(\mathbf{e}^s)^T & \mathbf{u}^T\end{bmatrix}^T$ is a ZMWSSI process. Also, it follows that the noise process $\mathbf{e}^s$ and the input $\mathbf{u}$ satisfy the condition of

$$E[\mathbf{e}^s(t)(\mathbf{u}(t))^T\boldsymbol{\mu}_\sigma^2(t)] = 0, \text{ for } \sigma \in \Sigma.$$

This means that the first condition of Definition 2.6 is satisfied.
Since $\hat{\mathcal{S}}_s$ and $\hat{\mathcal{S}}_d$ are both asGLSS, it follows that $\sum_{i=1}^{n_\mu} p_i\hat{A}_i \otimes \hat{A}_i$ is stable. Then, the third condition of Definition 2.6 is satisfied.
The fourth condition of the definition is satisfied by assumption, as $\hat{\mathcal{S}}_d$ and $\hat{\mathcal{S}}_s$ are both assumed to be asGLSS. More precisely, due to the latter assumption, $\hat{A}_{\sigma_2}\hat{A}_{\sigma_1} = 0$, for $\sigma_1, \sigma_2 \in \Sigma$ and $\sigma_1\sigma_2 \notin L$, and:

$$\begin{aligned}\hat{A}_{\sigma_2}\hat{B}_{\sigma_1}E\left[\mathbf{z}_{\sigma_1}^{\mathbf{u}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{u}}(t))^T\right] = 0,\\ \hat{A}_{\sigma_2}\hat{K}_{\sigma_1}E\left[\mathbf{z}_{\sigma_1}^{\mathbf{e}^s}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{e}^s}(t))^T\right] = 0.\end{aligned}$$

By Lemma 3.6, for $\mathbf{w} = \begin{bmatrix}\mathbf{u}^T & (\mathbf{e}^s)^T\end{bmatrix}^T$, we get:

$$E\left[\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t))^T\right] = \begin{bmatrix}E\left[\mathbf{z}_{\sigma_1}^{\mathbf{u}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{u}}(t))^T\right] & 0\\ 0 & E\left[\mathbf{z}_{\sigma_1}^{\mathbf{e}^s}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{e}^s}(t))^T\right]\end{bmatrix} = \begin{bmatrix}T_{\sigma_1,\sigma_1}^{\mathbf{u},\mathbf{u}} & 0\\ 0 & T_{\sigma_1,\sigma_1}^{\mathbf{e}^s,\mathbf{e}^s}\end{bmatrix}.$$

Then we can write:

$$\begin{aligned}\hat{A}_{\sigma_2}\begin{bmatrix}\hat{B}_{\sigma_1} & \hat{K}_{\sigma_1}\end{bmatrix}E[\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{w}}(t))^T] &= \hat{A}_{\sigma_2}\begin{bmatrix}\hat{B}_{\sigma_1} & \hat{K}_{\sigma_1}\end{bmatrix}\begin{bmatrix}T_{\sigma_1,\sigma_1}^{\mathbf{u},\mathbf{u}} & 0\\ 0 & T_{\sigma_1,\sigma_1}^{\mathbf{e}^s,\mathbf{e}^s}\end{bmatrix}\\ &= \begin{bmatrix}\hat{A}_{\sigma_2}\hat{B}_{\sigma_1}E\left[\mathbf{z}_{\sigma_1}^{\mathbf{u}}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{u}}(t))^T\right] & \hat{A}_{\sigma_2}\hat{K}_{\sigma_1}E\left[\mathbf{z}_{\sigma_1}^{\mathbf{e}^s}(t)(\mathbf{z}_{\sigma_1}^{\mathbf{e}^s}(t))^T\right]\end{bmatrix} = 0.\end{aligned}$$

Next, the second condition of Definition 2.6 is verified. Let $\hat{\mathbf{x}}^s$ and $\hat{\mathbf{x}}^d$ be the unique state processes of $\hat{\mathcal{S}}_s$ and $\hat{\mathcal{S}}_d$ respectively. We know that:

$$\begin{aligned}\hat{\mathbf{x}}(t) &= \hat{\mathbf{x}}^d(t) + \hat{\mathbf{x}}^s(t) \\ &= \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} \hat{A}_w \hat{B}_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t) + \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} \hat{A}_w \hat{K}_\sigma \mathbf{z}^{\mathbf{e}^s}_{\sigma w}(t) \\ &= \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma,\ w \in \Sigma^* \\ \sigma w \in L,\ |w| = N}} \sqrt{p_{\sigma w}} \hat{A}_w \begin{bmatrix} \hat{B}_\sigma & \hat{K}_\sigma \end{bmatrix} \begin{bmatrix} \mathbf{z}^{\mathbf{u}}_{\sigma w}(t) \\ \mathbf{z}^{\mathbf{e}^s}_{\sigma w}(t) \end{bmatrix}\end{aligned}$$

Then, by Lemma 2.2, $\hat{\mathbf{x}}(t)$ is the unique state process of $\hat{\mathcal{S}}$. Indeed,

$$\hat{\mathbf{x}}(t+1) = \sum_{i=1}^{n_\mu} (\hat{A}_i \hat{\mathbf{x}}(t) + \hat{B}_i \mathbf{u}(t) + \hat{K}_i \mathbf{e}^s(t)) \boldsymbol{\mu}_i(t)$$

holds and $\hat{\mathbf{x}}(t)$ is a ZMWSSI, as it is a sum of two ZMWSSI processes. Hence, the second condition of Definition 2.6 is satisfied and $\hat{\mathbf{x}}(t)$ is the unique state process of $\hat{\mathcal{S}}$. Finally, from

$$\mathbf{y}^d(t) = \hat{C}\hat{\mathbf{x}}^d(t) + \hat{D}\mathbf{u}(t)$$

(as $\hat{\mathcal{S}}_d$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$) and

$$\mathbf{y}^s(t) = \hat{C}\hat{\mathbf{x}}^s(t) + \mathbf{e}^s(t)$$

(as $\hat{\mathcal{S}}_s$ is an asGLSS of $(\mathbf{y}^s, \boldsymbol{\mu})$), it follows that $\mathbf{y}(t) = \hat{C}\hat{\mathbf{x}}(t) + \hat{D}\mathbf{u}(t) + \mathbf{e}^s(t)$, i.e., $\hat{\mathcal{S}}$ is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ since all the conditions of Definition 2.6 are satisfied. □

Chapter 4

On minimal GLSS state-space representations in innovation form: an algebraic characterization

4.1 Introduction

In this chapter, autonomous stationary Generalized Linear Switch System (asGLSS) are considered. *The main technical contributions are new algebraic conditions for an asGLSS to be minimal in innovation form.* These conditions depend only on the matrices of the system representation.

4.1.1 Motivation for studying asGLSS

Under suitable technical assumptions discussed in Chapter 3 any stochastic GLSS can be decomposed into a noiseless deterministic GLSS which is driven only by the control input, and asGLSS driven only by the noise. Moreover, at least for specific subclasses of GLSS, the identification of these two subsystems can be carried out separately [72]. Note that minimality and uniqueness of deterministic LPV state-space representations is well understood, [86], and the analogy can be easily made for the more global GLSS class. Hence, in order to understand the notion of minimality and innovation representation for stochastic GLSS with control input, the first step is to understand these notions for asGLSS. In turn, as it was argued in Chapter 1, minimal GLSSs in innovation form are instrumental for system identification

Recall from Section 1.6 of Chapter 1, that in the formulation of the identification problem for a class of GLSS state-space representation [122, 30, 125, 127, 21] the stated goal is usually to find a GLSS which is either isomorphic or at least input-output equivalent to the data generating system. In the former case, the isomorphism is often required to be linear and independent of the switching signal. However, in general there may exist GLSSs (in fact asGLSSs) which generate the same output for some switching signal, but which are not isomorphic (see Example 4.2 in Section 4.3 or

[58]), or which are not input-output equivalent. In the latter case the systems generate different outputs when the switching signal is changed (see Example 4.1 and Example 4.2 of Section 4.3 of this chapter).

This implies that in general the system identification problem for GLSSs, in fact, even for the subclass of asGLSSs, is ill-posed. However, it becomes well-posed, if we add the assumption that the underlying data generating system is a minimal asGLSS in innovation form.

Indeed, if the observed output has an asGLSS representation, it has a minimal one in innovation form, and all such representations are related by an isomorphism (independent of the switching signal). The necessity of this assumption is illustrated by the examples in Section 4.3. From Chapter 2 it follows that this assumption is sufficient, as the identification algorithm in [74] returns a minimal asGLSS in innovation form. We conjecture that the same will be true for most of the existing subspace identification algorithms [122, 30, 125, 127, 21, 27].

To deal only with minimal asGLSS representations in innovation form, simple conditions to check minimality and being in innovation form are needed. The latter is necessary in order to check if the elements of a parametrization of asGLSS are minimal and in innovation form, or to construct such parametrizations.

4.1.2 Related work to this chapter

It should be noted that there is a rich literature on subspace identification methods for stochastic Linear Parameter-Varying state-space representations [27, 29, 21, 122]. The results of the cited papers can be also applied to a GLSS. However, the cited papers do not deal with the problem of characterizing minimal stochastic GLSS state-space representations in innovation form. In [19, 21] the existence of an LPV state-space representation in innovation form was studied, but due to the specific assumptions (deterministic scheduling) and the definition of the innovation process, the resulting LPV state-space representation in innovation form had dynamic dependence on the scheduling parameters. Moreover, [19, 21] do not address the issue of minimality of the stochastic part of LPV state-space representations.

As it was mentioned in Chapter 2, Subsection 2.7.4, an asGLSS can be viewed as **GBS**, and stochastic realization theory of asGLSS follows from that of **GBS** [91]. The main novelty presented in this chapter with respect to [91] is the new algebraic characterization of minimal asGLSSs in innovation form, and that the results on existence and uniqueness of minimal **GBS**s are spelled out explicitly for GLSSs.

The paper [74] used the correspondence between **GBS**s and asGLSSs to state existence and uniqueness of minimal asGLSSs in innovation form. However, [74] did not provide an algebraic characterization of minimality or innovation form. Moreover, it considered only switching signals which were zero mean white noises. In contrast, in this chapter, more general switching signals are considered. The reader can consider this chapter as a complementary to [74]: here, we explain when the assumption that the data generating system is minimal asGLSS in innovation form could be true, while [74] presents an identification algorithm which is statistically consistent under the latter assumption.

4.1.3 Outline of the chapter

In Section 4.2 we present the first contribution of the thesis, namely, algebraic conditions for an asGLSS to be minimal in innovation form. Finally, in Section 4.3 numerical examples are developed to illustrate the contribution. These examples also demonstrate the problems which may arise when systems that are not in innovation form are used.

4.2 Main results: algebraic conditions for an asGLSS to be minimal in innovation form

For the main results, we invite the reader to recall the preliminaries in Chapter 2, especially Remark 2.6. In order to present the main contribution of the chapter, we will have to assume the following:

1. We are working with asGLSSs of the form (2.4), for which, in particular, Definition 2.7 from Section 2.5 holds.

2. The switching is considered to be admissible (see Definition 2.2 of Subsection 2.4.1).

Motivated by the challenges explained in Remark 2.6, we present, in this section, sufficient conditions for an asGLSS to be minimal and in innovation form. These conditions depend only on the matrices of the asGLSS in question and do not require any information on the noise processes.

The first result concerns an algebraic characterization of asGLSS in innovation form. In order to streamline the discussion, we introduce the following definition.

Definition 4.1 (Stably invertible w.r.t. $\boldsymbol{\mu}$)**.** *Assume that $\mathcal{S}$ is an asGLSS of the form* (2.4) *and $F = I_{n_y}$. We will call $\mathcal{S}$* stably invertible with respect to $\boldsymbol{\mu}$*, or* stably invertible *if $\boldsymbol{\mu}$ is clear from the context, if the matrix*

$$\sum_{i=1}^{n_\mu} p_i(A_i - K_iC) \otimes (A_i - K_iC) \tag{4.1}$$

is Schur.

Note that a system can be stably invertible w.r.t. one switching process, and not stably invertible w.r.t. another one. We can now state the result relating stable invertibility to asGLSSs in innovation forms. To this end, recall Definition 2.4 and Definition 2.10 from Chapter 2.

Theorem 4.1 (Innovation form condition)**.** *Assume that $\mathbf{y}$ is SII and $\mathbf{y}$ is full rank w.r.t. $\boldsymbol{\mu}$. If an asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ is stably invertible, then it is in innovation form.*

The proof can be found in the Appendix 4.5.1.

Stably invertible asGLSSs can be viewed as optimal predictors. Indeed, let $\mathcal{S}$ be the asGLSS of the form (2.1) which is in innovation form. Hence, the following equation holds:

$$\begin{aligned}\mathbf{x}(t+1) &= \sum_{i=1}^{n_\mu}(A_i - K_iC)\mathbf{x}(t) + K_i\mathbf{y}(t))\boldsymbol{\mu}_i(t),\\ \hat{\mathbf{y}}(t) &= C\mathbf{x}(t)\end{aligned} \tag{4.2}$$

where $\mathbf{x}$ is the state process of $\mathcal{S}$, and $\hat{\mathbf{y}}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}]$, i.e., $\hat{\mathbf{y}}$ is the best linear prediction of $\mathbf{y}(t)$ based on the predictors $\{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}$. Intuitively, (4.2) could be viewed as a filter, i.e., a dynamical system driven by past values of $\mathbf{y}$ and generating the best possible linear prediction $\hat{\mathbf{y}}(t)$ of $\mathbf{y}(t)$ based on $\{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}$. However, the solution of (4.2) is defined on the whole time axis $\mathbb{Z}$ and hence cannot be computed exactly. For stably invertible asGLSS we can approximate $\hat{\mathbf{y}}(t)$ as follows.

Lemma 4.1. *With the assumptions of Theorem 4.1, if $\mathcal{S}$ of the form* (2.4) *is a stably invertible realization of* $(\mathbf{y}, \boldsymbol{\mu})$*, and we consider the following dynamical system:*

$$\begin{aligned}\bar{\mathbf{x}}(t+1) &= \sum_{i=1}^{n_\mu}(A_i - K_iC)\bar{\mathbf{x}}(t) + K_i\mathbf{y}(t))\boldsymbol{\mu}_i(t),\\ \bar{\mathbf{y}}(t) &= C\bar{\mathbf{x}}(t),\\ \bar{\mathbf{x}}(0) &= 0,\end{aligned} \tag{4.3}$$

then $\lim_{t\to\infty}(\bar{\mathbf{x}}(t) - \mathbf{x}(t)) = 0$*, and* $\lim_{t\to\infty}(\bar{\mathbf{y}}(t) - \hat{\mathbf{y}}(t)) = 0$*, where* $\hat{\mathbf{y}}(t) = E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}]$ *and the limits are understood in the mean square sense.*

The proof of Lemma 4.1 can be found in Appendix 4.5.1. That is, the output $\bar{\mathbf{y}}(t)$ of the recursive filter (4.3) is an approximation of the optimal prediction $\hat{\mathbf{y}}(t)$ of $\mathbf{y}(t)$ for large enough t. That is, stably invertible asGLSS not only result in asGLSSs in innovation form, but they represent a class of asGLSSs for which recursive filters of the form (4.3) exist.

Next, we present algebraic conditions for minimality of an asGLSS in innovation form. To this end, recall from Subsection 2.7.1 of Chapter 2 the definition of a dLSS and its minimality. With that definition in mind we can state the following.

Theorem 4.2 (Minimality condition in innovation form)**.** *Assume that $\mathcal{S}$ is an asGLSS of the form* (2.4) *and that $\mathcal{S}$ is a realization of* $(\mathbf{y}, \boldsymbol{\mu})$ *in innovation form. Assume that* $\mathbf{y}$ *is full rank and* $\mathbf{y}$ *is SII. Then $\mathcal{S}$ is a minimal realization of* $(\mathbf{y}, \boldsymbol{\mu})$*, if and only if the dLSS* $\mathcal{D}_{\mathcal{S}} = (\{A_i, K_i\}_{i=1}^{n_\mu}, C, I_{n_y})$ *is minimal.*

The proof of Theorem 4.2 can also be found in Appendix 4.5.1. Theorem 4.2, in combination with Theorem 4.1, leads to the following corollary.

Corollary 4.1 (Minimality and innovation form)**.** *With the assumptions of Theorem 4.2, if $\mathcal{D}_{\mathcal{S}}$ is minimal and $\mathcal{S}$ if stably invertible, then $\mathcal{S}$ is a minimal asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form.*

Remark 4.1 (Checking minimality and innovation form)**.** *Recall that Algorithm 1 presented in Subsection 2.7.1 of Chapter 2, can be used to check minimality of $\mathcal{D}_{\mathcal{S}}$. Checking that $\mathcal{S}$ is stably invertible boils down to checking the eigenvalues of the matrix* (4.1)*. That is, Corollary 4.1 provides effective procedure for verifying that an asGLSS is minimal and in innovation form. Note that in contrast to the rank condition of Theorem 2.3, the procedure above uses only the matrices of the system.*

Remark 4.2 (Parametrizations of asGLSSs)**.** *Below we will sketch some ideas for applying the above results to parametrizations of asGLSSs. A detailed study of these issues remains a topic for future research.*

*For all the elements of a parametrization of asGLSSs to be minimal and in innovation form, by Corollary 4.1 it is necessary that **(A)** all elements of the parametrization, when viewed as dLSS, are minimal, and that **(B)** they are stably invertible and satisfy condition **(3)** of Definition 2.6. In order to deal with **(A)**, the techniques used in [1] or [84] could be applied. The condition **(B)** is equivalent to stability of a suitable parametrization of LTI systems, as we could view the matrices* (4.1) *and $\sum_{q=1}^{n_\mu} p_i A_i \otimes A_i$ as matrices of an LTI state-space representation. For the latter we can use standard techniques (see [96] and the references therein).*

*Finally, note that extension of argument of [1, Theorem 2] would also lead to identifiability conditions for parameterizations which satisfy the conditions **(A)** and **(B)** described above.*

Corollary 4.1 suggests the following minimization algorithm.

Algorithm 4: Minimization algorithm,
Input: A stably invertible asGLSS $\mathcal{S}$.
Output : asGLSS $\mathcal{S}_m = (\{A_i^m, K_i^m\}_{i=1}^{n_\mu}, C^m, I_{n_y}, \mathbf{e})$.

1. Transform the dLSS $\mathcal{D}_{\mathcal{S}}$ from Theorem 4.2 to a minimal dLSS $\mathcal{D}_m = (\{A_i^m, K_i^m\}_{i=1}^{n_\mu}, C^m, I_{n_y})$ using Algorithm 2.

Theorem 4.3. (Correctness of Algorithm 4) The asGLSS $\mathcal{S}_m$ is stably invertible and it is a minimal asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form.

The proof of this Theorem 4.3 is presented in Appendix 4.5.2. That is, Algorithm 4 is a simpler counterpart of Algorithm 3 which does not require computing the limits (2.21) and (2.22). The latter can create difficulties (see Remark 2.6). Theorem 4.3 implies that stable invertibility is preserved by minimization.

4.3 Numerical examples and counter examples

In this section, we present numerical examples in order to illustrate the main results. More precisely, Example 4.1 is an example of an asGLSS $\mathcal{S}$ which is not minimal. After using Algorithm 3 from Chapter 2, we get a minimal asGLSS $\mathcal{S}^m$, which has the same output as $\mathcal{S}$ for the chosen $\boldsymbol{\mu}$. However, for another switching $\boldsymbol{\mu}'$, $\mathcal{S}$ and $\mathcal{S}^m$ have different outputs.

Example 4.2 demonstrates that the phenomenon observed in Example 4.1 is not due to non-minimality. In this case, the asGLSS is already minimal (but not in innovation form), yet the same output discrepancy occurs: if the asGLSS from Example 4.2 is transformed by Algorithm 3 to a minimal asGLSS $\mathcal{S}^m$ in innovation form, then $\mathcal{S}$ and $\mathcal{S}^m$ generate the same output for the switching process $\boldsymbol{\mu}$, under which Algorithm 3 was applied. However, $\mathcal{S}$ and $\mathcal{S}^m$ still generate different outputs if the switching is changed.

Finally, Example 4.3 shows that for minimal and stably invertible asGLSSs in innovation form the problems arising in the two previous examples are not present.

Example 4.1. *Consider an asGLSS $\mathcal{S}$ of the form* (2.4)*, where $n_\mu = 2$ and*

$$A_1 = \begin{bmatrix} 0.4 & 0.4 & 0 \\ 0.2 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \quad K_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 10 & 0 & 0 \end{bmatrix}$$
$$A_2 = \begin{bmatrix} 0.1 & 0.1 & 0 \\ 0.2 & 0.3 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad F = 1$$

The switching signal process is defined as $\boldsymbol{\mu} = [\boldsymbol{\mu}_1 \quad \boldsymbol{\mu}_2]$ such that $\boldsymbol{\mu}_1(t) = 1$ and $\boldsymbol{\mu}_2(t)$ is a white-noise process with uniform distribution $\mathcal{U}(-1.5, 1.5)$. This corresponds to the parameters values $\{p_\sigma\}_{\sigma \in \{1,2\}}$ to be $p_1 = E[\mu_1^2(t)] = 1$ and $p_2 = E[\mu_2^2(t)] = 0.75$. The noise process is a white Gaussian noise with a variance equal to 1, i.e., $\mathbf{v} \sim \mathcal{N}(0,1)$. Note that not all eigenvalues λ_j of $\sum_{i=1}^{n_\mu} p_i(A_i - K_iC) \otimes (A_i - K_iC)$ are in the unit disk for $j = 1, 2, \ldots, n_x \times n_x$:

$$\begin{aligned} &\lambda_1 = 0.04, && \lambda_2 = -2.5991 + 0.6951, && \lambda_3 = -2.5991 - 0.6951, \\ &\lambda_4 = 0.0450 + 0.3130i, && \lambda_5 = 0.0450 - 0.3130i, && \lambda_6 = 2.4850, \\ &\lambda_7 = 2.9183, && \lambda_8 = 0.0450 + 0.3130i, && \lambda_9 = 0.0450 - 0.3130i. \end{aligned}$$

This means that the latter matrix is not Schur, hence,

$$\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$$

is not stably invertible and it is not clear if it is in innovation form. Using Algorithm 3, we can find *a minimal asGLSS in innovation form with the following matrices:*

$$\begin{aligned}
A_1^m &= \begin{bmatrix} 0.4007 & 0.3997 \\ 0.1997 & 0.0993 \end{bmatrix}, \quad K_1^m = \begin{bmatrix} -0.046 \\ -0.0541 \end{bmatrix} \\
A_2^m &= \begin{bmatrix} 0.1003 & 0.1002 \\ 0.2002 & 0.2997 \end{bmatrix}, \quad K_2^m = \begin{bmatrix} -0.0116 \\ -0.0578 \end{bmatrix} \\
C^m &= \begin{bmatrix} -10 & 0.0116 \end{bmatrix}, \quad \textit{and} \quad F^m = 1
\end{aligned}$$

Note that

$$\mathcal{S}^m = (\{A_\sigma^m, K_\sigma^m\}_{\sigma=1}^{n_\mu}, C^m, F^m, \mathbf{e}, \boldsymbol{\mu})$$

is guaranteed to be a minimal asGLSS in innovation form, since it is the output of Algorithm 3. As the dimension of $\mathcal{S}$ is higher than the dimension of $\mathcal{S}^m$, it follows that $\mathcal{S}$ is not minimal. We observe that the two systems have the same output trajectory when the chosen switching is applied. This is well illustrated in Figure 4.1. The values for the Variance Accounted For (VAF) and the Best Fit Ratio (BFR) are 99.9711% and 98.2948% respectively.

*In order to estimate a sample path of the innovation process **e** we proceeded as follows. First, we chose sample paths of $\mathbf{v}$ and $\boldsymbol{\mu}$, and simulated $\mathcal{S}$ with these sample paths of $\mathbf{v}$ and $\boldsymbol{\mu}$ to obtain a sample path of $\mathbf{y}$. Then, we used*

$$\hat{\mathbf{x}}(t+1) = \sum_{i=1}^{n_\mu} ((A_i^m - K_i^m C^m)\hat{\mathbf{x}}(t) + K_i^m \mathbf{y}(t))\boldsymbol{\mu}_i(t), \quad \boldsymbol{e}(t) = \mathbf{y}(t) - C^m \hat{\mathbf{x}}(t) \tag{4.4}$$

*to obtain a sample path of **e** corresponding to the chosen sample paths of $\mathbf{y}$ and $\boldsymbol{\mu}$.*

*On the other hand, if we replace the switching process $\boldsymbol{\mu}$ by $\boldsymbol{\mu}'$, where $\boldsymbol{\mu}_1' = \boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2'$ is also a white-noise with a uniform distribution $\mathcal{U}(-\sqrt{3}, \sqrt{3})$, and we use the realization of $\mathbf{v}$ and **e** computed in the previous simulation, then the respective outputs of*

$$\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v}, \boldsymbol{\mu}') \quad \textit{and} \quad \mathcal{S}^m = (\{A_\sigma^m, K_\sigma^m\}_{\sigma=1}^{n_\mu}, C^m, F^m, \mathbf{e}, \boldsymbol{\mu}')$$

will be different, as illustrated in Figure 4.2. In other words, the two systems are not input-output equivalent. Subsequently, the two representations are not isomorphic.

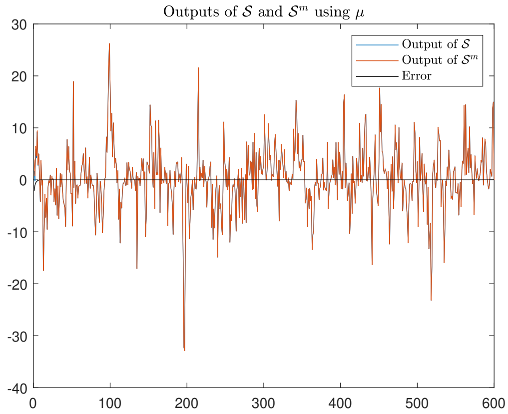

Figure 4.1: Outputs of $\mathcal{S}$ and $\mathcal{S}^m$ using μ

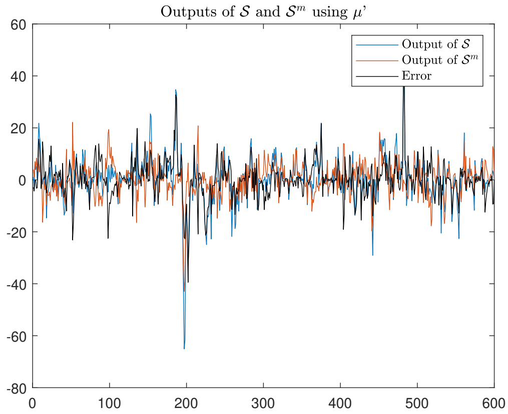

Figure 4.2: Outputs of $\mathcal{S}$ and $\mathcal{S}^m$ using μ'

Example 4.2. *We consider an asGLSS $\mathcal{S}$ of the form* (2.4)*, such that we use the same noise process* $\mathbf{v}$ *and switching process* $\boldsymbol{\mu}$ *as in Example 4.1. The system matrices are as follows.*

$$A_1 = \begin{bmatrix} 0.4 & 0.4 \\ 0.2 & 0.1 \end{bmatrix}, \quad K_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 10 & 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.1 & 0.1 \\ 0.2 & 0.3 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \textit{and} \quad F = 1$$

A direct computation reveals that

$$\sum_{i=1}^{n_\mu} p_i(A_i - K_iC) \otimes (A_i - K_iC) = \begin{bmatrix} 0.1675 & 0.1675 & 0.1675 & 0.1675 \\ -4.6550 & 0.0625 & -4.6550 & 0.0625 \\ -4.6550 & -4.6550 & 0.0625 & 0.0625 \\ 168.07 & -3.1850 & -3.1850 & 0.0775 \end{bmatrix}.$$

The eigenvalues λ_j of the above matrix are not all inside the unit disk for $j = 1, 2, \ldots, n_x \times n_x$ where

$$\begin{aligned} \lambda_1 &= -4.8790 + 1.2976i, & \lambda_2 &= -4.8790 - 1.2976i, \\ \lambda_3 &= 5.4104, & \lambda_4 &= 4.7175. \end{aligned}$$

This means that the latter matrix is not Schur, hence,

$$\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$$

is not stably invertible and it is not clear if it is in innovation form. As before, we use Algorithm 3 to obtain a minimal asGLSS innovation form with the following matrices

$$A_1^m = \begin{bmatrix} 0.4007 & -0.3997 \\ -0.1997 & 0.0993 \end{bmatrix}, \quad K_1^m = \begin{bmatrix} -0.046 \\ 0.0541 \end{bmatrix},$$

$$A_2^m = \begin{bmatrix} 0.1003 & -0.1002 \\ -0.2002 & 0.2997 \end{bmatrix}, \quad K_2^m = \begin{bmatrix} -0.0116 \\ 0.0578 \end{bmatrix}$$

$$C^m = \begin{bmatrix} -10 & -0.0116 \end{bmatrix}, \quad \textit{and} \quad F^m = 1$$

Note that

$$\mathcal{S}^m = (\{A_\sigma^m, K_\sigma^m\}_{\sigma=1}^{n_\mu}, C^m, F^m, \mathbf{e}, \boldsymbol{\mu})$$

is guaranteed to be a minimal asGLSS in innovation form, because it is the output of Algorithm 3. Since $\mathcal{S}$ and $\mathcal{S}^m$ have the same dimensions, it follows that $\mathcal{S}$ is minimal. When using the switching $\boldsymbol{\mu}$ (of the previous example) in Algorithm 3, the two systems get the same trajectory. This is illustrated in Figure 4.3. The VAF is equal to 99.97% and the BFR is equal to 98.3%.

On the other hand, $\mathcal{S}$ and $\mathcal{S}^m$ have different output trajectories, when using the switching process $\boldsymbol{\mu}'$ from Example 4.1 (as shown in Figure 4.4).

We used the same method as in Example 4.1 to simulate $\mathcal{S}$ and $\mathcal{S}_m$ both for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$, i.e., we chose a sample path of $\mathbf{v}$ and $\boldsymbol{\mu}$ in order to obtain a sample path of $\mathbf{y}$ by simulating $\mathcal{S}$, and then we used the sample path of $\mathbf{y}$ and (4.4) to obtain a sample path of the innovation process $\boldsymbol{e}$. We used the same sample paths of $\mathbf{v}$ and $\boldsymbol{e}$ both for simulating the systems for $\boldsymbol{\mu}$ and for $\boldsymbol{\mu}'$.

In particular, this means that the two systems are not isomorphic. From Example 4.2, we can conclude that minimality alone does not provide uniqueness.

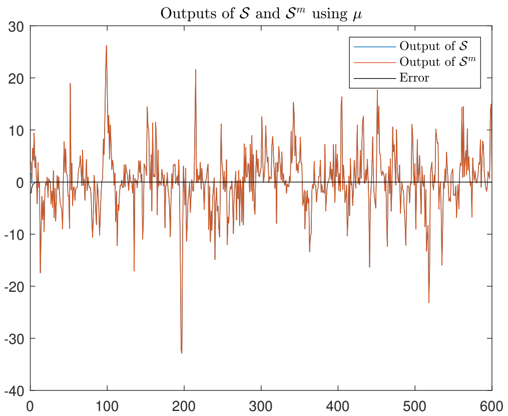

Figure 4.3: Outputs of $\mathcal{S}$ and $\mathcal{S}^m$ using μ

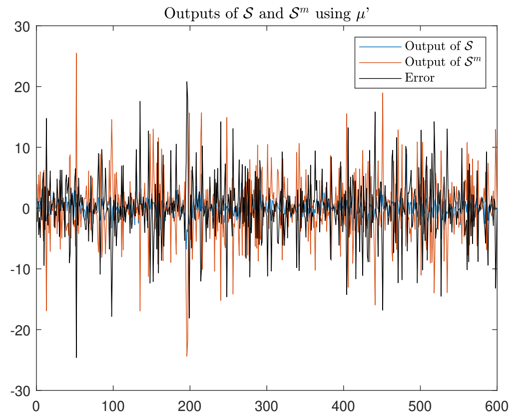

Figure 4.4: Outputs of $\mathcal{S}$ and $\mathcal{S}^m$ using μ'

Example 4.3. *We present a minimal asGLSS $\mathcal{S}$ in innovation form, where the state dimension $n_x = 2$ and its matrices are as follows:*

$$A_1 = \begin{bmatrix} 0.4 & 0.4 \\ 0.2 & 0.1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.1 & 0.1 \\ 0.2 & 0.3 \end{bmatrix}$$
$$K_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \textit{and} \quad F = 1$$

We use the same switching process $\boldsymbol{\mu}$ and noise process $\mathbf{v}$ as in Example 4.1, i.e., $p_1 = E\left[\mu_1^2(t)\right] = 1$,

$p_2 = E\left[\mu_2^2(t)\right] = 0.75$ *and* $\mathbf{v} \sim \mathcal{N}(0,1)$. *A direct computation reveals that*

$$\sum_{i=1}^{n_\mu} p_i(A_i - K_iC) \otimes (A_i - K_iC) = \begin{bmatrix} 0.1675 & 0.1675 & 0.1675 & 0.1675 \\ -0.38 & 0.0625 & -0.38 & 0.0625 \\ -0.38 & -0.38 & 0.0625 & 0.0625 \\ 1.12 & -0.26 & -0.26 & 0.0775 \end{bmatrix}.$$

The eigenvalues λ_j *of the above matrix are not all inside the unit disk for* $j = 1, 2, \ldots, n_x \times n_x$ *where*

$$\begin{aligned} &\lambda_1 = 0.5385, && \lambda_2 = -0.3055 + 0.3783i, \\ &\lambda_3 = -0.3055 - 0.3783i, && \lambda_4 = 0.4425. \end{aligned}$$

This means that the latter matrix is Schur, hence,

$$\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, F, \mathbf{v})$$

is stably invertible and it is in innovation form. If we apply Algorithm 3 to the system above, we get another asGLSS $\mathcal{S}^m$ *with the matrices*

$$\begin{aligned} A_1^m &= \begin{bmatrix} 0.4642 & -0.3581 \\ -0.1581 & 0.0358 \end{bmatrix}, \quad K_1^m = \begin{bmatrix} -0.1143 \\ 0.9934 \end{bmatrix} \\ A_2^m &= \begin{bmatrix} 0.1367 & -0.1188 \\ -0.2188 & 0.2633 \end{bmatrix}, \quad K_2^m = \begin{bmatrix} -0.1143 \\ 0.9934 \end{bmatrix} \\ C^m &= \begin{bmatrix} -0.9934 & -0.1143 \end{bmatrix}, \quad \textit{and} \quad F^m = 1 \end{aligned}$$

Note that

$$\mathcal{S}^m = (\{A_\sigma^m, K_\sigma^m\}_{\sigma=1}^{n_\mu}, C^m, F^m, \mathbf{e}, \boldsymbol{\mu})$$

is guaranteed to be a minimal asGLSS in innovation form, because it is the output of Algorithm 3. Since $\mathcal{S}$ *and* $\mathcal{S}^m$ *have the same dimensions, it follows that* $\mathcal{S}$ *is minimal. Moreover, as* $\mathcal{S}$ *is innovation form, then* $\mathcal{S}^m$ *is innovation form by Theorem 2.2 of Chapter 2. As expected, the two systems are isomorphic, since they are both minimal and in innovation form, and the transformation matrix is*

$$T = \begin{bmatrix} -0.9934 & -0.1143 \\ -0.1143 & 0.9934 \end{bmatrix}$$

Finally, we realize that the output trajectories of both systems are indeed the same not only for the chosen switching sequence (Figure 4.5), but also for any *other switching process (Figure 4.6).*

Here, we again used the same method as in Example 4.1 to simulate $\mathcal{S}$ *and* $\mathcal{S}_m$ *both for* $\boldsymbol{\mu}$ *and* $\boldsymbol{\mu}'$.

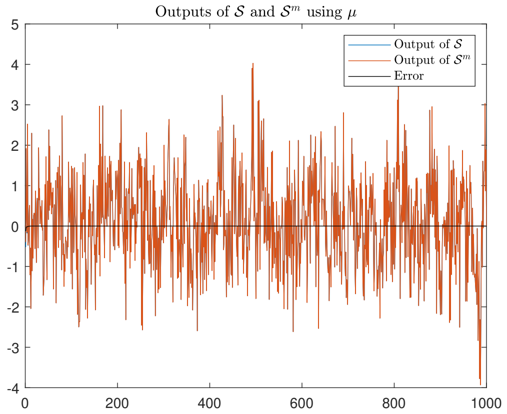

Figure 4.5: Outputs of $\mathcal{S}$ and $\mathcal{S}^m$ using μ

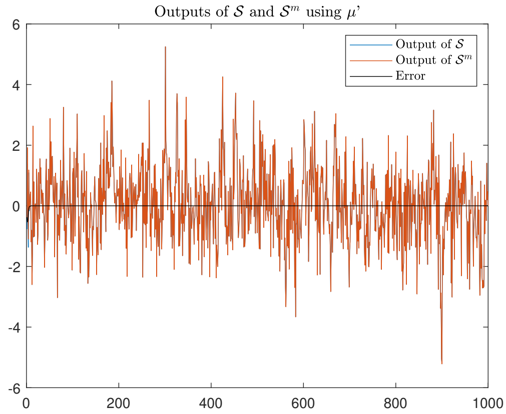

Figure 4.6: Outputs of $\mathcal{S}$ and $\mathcal{S}^m$ using μ'

4.4 Conclusion

In this chapter, we formulate conditions for a GLSS state-space representation to be minimal and in innovation form. These conditions depend only on the matrices of the GLSS representation. A minimization algorithm for transforming any GLSS representation to a minimal one in innovation form is formulated too. These results are expected to be useful for system identification, in particular, for making the identification problem mathematically well-posed. In the future, we will explore the application of the proposed results to concrete system identification algorithms.

4.5 Appendix

4.5.1 Proofs of Theorems 4.1 and 4.2

The proofs of Theorem 4.1-4.2 relies on the following series of technical results.

Lemma 4.2 **(Mean-square stability block-diagonal matrices).** *Consider $n_x \times n$ matrices $\{B_{i=1}\}_{i=1}^{n_\mu}$ and $n \times n$ matrices $\{F_i\}_{i=1}^{n_\mu}$ and $n_x \times n_x$ matrices $\{A_i\}_{i=1}^{n_\mu}$ such that that $\sum_{i=1}^{n_\mu} p_i(F_i \otimes F_i)$ and $\sum_{i=1}^{n_\mu} p_i(A_i \otimes A_i)$ are Schur. Consider the matrix*

$$\tilde{F}_i = \begin{bmatrix} F_i & B_i \\ 0 & A_i \end{bmatrix}.$$

Then $\sum_{i=1}^{n_\mu} p_i(\tilde{F}_i \otimes \tilde{F}_i)$ is Schur.

Proof of Lemma 4.2. From [92, Lemma 5] it follows that $\sum_{i=1}^{n_\mu} p_i(\tilde{F}_i \otimes \tilde{F}_i)$ being Schur is equivalent to the existence of a matrix $P > 0$ such that

$$P - \sum_{i=1}^{n_\mu} p_i \tilde{F}_i P \tilde{F}_i^T > 0.$$

We will construct such a matrix P. To this end, notice that as $\sum_{i=1}^{n_\mu} p_i(A_i \otimes A_i)$ is Schur from [92, Lemma 5] it follows that there exist $P_2 > 0$ such that

$$P_2 - \sum_{i=1}^{n_\mu} p_i A_i P_2 A_i^T > 0.$$

Define

$$S = \sum_{i=1}^{n_\mu} p_i(B_i P_2 B_i^T) + \mathscr{D}(P_2 - \sum_{i=1}^{n_\mu} p_i A_i P_2 A_i^T)^{-1} \mathscr{D}^T$$

where $\mathscr{D} = (\sum_{i=1}^{n_\mu} p_i B_i P_2 A_i^T)$. It then follows that $S \geq 0$, as $(P_2 - \sum_{i=1} n_\mu p_i A_i P_2 A_i^T)^{-1}$ is positive definite and $B_i P_2 B_i^T$ are postive semi-definite matrices for each $i = 1, \ldots, n_\mu$. From [92, Lemma 5] it follows that there exist $P_1 > 0$ such that

$$P_1 - \sum_{i=1}^{n_\mu} F_i P_i F_i^T - S > 0 \tag{4.5}$$

Consider $P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$. By standard calculation, it follows that

$$M(P) = P - \sum_{i=1}^{n_\mu} p_i \tilde{F}_i P \tilde{F}_i^T = \begin{bmatrix} P_1 - \sum_{i=1}^{n_\mu} p_i F_i P_1 F_i^T - \mathscr{D}_1 & -\mathscr{D} \\ -\mathscr{D}^T & P_2 - \sum_{i=1}^{n_\mu} p_i A_i P_2 A_i^T \end{bmatrix}$$

where $\mathscr{D}_1 = \sum_{i=1}^{n_\mu} p_i(B_i P_2 B_i^T)$. Note that $P_2 - \sum_{i=1} n_\mu p_i A_i P_2 A_i^T > 0$, hence, by using Schur complement [13], $M(P) > 0$ if (4.5) holds. The latter is true by the choice of P_1. $\square$

Lemma 4.3. *Assume that $(\mathbf{y}, \boldsymbol{\mu})$ has an asGLSS realization. Consider $n \times n_y$ matrices $\{B_{i=1}\}_{i=1}^{n_\mu}$ and $n \times n$ matrices $\{F_i\}_{i=1}^{n_\mu}$ such that that $\sum_{i=1}^{n_\mu} p_i(F_i \otimes F_i)$ is Schur. Then the infinite sum*

$$\mathbf{r}(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} F_w B_\sigma \mathbf{z}_{\sigma w}^{\mathbf{y}}(t) \tag{4.6}$$

converges in the mean-square sense and $\mathbf{r}(t)$ is a ZMWSSI process. Furthermore, $\mathbf{r}$ is the unique process $\bar{\mathbf{r}}$ which satisfies

$$\bar{\mathbf{r}}(t+1) = \sum_{i=1}^{n_\mu} (F_i \bar{\mathbf{r}}(t) + B_i \mathbf{y}(t)) \boldsymbol{\mu}_i(t) \tag{4.7}$$

where $\begin{bmatrix} \bar{\mathbf{r}}^T & \mathbf{y}^T \end{bmatrix}$ is ZMWSSI.

Proof of Lemma 4.3. Assume that $\mathcal{S} = (\{A_i, K_i\}_{i=1}^{n_\mu}, C, I, \mathbf{e})$ is a minimal asGLSS realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form of dimension n_x. Let us define the matrices

$$\tilde{F}_i = \begin{bmatrix} F_i & B_i C \\ 0 & A_i \end{bmatrix} \quad \tilde{B}_i = \begin{bmatrix} B_i \\ K_i \end{bmatrix} \quad \tilde{C} = \begin{bmatrix} I_n & 0 \end{bmatrix}$$

From Lemma 4.2 it follows that $\sum_{i=1}^{n_\mu} p_i(\tilde{F}_i \otimes \tilde{F}_i)$ is Schur, and hence by Lemma 2.2

$$\tilde{\mathbf{x}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} \tilde{F}_w B_\sigma \mathbf{z}^{\mathbf{e}}_{\sigma w}(t) \tag{4.8}$$

is convergent in the mean square sense and it is the unique state process of $(\{\tilde{F}_i, \tilde{B}_i\}_{i=1}^{n_\mu}, \tilde{C}, I, \mathbf{e})$. From [91, Lemma 2 and Lemma 9] it follows that

$$\mathbf{z}^{\mathbf{y}}_v(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} C A_w K_\sigma \mathbf{z}^{\mathbf{e}}_{\sigma w v}(t) + I \mathbf{z}^{\mathbf{e}}_v(t) \tag{4.9}$$

Notice that

$$\tilde{F}_w = \begin{bmatrix} F_w & \sum_{s_1, s_2 \in \Sigma^*, \sigma_1 \in \Sigma, w = s_1 \sigma s_2} F_{s_2} B_{\sigma_1} C A_{s_1} \\ 0 & A_w \end{bmatrix} \tag{4.10}$$

and there exists a process $\bar{\mathbf{x}}(t)$ such that $\tilde{\mathbf{x}}(t) = \begin{bmatrix} \bar{\mathbf{x}}(t) \\ \mathbf{x}(t) \end{bmatrix}$, where $\mathbf{x}$ is the unique state process of the asGLSS $\mathcal{S} = (\{A_i, K_i\}_{i=1}^{n_\mu}, C, I, \mathbf{e})$ which realizes $(\mathbf{y}, \boldsymbol{\mu})$. Indeed, let $\bar{\mathbf{x}}(t)$ be the vector formed by the first n entries of $\tilde{\mathbf{x}}(t)$ and let $\bar{\bar{\mathbf{x}}}(t)$ be formed by the last n_x rows of $\tilde{\mathbf{x}}(t)$. Then, $\tilde{\mathbf{x}}(t) = \begin{bmatrix} \bar{\mathbf{x}}(t) \\ \bar{\bar{\mathbf{x}}}(t) \end{bmatrix}$. We show that $\bar{\bar{\mathbf{x}}}(t) = \mathbf{x}(t)$, where $\mathbf{x}(t)$ is the state of the asGLSS $\mathcal{S}$. To this end, notice that

$$\begin{aligned}
\tilde{\mathbf{x}}(t+1) = \begin{bmatrix} \bar{\mathbf{x}}(t+1) \\ \bar{\bar{\mathbf{x}}}(t+1) \end{bmatrix} &= \sum_{i=1}^{n_\mu} \left(\begin{bmatrix} F_i & B_i C \\ 0 & A_i \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}(t) \\ \bar{\bar{\mathbf{x}}}(t) \end{bmatrix} + \begin{bmatrix} B_i \\ K_i \end{bmatrix} \mathbf{e}(t) \right) \boldsymbol{\mu}_i(t) \\
&= \sum_{i=1}^{n_\mu} \begin{bmatrix} F_i \bar{\mathbf{x}}(t) + B_i (C \bar{\bar{\mathbf{x}}}(t) + \mathbf{e}(t)) \\ A_i \bar{\bar{\mathbf{x}}}(t) + K_i \mathbf{e}(t) \end{bmatrix} \boldsymbol{\mu}_i(t) \\
&= \sum_{i=1}^{n_\mu} \begin{bmatrix} F_i \bar{\mathbf{x}}(t) + B_i \bar{\bar{\mathbf{y}}}(t) \\ A_i \bar{\bar{\mathbf{x}}}(t) + K_i \mathbf{e}(t) \end{bmatrix} \boldsymbol{\mu}_i(t),
\end{aligned}$$

where $\bar{\bar{\mathbf{y}}}(t) = C \bar{\bar{\mathbf{x}}}(t) + \mathbf{e}(t)$. Then

$$\bar{\bar{\mathbf{x}}}(t) = \sum_{i=1}^{n_\mu} \Big(A_i \bar{\bar{\mathbf{x}}}(t) + K_i \mathbf{e}(t) \Big) \boldsymbol{\mu}_i(t).$$

Since $\tilde{\mathbf{x}}$ is ZMWSSI, then $\bar{\bar{\mathbf{x}}}$ is also ZMWSSI. Then, by Lemma 2.2 $\bar{\bar{\mathbf{x}}}$ is the unique state process of $\mathcal{S}$. Then $\bar{\bar{\mathbf{x}}} = \mathbf{x}$ and $\bar{\bar{\mathbf{y}}} = \mathbf{y}$.
We also know that $\begin{bmatrix}\bar{\mathbf{x}}^T, & \mathbf{y}^T\end{bmatrix}^T$ is a ZMWSSI since,

$$\begin{bmatrix}\bar{\mathbf{x}}(t) \\ \mathbf{y}(t)\end{bmatrix} = \begin{bmatrix}I & 0 \\ 0 & C\end{bmatrix}\begin{bmatrix}\bar{\mathbf{x}} \\ \mathbf{x}\end{bmatrix} = \begin{bmatrix}I & 0 \\ 0 & C\end{bmatrix}\tilde{\mathbf{x}}(t)$$

and $\tilde{\mathbf{x}}(t)$ is ZMWSSI, as it is the state process of $(\{\tilde{F}_i, \tilde{B}_i)\}_{i=1}^{n_\mu}, \tilde{C}, I, \mathbf{v})$. We will show that if $\bar{\mathbf{r}}(t)$ is such that

$$\bar{\mathbf{r}}(t+1) = \sum_{i=1}^{n_\mu}\Big(F_i\bar{\mathbf{r}}(t) + B_i\mathbf{y}(t)\Big)\boldsymbol{\mu}_i(t)$$

and $\begin{bmatrix}\bar{\mathbf{r}}^T & \mathbf{y}^T\end{bmatrix}^T$ is ZMWSSI, then $\bar{\mathbf{r}} = \mathbf{r}$, and $\mathbf{r}$ is mean square convergent. This in particular, implies that $\bar{\mathbf{x}} = \mathbf{r}$. In order to to show that $\bar{\mathbf{r}} = \mathbf{r}$ and that the right-hand side of (4.6) is convergent, we proceed as follows:

$$\bar{\mathbf{r}}(t) = \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} F_w\sqrt{p_w}\mathbf{z}_w^{\bar{r}}(t) + \sum_{k=0}^{N-1}\sum_{\substack{|v| = N \\ v \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma v \in L}} \sqrt{p_{\sigma v}}F_vB_\sigma\mathbf{z}_{\sigma v}^{\mathbf{y}}(t).$$

If we define

$$\mathbf{r}_N(t) := \sum_{k=0}^{N-1}\sum_{\substack{|v| = N \\ v \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma v \in L}} \sqrt{p_{\sigma v}}F_vB_\sigma\mathbf{z}_{\sigma v}^{\mathbf{y}}(t),$$

then we get the following:

$$\bar{\mathbf{r}}(t) - \mathbf{r}_N(t) = \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} F_w\sqrt{p_w}\mathbf{z}_w^{\bar{r}}(t).$$

If we show that $\lim_{N\to\infty} E\left[\|\bar{\mathbf{r}}(t) - \mathbf{r}_N(t)\|_2^2\right] = 0$, then this means exactly that the right-hand side of (4.6) is convergent in the mean square sense, and $\bar{\mathbf{r}}(t) = \mathbf{r}(t)$. To this end, notice that

$$\begin{aligned}
E\left[\|\bar{\mathbf{r}}(t) - \mathbf{r}_N(t)\|_2^2\right] &= trace\Big(E\left[(\bar{\mathbf{r}}(t) - \mathbf{r}_N(t))(\bar{\mathbf{r}}(t) - \mathbf{r}_N(t))^T\right]\Big) \\
&= trace\Bigg(\sum_{\substack{|w| = N \\ w \in \Sigma^*}}\sum_{\substack{|v| = N \\ v \in \Sigma^*}} \sqrt{p_w}\sqrt{p_v}F_wE\left[\mathbf{z}_w^{\bar{r}}(t)(\mathbf{z}_v^{\bar{\mathbf{r}}}(t))^T\right]F_v^T\Bigg) \\
&= trace\Bigg(\sum_{\substack{|w| = N \\ w \in \Sigma^*}}\sum_{\substack{|v| = N \\ v \in \Sigma^*}} \sqrt{p_w}\sqrt{p_v}F_wT_{w,v}^{\bar{\mathbf{r}},\bar{\mathbf{r}}}F_v^T\Bigg).
\end{aligned}$$

Because $\bar{\mathbf{r}}$ is ZMWSSI, it follows that $T^{\bar{\mathbf{r}},\bar{\mathbf{r}}}_{w,v} = 0$ when $w \neq v$ and $|w| = |v|$. Then we can write:

$$E\left[\|\bar{\mathbf{r}}(t) - \mathbf{r}_N(t)\|_2^2\right] = trace\left(\sum_{\substack{|w_1| = N \\ w_1 \in \Sigma^*}} p_{w_1} F_{w_1} T^{\bar{\mathbf{r}},\bar{\mathbf{r}}}_{w_1,w_1} F^T_{w_1}\right)$$

$$= trace\left(\sum_{\substack{|w| = N-1 \\ w \in \Sigma^*}} p_w F_w \left(\sum_{\sigma \in \Sigma} F_\sigma T^{\bar{\mathbf{r}},\bar{\mathbf{r}}}_{\sigma,\sigma} F^T_\sigma\right) F^T_w\right).$$

Consider the following linear map

$$\mathcal{R} : \mathbb{R}^{n \times n} \ \ni \ V \longrightarrow \sum_{\sigma \in \Sigma} p_\sigma F_\sigma V F^T_\sigma.$$

Also consider that:

$$S = \sum_{\sigma \in \Sigma} p_\sigma F_\sigma V F^T_\sigma.$$

This leads to:

$$E\left[\|\bar{\mathbf{r}}(t) - \mathbf{r}_N(t)\|_2^2\right] = trace\,\underbrace{\mathcal{R} \circ \mathcal{R} \cdots \circ \mathcal{R}}_{N-1 \ \text{times}}(S) = \mathcal{R}^{N-1}(S).$$

From [18, Proposition 2.5], it follows that $\sum_{\sigma \in \Sigma} p_\sigma F_\sigma \otimes F_\sigma$ being Schur implies that $\lim_{k \to \infty} \mathcal{R}^k(V) = 0$ for any V, in particular, $V = S$. Hence, $\lim_{N \to \infty} trace\Big(\mathcal{R}^{N-1}(S)\Big) = 0$. Then $E\left[\|\bar{\mathbf{r}}(t) - \mathbf{r}_N(t)\|_2^2\right] \longrightarrow 0$ as $N \to \infty$.

□

Proof of Lemma 4.1. From Lemma 4.3 it follows that

$$\tilde{\mathbf{x}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t),$$

where $F_i = (A_i - K_i C)$, $i = 1, \ldots, n_\mu$ is convergent in the mean-square sense. Hence $\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t)$ converges to zero in the mean square sense. In order to see this, notice that

$$\mathbf{x}(t) = \lim_{N \to \infty} \sum_{k=0}^{N} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t),$$

where the convergence is understood in the mean square sense (in $\mathcal{H}_1$). In addition, notice that

$$\begin{aligned}
\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t) &= \lim_{N\to\infty} \left[\sum_{k=0}^{N} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) - \sum_{k=0}^{t-1} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) \right], \\
&= \lim_{N\to\infty} b_N
\end{aligned}$$

where

$$b_N = \left[\sum_{k=0}^{N} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) - \sum_{k=0}^{t-1} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) \right]$$

By convergence of subsequences of a convergent sequence, it follows that $\lim_{N\to\infty} b_N = \lim_{N\to\infty} b_{N+t-1}$ and hence

$$\begin{aligned}
\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t) &= \lim_{N\to\infty} \left[\sum_{k=0}^{N+t-1} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) - \sum_{k=0}^{t-1} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) \right] \\
&= \lim_{N\to\infty} \left[\sum_{k=t}^{N+t-1} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) \right].
\end{aligned}$$

Because the limit is convergent in the mean square sense, then

$$E\left[\|\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t)\|_2^2\right] = \lim_{N\to\infty} E\left[\left\| \sum_{k=t}^{N+t-1} \sum_{\substack{|w| = k \\ w \in \Sigma^*}} \sum_{\substack{\sigma \in \Sigma \\ \sigma w \in L}} \sqrt{p_{\sigma w}} F_w K_\sigma \mathbf{z}^{\mathbf{y}}_{\sigma w}(t) \right\|_2^2 \right].$$

However, as $\{\mathbf{z}^{\mathbf{y}}_{\sigma w}\}_{|w|\geq N}$ are jointly wide-sense stationary it follows that

$$E\left[\left\|\sum_{k=t}^{N+t-1}\sum_{\substack{|w|=k\\ w\in\Sigma^*}}\sum_{\substack{\sigma\in\Sigma\\ \sigma w\in L}}\sqrt{p_{\sigma w}}F_wK_\sigma\mathbf{z}^{\mathbf{y}}_{\sigma w}(t)\right\|_2^2\right]$$
$$=\sum_{k=t}^{N+t-1}\sum_{k_1=t}^{N+t-1}\sum_{\substack{|w|=k\\ |w_1|=k_1\\ w,w_1\in\Sigma^*}}\sum_{\substack{\sigma,\sigma_1\in\Sigma\\ \sigma w,\sigma_1 w_1\in L}}\sqrt{p_{\sigma w}}\sqrt{p_{\sigma_1 w_1}}trace\Big(F_wK_\sigma E\left[\mathbf{z}^{\mathbf{y}}_{\sigma w}(t)(\mathbf{z}^{\mathbf{y}}_{\sigma_1 w_1}(t))^T\right]K^T_{\sigma_1}F^T_{w_1}\Big)$$
$$=\sum_{k=t}^{N+t-1}\sum_{k_1=t}^{N+t-1}\sum_{\substack{|w|=k\\ |w_1|=k_1\\ w,w_1\in\Sigma^*}}\sum_{\substack{\sigma,\sigma_1\in\Sigma\\ \sigma w,\sigma_1 w_1\in L}}\sqrt{p_{\sigma w}}\sqrt{p_{\sigma_1 w_1}}trace\Big(F_wK_\sigma E\left[\mathbf{z}^{\mathbf{y}}_{\sigma w}(0)(\mathbf{z}^{\mathbf{y}}_{\sigma_1 w_1}(0))^T\right]K^T_{\sigma_1}F^T_{w_1}\Big)$$
$$=E\left[\left\|\sum_{k=t}^{N+t-1}\sum_{\substack{|w|=k\\ w\in\Sigma^*}}\sum_{\substack{\sigma\in\Sigma\\ \sigma w\in L}}\sqrt{p_{\sigma w}}F_wK_\sigma\mathbf{z}^{\mathbf{y}}_{\sigma w}(0)\right\|_2^2\right].$$

Hence

$$E\left[\|\bar{\mathbf{x}}(t)-\tilde{\mathbf{x}}(t)\|_2^2\right]=\lim_{N\to\infty}E\left[\left\|\sum_{k=t}^{N+t-1}\sum_{\substack{|w|=k\\ w\in\Sigma^*}}\sum_{\substack{\sigma\in\Sigma\\ \sigma w\in L}}\sqrt{p_{\sigma w}}F_wK_\sigma\mathbf{z}^{\mathbf{y}}_{\sigma w}(0)\right\|_2^2\right]. \tag{4.11}$$

Let us define the following series:

$$a_k=\sum_{\substack{|w|=k\\ w\in\Sigma^*}}\sum_{\substack{\sigma\in\Sigma\\ \sigma w\in L}}\sqrt{p_{\sigma w}}F_wK_\sigma\mathbf{z}^{\mathbf{y}}_{\sigma w}(0).$$

Then by Lemma 4.3 $\sum_{k=0}^{\infty}a_k$ is convergent in the mean-square sense. By the classical Cauchy property of convergent series in Hilbert-spaces, we get:

$$\lim_{t\to\infty}\sum_{k=t}^{\infty}a_k=0.$$

Hence, the norm of $\sum_{k=t}^{\infty}a_k$ in $\mathcal{H}_1$ converges to zero as $t\to\infty$. But the square of the norm of $\sum_{k=t}^{\infty}a_k$ in $\mathcal{H}_1$ is $E\,[\|\sum_{k=t}^{\infty}a_k\|_2^2]$ and hence

$$\lim_{t\to\infty}E\left[\|\sum_{k=t}^{\infty}a_k\|_2^2\right]=0.$$

From (4.11) it follows that $E\left[\|\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t)\|_2^2\right] = E\left[\| \sum_{k=t}^{\infty} a_k \|_2^2\right]$, and hence

$$\lim_{t\to\infty} E\left[\|\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t)\|_2^2\right] = \lim_{t\to\infty} E\left[\| \sum_{k=t}^{\infty} a_k \|_2^2\right] = 0.$$

This means that $\bar{\mathbf{x}}(t) - \tilde{\mathbf{x}}(t)$ converges to zero in the mean-square sense.
It remains to show that $\tilde{\mathbf{x}}(t) = \mathbf{x}(t)$. To this end, notice that $\mathbf{x}(t)$ satisfies (4.2), hence by Lemma 4.3, $\mathbf{x}(t) = \bar{\mathbf{x}}(t)$. □

Proof of Theorem 4.1. Note that we can write $\mathbf{v}(t) = \mathbf{y}(t) - C\mathbf{x}(t)$ and hence the first equation of (2.1) holds, i.e., $\mathbf{x}(t+1) = \sum_{i=1}^{n_\mu}(A_i - K_iC)\mathbf{x}(t) + K_i\mathbf{y}(t))\boldsymbol{\mu}_i(t)$. Since the matrix

$$\sum_{i=1}^{n_\mu} \boldsymbol{\mu}_i (A_i - K_iC) \otimes (A_i - K_iC) \text{ is Schur,}$$

from Lemma 4.3 it follows that

$$\mathbf{x}(t) = \sum_{w\in\Sigma^*, i\in\Sigma} \sqrt{p_{iw}} A_w K_i \mathbf{z}_{iw}^{\mathbf{y}}(t),$$

and hence the elements of $\mathbf{x}(t)$ belong to the Hilber-space generated by $\{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}$. Note that $E_l[\mathbf{v}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}] = 0$, see the proof of [91, eq. (37), proof of Theorem 4], hence, $E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t)\}_{w\in\Sigma^+}] = C\mathbf{x}(t)$ and therefore $\mathbf{e}(t) = \mathbf{v}(t)$. □

Proof of Theorem 4.2. Note that $\mathcal{S}$ is minimal if and only if the associated dLSS

$$\mathscr{S}_{\mathcal{S}} = (\{\sqrt{p_i}A_i, G_i\}_{i=1}^{n_\mu}, C, I_{n_y})$$

is minimal. In turn, the latter holds if and only if the extended $n_x - 1$-step observability and reachability matrices of the associated dLSS satisfy the following rank conditions $\operatorname{rank} \mathcal{O}_{\mathscr{S}_{\mathcal{S}}, n_x-1} = n_x$ and $\operatorname{rank} \mathcal{R}_{\mathscr{S}_{\mathcal{S}}, n_x-1} = n_x$. Also note that the rows of the extended observability matrix $\mathcal{O}_{\mathcal{D}_{\mathcal{S}}, n_x-1}$ of the dLSS $\mathcal{D}_{\mathcal{S}}$ defined in the statement of Theorem 4.2 coincide with the rows of the observability matrix $\mathcal{O}_{\mathscr{S}_{\mathcal{S}}, n_x-1}$ multiplied by a positive constant. Indeed, the set of row of $\mathcal{O}_{\mathcal{D}_{\mathcal{S}}, n_x-1}$ is the unions of the set of rows of the matrices CA_w for some $w \in \Sigma^*$, $|w| \le n_x - 1$. Likewise, the set of rows of $\mathcal{O}_{\mathscr{S}_{\mathcal{S}}, n_x-1}$ is the union of the set of rows the matrices $\sqrt{p_w}CA_w$ for some $w \in \Sigma^*$, $|w| \le n_x - 1$. Hence, the linear space spanned by the rows of these two matrices are the same and hence $\operatorname{rank} \mathcal{O}_{\mathscr{S}_{\mathcal{S}}, n_x-1} = \operatorname{rank} \mathcal{O}_{\mathcal{D}_{\mathcal{S}}, n_x-1}$. That is, $\mathcal{S}$ satisfies the observability rank condition if and only if the dLSS $\mathcal{D}_{\mathcal{S}}$ is observable. We will show that $\boldsymbol{Im}\mathcal{R}_{\mathscr{S}_{\mathcal{S}}, n_x-1} = \boldsymbol{Im}\mathcal{R}_{\mathcal{D}_{\mathcal{S}}, n_x-1}$, where $\mathcal{R}_{\mathcal{D}_{\mathcal{S}}, n-1}$ is the $n_x - 1$-step extended controllability matrix of the dLSS $\mathcal{D}_{\mathcal{S}}$. From this, it follows that $\mathcal{S}$ satisfies the reachability rank condition if and only if $\mathcal{D}_{\mathcal{S}}$ is reachable.

Now we will show that $\boldsymbol{Im}\ \mathcal{R}_{\mathscr{S}_{\mathcal{S}}, n_x-1} = \boldsymbol{Im}\mathcal{R}_{\mathcal{D}_{\mathcal{S}}, n_x-1}$. To this end, we recall that $\mathbf{x}(t)$ belongs to the linear space generated by the columns of A_wK_σ, $w \in \Sigma^*$, $\sigma \in \Sigma$, due to the fact that $\mathbf{x}(t)$ can be expressed as the infinite sum (2.10) with $B_\sigma = 0$, $\sigma \in \Sigma$. Since, by Remark 2.5,

$$G_\sigma = E[\mathbf{x}(t)(\mathbf{z}_\sigma^{\mathbf{y}}(t))^T],$$

it then follows that the columns of G_σ also belong to the linear space generated by the columns of $A_w K_\sigma$, $w \in \Sigma^*, \sigma \in \Sigma$. Therefore, the columns of $A_v G_\sigma$, $v \in \Sigma^*, \sigma \in \Sigma$ also belong to the linear space generated by the columns of $A_w K_\sigma$, $w \in \Sigma^*, \sigma \in \Sigma$. In turn, it is easy to see that latter subspace equals $\boldsymbol{Im}\mathcal{R}_{\mathcal{D}_\mathcal{S},n_x-1}$. That is, $\boldsymbol{Im}\ A_v G_\sigma$ is a subspace of $\boldsymbol{Im}\mathcal{R}_{\mathcal{D}_\mathcal{S},n_x-1}$, and therefore $\boldsymbol{Im}\ \mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1} \subseteq \boldsymbol{Im}\mathcal{R}_{\mathcal{D}_\mathcal{S},n_x-1}$. Conversely, from [91, eq. (37), proof of Theorem 4] it follows that

$$E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_{\sigma v}(t))^T] = \sqrt{p_w} A_v G_\sigma,$$

i.e., for every $w \in \Sigma^+$, the columns $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_w(t))^T]$ belong to the space generated by $A_v G_\sigma$, $\sigma \in \Sigma$, $v \in \Sigma^*$. Notice that by [81, Theorem 2 and Remark 1] applied to the dLSS $\mathscr{S}_\mathcal{S}$, the latter space equals $\boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$. Since the elements of $\mathbf{z}^{\mathbf{e}}_\sigma(t)$ are limits of finite linear combinations of the rows of $\{\mathbf{z}^{\mathbf{y}}_w(t)\}_{w\in\Sigma^+}$, it then follows that the columns of $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{e}}(t))^T]$ are the limits of finite linear combinations of columns of $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_w(t))^T]$, $w \in \Sigma^+$, and hence the columns of $E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{y}}_w(t))^T]$, $w \in \Sigma^+$ also belong to $\boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$. From [91, Proof of Theorem 4] it follows that

$$K_\sigma Q_\sigma = E[\mathbf{x}(t)(\mathbf{z}^{\mathbf{e}}_\sigma(t))^T],$$

where $Q_\sigma = E[\mathbf{e}(t)\mathbf{e}^T(t)\boldsymbol{\mu}^2_\sigma(t)]$, and hence the columns of $K_\sigma Q_\sigma$ belong to $\boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$. Since Q_σ is non-singular, it then follows that the columns of K_σ belong to $\boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$. Since by [81, Theorem 2 and Remark 1] the space $\boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$ is A_σ-invariant for all $\sigma \in \Sigma$, it then follows that $\boldsymbol{Im}A_v K_\sigma \subseteq \boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$ for all $v \in \Sigma^*$, $\sigma \in \Sigma$, and thus $\boldsymbol{Im}\mathcal{R}_{\mathcal{D}_\mathcal{S},n_x-1} \subseteq \boldsymbol{Im}\mathcal{R}_{\mathscr{S}_\mathcal{S},n_x-1}$. $\square$

4.5.2 Proof of Theorem 4.3

Before presenting the proof of Theorem 4.3, we need to define some necessary maps used in the proof.

Definition 4.2 (Necessary Maps)**.** *Define, for the sequel, the following maps:*

$$\mathscr{F} : \mathcal{D} \longmapsto \mathscr{F}(\mathcal{D})$$

$$\mathscr{D} : \mathcal{D} \longmapsto \mathscr{D}(\mathcal{D})$$

where $\mathcal{D} = (\{A_i, K_i\}_{i=1}^{n_\mu}, C, I)$ *is a dLSS and* $\mathscr{F}(\mathcal{D})$ *and* $\mathscr{D}(\mathcal{D})$ *are two transformed dLSSs with* $\mathscr{F}(\mathcal{D}) = (\{A_i - K_i C, K_i\}_{i=1}^{n_\mu}, C, I)$ *and* $\mathscr{D}(\mathcal{D}) = (\{A_i, K_i\}_{i=1}^{n_\mu}, -C, I)$*. For a dLSS* $\mathcal{D} = (\{A_i, K_i\}_{i=1}^{n_\mu}, C, I)$*, let us denote by* $\mathscr{Y}_\mathcal{D}$ *the input-output map of* $\mathcal{D}$ *induced by the initial state zero (see [86]). In other words,* $\mathscr{Y}_\mathcal{D}$ *is defined as follows, for every* $\mathrm{u} : \mathbb{Z} \longrightarrow \mathbb{R}^{n_u}$*,* $\mu : \mathbb{Z} \longrightarrow \mathbb{R}^{n_\mu}$ *with finite support,* $\mathscr{Y}_\mathcal{D}(\mathrm{u}, \mu) = \mathrm{y}$*, where* $\mathrm{y} : \mathbb{Z} \longrightarrow \mathbb{R}^{n_y}$ *is the unique function for which there exists* $\mathrm{x} : \mathbb{Z} \longrightarrow \mathbb{R}^{n_x}$ *with finite support such that* (2.2) *holds*[1].

Note that if $\mathcal{X}$ is a dLSS, then it is clear that $\mathscr{D}(\mathscr{F}(\mathscr{D}(\mathscr{F}(\mathcal{X})))) = \mathcal{X}$.

[1]Note that in [86] solutions of dLSS were defined for positive times. However, that definition is equivalent to the one described above in Subsection 2.7.1 of Chapter 2.

Lemma 4.4. *Let $\mathcal{D}$ and $\mathcal{D}'$ be two dLSSs. If $\mathcal{D}$ and $\mathcal{D}'$ are input-output equivalent, then*
1) *$\mathscr{F}(\mathcal{D})$ and $\mathscr{F}(\mathcal{D}')$ are input-output equivalent.*
2) *$\mathscr{D}(\mathcal{D})$ and $\mathscr{D}(\mathcal{D}')$ are input-output equivalent.*

Proof of Lemma 4.4. Recall the terminology and notation of Section 2.7.1. In particular, recall the notion of a Markov function $M_{\mathcal{D}}$ realized by a dLSS $\mathcal{D}$ defined in (2.13). We will show that if $M_{\mathcal{D}} = M_{\mathcal{D}'}$, then $M_{\mathscr{F}(\mathcal{D})} = M_{\mathscr{F}(\mathcal{D}')}$ and $M_{\mathscr{D}(\mathcal{D})} = M_{\mathscr{D}(\mathcal{D}')}$.
Recall from [86] the notion of a solution of a dLSS in discrete time, and consider that $(\mathrm{x}, \mathrm{u}, \mu, \mathrm{y})$ is a solution of the system $\mathcal{D}$. It then follows that

$$\begin{aligned}\mathrm{x}(t+1) &= \sum_{i=1}^{n_\mu}(A_i\mathrm{x}(t) + K_i\mathrm{u}(t))\mu_i(t)\\ \mathrm{y}(t) &= C\mathrm{x}(t).\end{aligned}$$

It can then be modified to

$$\begin{aligned}\mathrm{x}(t+1) &= \sum_{i=1}^{n_\mu}(A_i\mathrm{x}(t) + K_i(C\mathrm{x}(t) - C\mathrm{x}(t) + \mathrm{u}(t)))\mu_i(t)\\ &= \sum_{i=1}^{n_\mu}((A_i - K_iC)\mathrm{x}(t) + K_i(\mathrm{y}(t) + \mathrm{u}(t)))\mu_i(t).\end{aligned}$$

Notice that $(\mathrm{x}, \mathrm{v}, \mu, \mathrm{y})$ is a solution of the system $\mathscr{F}(\mathcal{D})$ with $\mathrm{v} = \mathrm{y} + \mathrm{u}$. It can also be shown, in the same manner, that if $(\mathrm{x}, \mathrm{v}, \mu, \mathrm{y})$ is a solution of $\mathscr{F}(\mathcal{D})$, then $(\mathrm{x}, \mathrm{u}, \mu, \mathrm{y})$ is a solution of $\mathcal{D}$, with $\mathrm{u} = \mathrm{v} - \mathrm{y}$. That is, we conclude that $\mathscr{Y}_{\mathcal{D}}(\mathrm{u}, \mu) = \mathrm{y} = \mathscr{Y}_{\mathscr{F}(\mathcal{D})}(\mathrm{v}, \mu)$. In the same manner, if $(\bar{\mathrm{x}}, \mathrm{u}, \mu, \bar{\mathrm{y}})$ is a solution of $\mathcal{D}'$ then $(\bar{\mathrm{x}}, \bar{\mathrm{v}}, \mu, \bar{\mathrm{y}})$ is a solution of $\mathscr{F}(\mathcal{D}')$, with $\bar{\mathrm{v}} = \mathrm{u} + \bar{\mathrm{y}}$. Since $\mathcal{D}$ and $\mathcal{D}'$ are both realization of the sub-Markov parameters $M_{\mathcal{D}}$, i.e., $M_{\mathcal{D}} = M_{\mathcal{D}'}$, by [86], it then follows that $\mathscr{Y}_{\mathcal{D}} = \mathscr{Y}_{\mathcal{D}'}$. However, from the discussion above, it follows that $\mathscr{Y}_{\mathcal{D}}$ and $\mathscr{Y}_{\mathcal{D}'}$ determine uniquely $\mathscr{Y}_{\mathscr{F}(\mathcal{D})}$ and $\mathscr{Y}_{\mathscr{F}(\mathcal{D}')}$ respectively. More precisely:

$$\begin{aligned}&\mathscr{Y}_{\mathscr{F}(\mathcal{D})}(\mathrm{v}, \mu) = \mathrm{y} \iff\\ &\mathrm{y} = \mathscr{Y}_{\mathcal{D}}(\mathrm{v} - \mathrm{y}, \mu) = \mathscr{Y}'_{\mathcal{D}}(\mathrm{v} - \mathrm{y}, \mu) \iff\\ &\mathrm{y} = \mathscr{Y}_{\mathscr{F}(\mathcal{D}')}(\mathrm{v}, \mu)\end{aligned}$$

That is, for all v, μ, $\mathscr{Y}_{\mathscr{F}(\mathcal{D})}(\mathrm{v}, \mu) = \mathscr{Y}_{\mathscr{F}(\mathcal{D}')}(\mathrm{v}, \mu)$, i.e., $\mathscr{Y}_{\mathscr{F}(\mathcal{D})} = \mathscr{Y}_{\mathscr{F}(\mathcal{D}')}$. In other words, from [86] it follows that $\mathscr{F}(\mathcal{D}')$ is a realization of $M_{\mathscr{F}(\mathcal{D})}$.
Finally, note that $M_{\mathscr{D}(\mathcal{D})} = -M_{\mathcal{D}}$, hence

$$M_{\mathcal{D}'} = M_{\mathcal{D}} \Longrightarrow M_{\mathscr{D}(\mathcal{D}')} = -M_{\mathcal{D}'} = -M_{\mathcal{D}} = M_{\mathscr{D}(\mathcal{D})}.$$

□

Corollary 4.2. *An dLSS Σ is a minimal realization of M_Σ, then $\mathscr{F}(\Sigma)$ is a minimal realization of $M_{\mathscr{F}(\Sigma)}$.*

Proof of corollary 4.2. Suppose that Σ is minimal and $\mathscr{F}(\Sigma)$ is not. This means that, there exists a minimal dLSS Σ^m such that $dim\Sigma^m < dim\mathscr{F}(\Sigma) = dim\Sigma$ and Σ^m is a realization of $M_{\mathscr{F}(\Sigma)}$. Define the dLSS $\hat{\Sigma}^m := \mathscr{D}(\mathscr{F}(\mathscr{D}(\Sigma^m)))$. Then, by using Lemma 4.4 and the fact that $M_{\mathscr{F}(\Sigma)} = M_{\Sigma^m}$, it follows that:

$$M_{\mathscr{D}(\mathscr{F}(\mathscr{D}(\Sigma^m)))} = M_{\mathscr{D}(\mathscr{F}(\mathscr{D}(\mathscr{F}(\Sigma))))}$$

Notice that

$$\mathscr{D}(\mathscr{F}(\mathscr{D}(\Sigma^m))) = \hat{\Sigma}^m$$
$$\mathscr{D}(\mathscr{F}(\mathscr{D}(\mathscr{F}(\Sigma)))) = \Sigma$$

Hence

$$M_{\hat{\Sigma}^m} = M_\Sigma$$

i.e., $\hat{\Sigma}^m$ is a realization of M_Σ. Notice that $dim\Sigma = dim\mathscr{F}(\Sigma) > dim\Sigma^m = dim\hat{\Sigma}^m$, which implies that Σ is not minimal, and leads to a contradiction. □

Proof of Theorem 4.3. In order to prove Theorem 4.3, two steps are needed: **(1)** proving that $\mathcal{S}_m$ is stably invertible, and **(2)** proving that $\mathcal{S}_m$ is a realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form.

Step(1): From Lemma 4.4 it follows that $\mathscr{F}(\mathcal{D})$ and $\mathscr{F}(\mathcal{D}_m)$ are both a realization of $M_{\mathcal{F}(\mathcal{D})}$, this is due to the fact that $\mathcal{D}$ and $\mathcal{D}_m$ are both a realization of $M_{\mathcal{D}}$. Recall from [92, Lemma 6] [2] that if the matrix

$$\sum_{i=1}^{n_\mu} (\sqrt{p_i}\mathscr{A}_i^T)^T \otimes (\sqrt{p_i}\mathscr{A}_i^T)^T = \sum_{i=1}^{n_\mu} p_i \mathscr{A}_i \otimes \mathscr{A}_i \text{ is Schur,}$$

then the matrix

$$\sum_{i=1}^{n_\mu} p_i \mathscr{A}_i^m \otimes \mathscr{A}_i^m \text{ is also Schur,}$$

where $\mathscr{A}_i$ are the state matrices of an dLSS Σ_1 and $\mathscr{A}_i^m$ are the state matrices of the minimal dLSS Σ_1^m which is equivalent to Σ_1. Recall that Algorithm 4 assumes that its input $\mathcal{S} = (\{A_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, I_{n_y})$ is stably invertible. It follows that the matrix

$$\sum_{i=1}^{n_\mu} p_i(A_i - K_iC) \otimes (A_i - K_iC) \text{ is Schur,}$$

[2] The termenology of [92] refers to the switch systems representations not the GLSS representations. However, the two are equivalent using the correspondence from [86, Appendix B]

where $(A_i - K_iC)$ are the state matrices of $\mathscr{F}(\mathcal{D})$. Note that $\mathscr{F}(\mathcal{D}_m)$ is minimal due to the minimality of $\mathcal{D}_m$ and Corollary 4.2 and $\mathscr{F}(\mathcal{D}_m)$ is equivalent to $\mathscr{F}(\mathcal{D})$. It follows, from [92, Lemma 6], that the matrix

$$\sum_{i=1}^{n_\mu} p_i(A_i^m - K_i^m C^m) \otimes (A_i^m - K_i^m C^m) \text{ is Schur,}$$

where $(A_i^m - K_i^m C^m)$ are the state matrices of $\mathscr{F}(\mathcal{D}_m)$. Therefore, the algorithm's output $\mathcal{S}_m$ is indeed stably invertible.

Step(2): Since $\mathcal{S}$ is a realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form, then there exists a unique ZMWSSI process $\mathbf{x}$ such that (2.1) holds with $\mathbf{v} = \mathbf{e}$. Let us apply the linear transformation T of the Kalman decomposition as described in [86, Corollary 1] to $\mathcal{S}$. Then, let

$$\hat{\mathbf{x}}(t) = T\mathbf{x}(t), \;\; \hat{A}_i = TA_i, \;\; \hat{K}_i = TK_i, \;\text{ for }\; i = 1, \ldots, n_\mu, \;\text{ and }\; \hat{C} = CT^{-1}.$$

Then the asGLSS $\hat{\mathcal{S}} = (\{\hat{A}_i, \hat{K}_i\}_{i=1}^{n_\mu}, \hat{C}, I, \mathbf{e})$ is also a realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form. Now, in order to get the Kalman decomposition, recall form [86, Corollary 1] that:

$$\hat{A}_i = \begin{bmatrix} A_i^m & 0 & A_i' \\ A_i'' & A_i''' & A_i'''' \\ 0 & 0 & A_i^{uc} \end{bmatrix}, \quad \hat{K}_i = \begin{bmatrix} K_i^m \\ K_i' \\ 0 \end{bmatrix}, \quad \hat{C} = \begin{bmatrix} C^m & 0 & C' \end{bmatrix},$$

for $i = 1, \ldots, n_\mu$, and for suitable block matrices. In particular, for all $w \in \Sigma^*$ we found:

$$\hat{A}_w = \begin{bmatrix} A_w^m & 0 & * \\ * & A_w''' & * \\ 0 & 0 & A_w^{uc} \end{bmatrix}, \tag{4.12}$$

where $*$ refers to any block matrix. Indeed, it can be shown that for $w = \epsilon$, $\hat{A}_w = I$. Now if (4.12) holds for $w = v$, then for $w = v\sigma'$:

$$\hat{A}_w = \begin{bmatrix} A_{\sigma'}^m & 0 & * \\ * & A_{\sigma'}''' & * \\ 0 & 0 & A_{\sigma'}^{uc} \end{bmatrix} \begin{bmatrix} A_v^m & 0 & * \\ * & A_v''' & * \\ 0 & 0 & A_v^{uc} \end{bmatrix} = \begin{bmatrix} A_{\sigma' v}^m & 0 & * \\ * & A_{\sigma' v}''' & * \\ 0 & 0 & A_{\sigma' v}^{uc} \end{bmatrix}.$$

Hence, by induction of then length of w, (4.12) holds. It then follows that:

$$\hat{A}_w \hat{K}_\sigma = \begin{bmatrix} A_w^m K_\sigma^m \\ * \\ 0 \end{bmatrix}.$$

Consider the following decomposition:

$$\hat{\mathbf{x}}(t) = T\mathbf{x}(t) = \begin{bmatrix} \mathbf{x}^m(t) \\ \mathbf{x}^c(t) \\ \mathbf{x}^{uc}(t) \end{bmatrix}.$$

From Lemma 2.2, $\hat{\mathbf{x}}(t)$ can be expressed as:

$$\hat{\mathbf{x}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma, \sigma w \in L \\ w \in \Sigma^*, |w| = N}} \sqrt{p_{\sigma w}} \hat{A}_w \hat{K}_\sigma \mathbf{z}^{\mathbf{e}}_{\sigma w}(t).$$

It follows that:

$$\begin{gathered}\hat{\mathbf{x}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma, \sigma w \in L \\ w \in \Sigma^*, |w| = N}} \sqrt{p_{\sigma w}} \begin{bmatrix} A^m_w K^m_\sigma \mathbf{z}^{\mathbf{e}}_{\sigma w}(t) \\ * \\ 0 \end{bmatrix} = \\ \begin{bmatrix} \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma, \sigma w \in L \\ w \in \Sigma^*, |w| = N}} \sqrt{p_{\sigma w}} A^m_w K^m_\sigma \mathbf{z}^{\mathbf{e}}_{\sigma w}(t) \\ * \\ 0 \end{bmatrix}. \end{gathered} \tag{4.13}$$

From (4.13) it follows that

$$\mathbf{x}^m(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma \in \Sigma, \sigma w \in L \\ w \in \Sigma^*, |w| = N}} \sqrt{p_{\sigma w}} A^m_w K^m_\sigma \mathbf{z}^{\mathbf{e}}_{\sigma w}(t).$$

Then from Lemma 2.2 it follows that $\mathbf{x}^m(t)$ is the state process of $\mathcal{S}_m$. Finally, notice that, as $\hat{\mathcal{S}}$ is a realization of $(\mathbf{y}, \boldsymbol{\mu})$,

$$\mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{e}(t) = \hat{C}T\mathbf{x}(t) + \mathbf{e}(t) = \hat{C}\hat{\mathbf{x}}(t) + \mathbf{e}(t).$$

The output can then be expressed as:

$$\mathbf{y}(t) = \begin{bmatrix} C^m & 0 & C' \end{bmatrix} \begin{bmatrix} \mathbf{x}^m(t) \\ \mathbf{x}^c(t) \\ \mathbf{x}^{uc}(t) \end{bmatrix} + \mathbf{e}(t). \tag{4.14}$$

From (4.13) it follows that $\mathbf{x}^{uc}(t) = 0$. Hence, from (4.14) it follows that $\mathbf{y}(t) = C^m\mathbf{x}^m(t) + \mathbf{e}(t)$. Since $\mathbf{x}^m$ is the state process of $\mathcal{S}_m$, it then follows that $\mathcal{S}_m$ is a realization of $(\mathbf{y}, \boldsymbol{\mu})$. In addition, because $\mathcal{S}_m$ is stably invertible, as proven in **(1)**, then $\mathcal{S}_m$ is a realization of $(\mathbf{y}, \boldsymbol{\mu})$ in innovation form by Theorem 4.1. $\square$

Chapter 5

Towards stochastic realization theory for GLSS models with i.i.d. inputs: minimal covariance realization algorithm and its application to subspace system identification

5.1 Introduction

In this chapter, we extend the concept of minimal (dimensional) systems in innovation form to the class of GLSS and apply it to system identification. For the sake of simplicity, we assume that $\boldsymbol{\mu}$ is i.i.d. process and $\mathbf{u}$ is a white noise process w.r.t. $\boldsymbol{\mu}$. These assumptions represent the simplest case of persistently exciting data, and there is little hope to tackle the more general case without solving this one first.

We present a necessary and sufficient condition for minimality GLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form, and we show that all such realizations of an output are isomorphic. In addition, we present a realization algorithm for computing such GLSSs from covariances of inputs and outputs.

Finally, we present a statistically consistent system identification algorithm which is based on the latter realization algorithm, and which returns, in the limit, a minimal GLSS in innovation form. In particular, if the true system is a minimal GLSS in innovation form, then in the limit this identification algorithm returns a GLSS which is isomorphic to the true GLSS, hence, it has the same deterministic behavior. We further improve this identification algorithm by combining it with Gradient-Based (GB) methods [20].

5.1.1 Related work to the chapter

Realization theory of deterministic switched systems was studied in [83], for stochastic switched systems with no inputs in [90, 91]. To the best of our knowledge, our results on realization theory of stochastic GLSS (precisely for the LSS subclass) with inputs are completely new.

Identification of switched and jump-Markov systems is an active research area, see the [62, 33, 8, 16, 4, 6], and the references therein. However, most of the literature assumes that the switching signal is unobserved, which makes the problem more challenging but also leads to lack of consistency results for state-space representations. A consistent subspace identification algorithms was presented in [82] for noiseless LSSs. In [102] a consistent identification algorithm for noisy LSSs was presented, but the noise gain matrix and the noise covariance were not estimated, and the identification was based on several i.i.d. time series.

Moreover, there is a wealth of literature on system identification for LPV systems, including subspace methods, [113, 21, 72, 126, 125] and the references therein. Stochastic realization theory of LPV systems was investigated in [72, 19]. In [19] the existence of LPV systems in innovation form was investigated, but the noise gain matrices of the obtained system depended on the current and past scheduling. Moreover, [19] did not address minimality and uniqueness of systems in innovation form. Unlike the identification algorithm of this chapter, the existing subspace methods, see the overview [21], are not proven to be consistent or to return a minimal system in innovation form.

A statistically consistent identification algorithm was presented in [74] for noisy LPV systems with no inputs, and in [72] for noisy LPV systems with inputs. This chapter is an extension of [72], the main difference w.r.t. [72] is as follows:

(1) The conditions on the scheduling signals from [72] do not directly apply for i.i.d. switching signals (see Remark 5.1 at the end of the introduction).

(2) We characterize minimality and uniqueness (up to isomorphism) of systems in innovation form.

(3) The proposed identification algorithm returns a minimal system in innovation form, while the one of [72] returns a non-minimal one; that is, the system returned by the algorithm from [72] need not have the deterministic behavior as the true system.

(4) We present novel modifications to improve the empirical behavior of the identification algorithm.

To sum up, the identification algorithm of this chapter is the first algorithm for noisy GLSSs that has all of the following properties: **(a)** it estimates the noise gain and the noise covariance, **(b)** returns a minimal GLSSs in innovation form, and **(c)** applies to one single long time series.

This chapter uses the technical results from Chapter 3 on the decomposition of outputs of GLSSs into stochastic and deterministic components. The latter is used in Chapter 3 to show existence of a realization in innovation form and to provide *sufficient* conditions for minimality and uniqueness of GLSSs in innovation form. In contrast to Chapter 3, we formulate *necessary and sufficient* conditions for minimality, and we show isomorphism of a much wider class of GLSSs in innovation forms, see Remark 5.3 for a detailed explanation of the gap between the two results of the chapters.

5.1.2 Outline of the chapter

This chapter is structured as follows. Section 5.3 shows the association between a dLSS and its correspondent sGLSS in order to present the *necessary* and *sufficient* conditions mentioned above. In Section 5.4, the selection Ho-Kalman algorithm is recalled from the literature in the aim of presenting the new covariance realization algorithm. Finally, in Section 5.5, the identification algorithm, which is based on the previous covariance algorithm, is introduced. In addition, the gradient search algorithm is added as a last step of this new identification algorithm. The proofs of the theorems and Lemma presented in this chapter are discussed in the Appendixes.

5.2 Technical assumptions on GLSS

In this chapter we consider stochastic processes $\mathbf{y}, \mathbf{u}, \boldsymbol{\mu}$ such that the following holds.

Assumption 5.1. *The tuple* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *has a realization by a sGLSSs of the form* (2.1)

In particular, Assumption 5.1 tacitly implies that $\boldsymbol{\mu}$ is an admissible switching process. For the sake of simplicity, in this chapter we will further restrict the class of switching sequences by considering i.i.d switching processes, i.e., we impose the following additional restriction on $\boldsymbol{\mu}$.

Assumption 5.2 (i.i.d. switching)**.** *The switching process* $\boldsymbol{\mu}$ *is admissible and is of the type as described in Example 2.8, Chapter 2. That is,* $\boldsymbol{\mu}$ *arises as follows. Let* $\boldsymbol{\theta}$ *be an i.i.d process with values in* $\Sigma = \{1, \ldots, n_\mu\}$*. Then* $\boldsymbol{\mu}_\sigma(t) = \chi(\boldsymbol{\theta}(t) = \sigma)$ *for all* $\sigma \in \Sigma$*,* $t \in \mathbb{Z}$*, where* χ *is the indicator function, i.e.,*

$$\chi(\boldsymbol{\theta}(t) = \sigma) = \begin{cases} 1, & \textit{if } \boldsymbol{\theta}(t) = \sigma \\ 0, & \textit{otherwise.} \end{cases}$$

Moreover, we set $\mathcal{E} = \Sigma \times \Sigma$ *and thus* $L = \Sigma^+$ *holds. Furthermore, the constants* $\{p_\sigma\}_{\sigma\in\Sigma}$ *are defined as* $p_\sigma = P(\boldsymbol{\theta}(t) = \sigma)$*,* $\alpha_\sigma = 1$ *for all* $\sigma \in \Sigma$*. Finally, with* $\alpha_\sigma = 1$*,* $\sigma \in \Sigma$ *it holds* $\sum_{\sigma\in\Sigma} \alpha_\sigma \boldsymbol{\mu}_\sigma(t) = 1$.

This assumption imposes restrictions on the data used for system identification, but not necessarily for the model class which will be identified depending on the switching $\boldsymbol{\mu}$ as discussed in the preliminary chapter. It can be thought of as a persistence of excitation condition.

The assumption that $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ has a sGLSS realization implicitly implies that $\mathbf{u}$ is a white noise w.r.t. $\boldsymbol{\mu}$ and that $\mathbf{y}, \mathbf{u}$ are jointly ZMWSSI. In addition, in order to avoid cumbersome expressions in the statement of the main results and their proofs, we will assume that $T^{\mathbf{u},\mathbf{u}}_{\sigma,\sigma} = Q_{\mathbf{u}} > 0$ does not depend on $\sigma \in \Sigma$. This is the case, for example, if $\mathbf{u}$ is an i.i.d. process with $Q_{\mathbf{u}} = E[\mathbf{u}(t)\mathbf{u}^T(t)]$, such that the σ-algebra generated by $\{\mathbf{u}(s)\}_{s\in\mathbb{Z}}$ is independent of the σ-algebra generated by $\{\boldsymbol{\mu}(s)\}_{s\in\mathbb{Z}}$. For the convenience of the reader we gather these facts and assumptions in the following assumption.

Assumption 5.3 (Inputs and outputs)**.** ***(1)*** $\mathbf{u}$ *is a white noise w.r.t.* $\boldsymbol{\mu}$ *and there exists a positive definite matrix* $Q_{\mathbf{u}} > 0$ *such that for all* $\sigma \in \Sigma$, $T^{\mathbf{u},\mathbf{u}}_{\sigma,\sigma} = Q_{\mathbf{u}}$. ***(2)*** *and the process* $\begin{bmatrix}\mathbf{y}^T, & \mathbf{u}^T\end{bmatrix}^T$ *is a ZMWSSI.*

Remark 5.1 (Relationship with [72])**.** *In this chapter, if we view the switching* $\boldsymbol{\mu}$ *of the GLSS we are working with as a scheduling signal, then it does not satisfy the assumptions of [72]: while it is an i.i.d. process , it is not zero mean,* $\boldsymbol{\mu}_i(t), \boldsymbol{\mu}_j(t)$ *are not uncorrelated for* $i, j \in \Sigma$*, and* $\boldsymbol{\mu}_1(t) \neq 1$*. By applying a suitable normalizing affine transformation to* $\boldsymbol{\mu}$ *we can bring it to a form which satisfies [72]; for example, for* $n_\mu = 2$*, we can take*

$$\tilde{\boldsymbol{\mu}}(t) = (\boldsymbol{\mu}_1(t) + \boldsymbol{\mu}_2(t), \boldsymbol{\mu}_2(t) - p_2)^T = (1, \boldsymbol{\mu}_2(t) - p_2)^T,$$

where p_2 *is defined in Definition 2.2. The system matrices of the thus arising LPV system are affine combinations of the system matrices of original GLSS. However, this transformation does not appear to simplify the derivation of the results. To the contrary, it would require handling the affine transformation between system matrices when analyzing minimality.*

5.3 Minimality of sGLSS in innovation form

5.3.1 Relationship between sGLSSs and dLSSs

The idea behind relating sGLSSs and dLSSs is to express the covariances of outputs of sGLSSs as Markov parameters of dLSSs. To this end, we need to recall from Chapter 3 the decomposition of the output $\mathbf{y}$ into deterministic and stochastic parts, and the notion of orthogonal projection from Chapter 2. We recall that, intuitively, the orthogonal projection $E_l[\mathbf{z} \mid \mathfrak{S}]$ is *minimal variance linear prediction* of $\mathbf{z}$ based on the elements of $\mathfrak{S}$.

We also recall from Chapter 3, Section 3.3, (3.1)-(3.2) that the *deterministic component* $\mathbf{y}^d$ and the *stochastic component* $\mathbf{y}^s$ of $\mathbf{y}$ are defined as

$$\begin{aligned} \mathbf{y}^d(t) &= E_l[\mathbf{y}(t) \mid \{\mathbf{z}^{\mathbf{u}}_w(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}], \\ \mathbf{y}^s(t) &= \mathbf{y}(t) - \mathbf{y}^d(t). \end{aligned} \tag{5.1}$$

Recall from Section 3.3 that intuitively, $\mathbf{y}^d$ is the best linear prediction of $\mathbf{y}$ based on the present and past values of $\mathbf{u}$ multiplied by indicator functions of past discrete modes. In fact, by Theorem 3.1 of Chapter 3, $\mathbf{y}^d$ (resp. $\mathbf{y}^s$) is the output of the asGLSS obtained from (2.1), by setting $K_\sigma = 0$ (resp. $B_\sigma = 0$), for all $\sigma \in \Sigma$.

Next, we define the following Markov functions (see Chapter 2, Section 2.7.1 for the definition of a Markov function) values of which are formed by suitable covariances of $\mathbf{y}^s, \mathbf{y}^d$ and $\mathbf{u}$.

Definition 5.1 (Markov function)**.** *Define the Markov function* $M_{\mathbf{y},\mathbf{u}} : \Sigma^* \to \mathbb{R}^{n_y\times(n_y+n_u)}$ *by*

$$M_{\mathbf{y},\mathbf{u}}(w) = \begin{cases} \begin{bmatrix}\Lambda^{\mathbf{y}^d,\mathbf{u}}_w Q^{-1}_{\mathbf{u}}, & \Lambda^{\mathbf{y}^s,\mathbf{y}^s}_w\end{bmatrix} & w \neq \epsilon, \\ \begin{bmatrix}\Lambda^{\mathbf{y}^d,\mathbf{u}}_\epsilon Q^{-1}_{\mathbf{u}}, & I_{n_y}\end{bmatrix} & w = \epsilon. \end{cases} \tag{5.2}$$

Definition 5.2 (Auxiliary covariances)**.** *Define the Markov function $\Psi_{\mathbf{u},\mathbf{y}}$ such that $\Psi_{\mathbf{u},\mathbf{y}}(w)$ is formed by the first n_u columns of $M_{\mathbf{y},\mathbf{u}}(w)$, i.e. $\Psi_{\mathbf{u},\mathbf{y}}(w) = \Lambda_w^{\mathbf{y}^d,\mathbf{u}} Q_{\mathbf{u}}^{-1}$, and define the Markov function $\Psi_{\mathbf{y}^s}$ such that $\Psi_{\mathbf{y}^s}(w)$ is formed by the first n_y columns of $M_{\mathbf{y},\mathbf{u}}(w)$, i.e. $\Psi_{\mathbf{y}^s}(w) = \Lambda_w^{\mathbf{y}^s,\mathbf{y}^s}$.*

The intuition behind the Markov function definition is as follows. It can be shown that $\Psi_{\mathbf{u},\mathbf{y}}(w)$ can be expressed as Markov parameters of a suitably defined dLSS. Since, by Theorem 3.1 and Theorem 3.2 of Chapter 3, $\mathbf{y}^s$ is the output of an asGLSS, and hence of a GBS in terminology of [91], then by [91, Lemma 4] $\Psi_{\mathbf{y}^s}$ can be shown to be the Markov function of a suitable dLSS too. Subsequently, by combining the dLSS realizations of $\Psi_{\mathbf{u},\mathbf{y}}$ and $\Psi_{\mathbf{y}^s}$, it follows that $M_{\mathbf{y},\mathbf{u}}$ has a dLSS realization.

Formally, consider the process $\mathbf{x}^s(t) = \mathbf{x}(t) - E_l[\mathbf{x}(t) \mid \{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}]$ from Theorem 3.1, Chapter 3, and consider the covariances $T_{\sigma,\sigma}^{\mathbf{x}^s,\mathbf{x}^s} = E[\mathbf{z}_\sigma^{\mathbf{x}^s}(t)(\mathbf{z}_\sigma^{\mathbf{x}^s}(t))^T]$. Define the matrices

$$G_\sigma = \sqrt{p_\sigma}(A_\sigma T_{\sigma,\sigma}^{\mathbf{x}^s,\mathbf{x}^s} C^T + K_\sigma T_{\sigma,\sigma}^{\mathbf{v},\mathbf{v}} F^T) \tag{5.3}$$

Then define the *dLSS associated with sGLSS* as

$$\mathscr{S}_{\mathcal{S}} = (\{\sqrt{p_\sigma}A_\sigma, \begin{bmatrix}\sqrt{p_\sigma}B_\sigma & G_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, C, \begin{bmatrix}D & I_{n_y}\end{bmatrix}), \tag{5.4}$$

Note that, by the proof of Theorem 3.1, presented in Section 3.3 of Chapter 3, $\mathbf{x}^s(t)$ is the state of the asGLSS $\mathcal{S}_s$ defined in (3.6), which is obtained from $\mathcal{S}$ by considering $B_\sigma = 0$, and hence $\mathbf{x}^s$ is a ZMWSSI process and $T_{\sigma,\sigma}^{\mathbf{x}^s,\mathbf{x}^s}$ is well-defined.

Lemma 5.1. *If $\mathcal{S}$ is a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, then the associated dLSS $\mathscr{S}_{\mathcal{S}}$ is a realization of $M_{\mathbf{y},\mathbf{u}}$.*

The proof follows from [91, Lemma 4] and Theorem 3.2 of Chapter 3. For the convenience of the reader, the proof is represented in Appendix 5.8.1

Remark 5.2 (Computing G_σ)**.** *In order to compute G_σ, we can use the fact that $\mathbf{x}^s$ is the state of the asGLSS obtained from* (2.1) *by setting $B_\sigma = 0$, $\sigma \in \Sigma$ (see the proof of Theorem 3.1 in Section 3.5.1 of Chapter 3). Hence, by [91, Lemma 5], the covariance of $\mathbf{x}^s$ satisfies*

$$p_\sigma T_{\sigma,\sigma}^{\mathbf{x}^s,\mathbf{x}^s} = \lim_{N\to\infty} P_\sigma^N,$$

where $P_\sigma^0 = 0$ and

$$P_\sigma^{N+1} = p_\sigma \sum_{\sigma_1\in\Sigma} (A_{\sigma_1} P_{\sigma_1}^N A_{\sigma_1}^T + p_{\sigma_1} K_{\sigma_1} T_{\sigma_1,\sigma_1}^{\mathbf{v},\mathbf{v}} K_{\sigma_1}^T).$$

Conversely, we can associate with any dLSS realization of the Markov function $M_{\mathbf{y},\mathbf{u}}$ a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. In order to define the latter, we need to specify its noise process, which happens to be the *innovation noise* $\mathbf{e}^s$ of $\mathbf{y}^s$ as defined in [91, eq. (16)]. By Theorem 3.2 of Chapter 3,

$$\begin{aligned} \mathbf{e}^s(t) &= \mathbf{y}(t) - \hat{\mathbf{y}}(t) \\ \hat{\mathbf{y}}(t) &= E_l[\mathbf{y}(t) \mid \{\mathbf{z}_w^{\mathbf{y}}(t), \mathbf{z}_w^{\mathbf{u}}(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}]. \end{aligned} \tag{5.5}$$

That is, $\mathbf{e}^s(t)$ is the prediction error of the best linear predictor $\hat{\mathbf{y}}(t)$ of $\mathbf{y}(t)$ given the products of past outputs, inputs and discrete modes, i.e., $\mathbf{e}^s$ is direct extension of the classical innovation noise.

Definition 5.3 (Full rank property of a sGLSS output process)**.** *The process* $\mathbf{y}$ *is said to be* full rank *w.r.t.* $\mathbf{u}$ *and* $\boldsymbol{\mu}$*, if for all* $\sigma \in \Sigma$*, the covariance* $T^{\mathbf{e}^s,\mathbf{e}^s}_{\sigma,\sigma}$ *is invertible.*

Note that the difference between the definition above and Definition 2.10 is that, in Chapter 2, the output was assumed to be an output of an asGLSS. However, in what follows, we will only mention that $\mathbf{y}$ is full rank when it is clear from the context that it is w.r.t. $\mathbf{u}$ and $\boldsymbol{\mu}$ or only w.r.t. $\boldsymbol{\mu}$. If $\mathbf{y}$ is full rank, then, for an observable dLSS

$$\mathscr{S} = (\{\hat{A}_\sigma, \begin{bmatrix}\hat{B}_\sigma & \hat{G}_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, \hat{C}, \begin{bmatrix}\hat{D} & I_{n_y}\end{bmatrix}) \tag{5.6}$$

for which $\sum_{\sigma=1}^{n_\mu} \hat{A}_\sigma \otimes \hat{A}_\sigma$ is a Schur matrix, define the *sGLSS* $\mathcal{S}_{\mathscr{S}}$ *associated with* $\mathscr{S}$ *as*

$$\mathcal{S}_{\mathscr{S}} = (\{\frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma, \frac{1}{\sqrt{p_\sigma}}\hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, I_{n_y}, \mathbf{e}^s), \tag{5.7}$$

where for all $\sigma \in \Sigma$,

$$\hat{K}_\sigma = \lim_{\mathcal{I}\to\infty} \hat{K}^{\mathcal{I}}_\sigma, \tag{5.8}$$

and the matrices $\{\hat{K}^{\mathcal{I}}_\sigma\}_{\sigma\in\Sigma,\mathcal{I}\in\mathbb{N}}$ satisfy the recursion

$$\begin{aligned}
\hat{P}^{\mathcal{I}+1}_\sigma &= p_\sigma \sum_{\sigma_1\in\Sigma} \frac{1}{p_{\sigma_1}} \left(\hat{A}_{\sigma_1}\hat{P}^{\mathcal{I}}_{\sigma_1}\hat{A}^T_{\sigma_1} + \hat{K}^{\mathcal{I}}_{\sigma_1}\hat{Q}^{\mathcal{I}}_{\sigma_1}(\hat{K}^{\mathcal{I}}_{\sigma_1})^T\right)\\
\hat{Q}^{\mathcal{I}}_\sigma &= p_\sigma T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma} - \hat{C}\hat{P}^{\mathcal{I}}_\sigma(\hat{C})^T\\
\hat{K}^{\mathcal{I}}_\sigma &= \left(\hat{G}_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma\hat{P}^{\mathcal{I}}_\sigma(\hat{C})^T\right)\left(\hat{Q}^{\mathcal{I}}_\sigma\right)^{-1}
\end{aligned} \tag{5.9}$$

with $\hat{P}^0_\sigma = 0$. We will argue below that the limit (5.8) exists and that the matrices $\hat{Q}^{\mathcal{I}}_\sigma, \sigma \in \Sigma$, $\mathcal{I} \in \mathbb{N}$ are invertible, and hence the recursion (5.9) is well-posed. Note that the noise process of $\mathcal{S}_{\mathscr{S}}$ is the innovation noise, and we will argue below that its noise and state covariances satisfy

$$p_\sigma T^{\mathbf{e}^s,\mathbf{e}^s}_{\sigma,\sigma} = \lim_{\bar{\mathcal{I}}\to\infty} \hat{Q}^{\bar{\mathcal{I}}}_\sigma, \quad p_\sigma T^{\hat{\mathbf{x}}^s,\hat{\mathbf{x}}^s}_{\sigma,\sigma} = \lim_{\bar{\mathcal{I}}\to\infty} \hat{P}^{\bar{\mathcal{I}}}_\sigma, \tag{5.10}$$

where

$$\hat{\mathbf{x}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{\sigma\in\Sigma,\ w\in\Sigma^* \\ \sigma w\in L,\ |w|=N}} \sqrt{p_\sigma}\hat{A}_w\hat{K}_\sigma\mathbf{z}^{\mathbf{e}^s}_{\sigma w}(t) + \frac{\sqrt{p_\sigma}}{\sqrt{p_w}}\hat{A}_w\hat{B}_\sigma\mathbf{z}^{\mathbf{u}}_{\sigma w}(t)$$

is the unique state process of $\mathcal{S}_{\mathscr{S}}$, and

$$\hat{\mathbf{x}}^s(t) = \hat{\mathbf{x}}(t) - E_l[\hat{\mathbf{x}}(t) \mid \{\mathbf{z}^{\mathbf{u}}_w(t)\}_{w\in\Sigma^+} \cup \{\mathbf{u}(t)\}]$$

and $\hat{\mathbf{x}^s}$ is the state of the asGLSS stochastic component $(\mathcal{S}_{\mathscr{S}})_s$ of $\mathcal{S}_{\mathscr{S}}$ defined as in (3.6). The well-posedness of (5.9) (non-singularity of the matrices $\hat{Q}^{\mathcal{I}}_\sigma$) convergence of (5.8), (5.9) and (5.10) follow from the proof of [91, Theorem 3] and it is shown in the proof of Lemma 5.2 below.

Lemma 5.2. *If* $\mathbf{y}$ *is full rank w.r.t.* $\boldsymbol{\mu}$*, and* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *has a realization by sGLSS, and* $\mathscr{S}$ *is a minimal dLSS realization of* $M_{\mathbf{y},\mathbf{u}}$*, then the recursion* (5.9) *is well-posed, i.e., the matrices* $\{\hat{Q}_{\sigma}^{\mathcal{I}}\}_{\sigma\in\Sigma,\mathcal{I}\in\mathbb{N}}$ *are invertible, the limits* (5.8) *exist, and the associated sGLSS* $\mathcal{S}_{\mathscr{S}}$ *is a stationary, and it is a sGLSS realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$*, and* (5.9) *and* (5.10) *hold.*

In the rest of the chapter, we will refer to $\mathcal{S}_{\mathscr{S}}$ as the *sGLSS associated with the dLSS* $\mathscr{S}$. The proof of Lemma 5.2 is presented in Appendix 5.8.2, it relies on Theorem 3.2 of Chapter 3 and [91, Theorem 3].

5.3.2 Main result on minimal sGLSSs in innovation form

Recall from Chapter 3, Definition 3.1 that a sGLSS of the form (2.1) is said to be in *innovation form*, if its noise process $\mathbf{v}$ equals $\mathbf{e}^s$ from (5.5) and $F = I_{n_y}$. Note that the sGLSS associated with a dLSS is in innovation form. Recall from The sGLSS (2.1) in innovation form gives rise to the predictor (3.12), i.e.,

$$\begin{aligned}\mathbf{x}(t+1) &= \sum_{i=1}^{n_\mu}\Big((A_i - K_iC)\mathbf{x}(t) + (B_i - K_iD)\mathbf{u}(t) + K_i\mathbf{y}(t)\Big)\boldsymbol{\mu}_i(t)\\ \hat{\mathbf{y}}(t) &= C\mathbf{x}(t) + D\mathbf{u}(t),\end{aligned} \tag{5.11}$$

which generates the best linear prediction $\hat{\mathbf{y}}(t)$ of the output $\mathbf{y}(t)$, while being driven by past outputs and inputs. Indeed, since $F = I_{n_y}$ from (2.1) it follows that

$$\mathbf{v}(t) = \mathbf{y}(t) - C\mathbf{x}(t) - D\mathbf{u}(t)$$

and if $\mathbf{v}(t) = \mathbf{e}^s(t)$, then by substituting the expression above into (2.1) the equation (5.11) follows.

Recall from Definition 2.8 that we say that a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ is *minimal*, if it has the smallest dimension among all sGLSS realizations of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. In order to study minimality, we need the concepts of reachability and observability. We call a sGLSS $\mathcal{S}$ *reachable* (resp. *observable*), if the associated dLSS $\mathscr{S}_{\mathcal{S}}$ is reachable, (resp. observable) according to the definition of [81, Definition 19], applied to zero initial state.

In order to investigate uniqueness of minimal sGLSS in innovation form, we need to recall the concept of isomorphism from Chapter 3, (3.14). Let $\mathcal{S}$ and $\hat{\mathcal{S}}$ be sGLSSs realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form such that $\mathcal{S}$ is of the form (2.1) and $\hat{\mathcal{S}} = (\{\hat{A}_\sigma, \hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, I_{n_y}, \mathbf{e}^s)$. We say that $\mathcal{S}$ and $\hat{S}$ are *isomorphic*, if there exists a nonsingular matrix T such that

$$\begin{aligned}&\hat{A}_\sigma = TA_\sigma T^{-1}, \hat{K}_\sigma = TK_\sigma,\ \hat{B}_\sigma = TB_\sigma, \quad \sigma\in\Sigma\\ &\hat{C} = CT^{-1}, D = \hat{D}\end{aligned} \tag{5.12}$$

Theorem 5.1 (Main result: minimal innovation form)**.** *Assume that* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *has a sGLSS realization and that* $\mathbf{y}$ *is full rank w.r.t.* $\mathbf{u}$ *and* $\boldsymbol{\mu}$*. Then the following holds.*

1. *A sGLSS realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *is minimal if and only if it is reachable and observable.*

2. *If a dLSS* $\mathscr{S}_m$ *of the form* (5.6) *is a minimal realization of* $M_{\mathbf{y},\mathbf{u}}$*, then the associated sGLSS* $\mathcal{S}_{\mathscr{S}_m}$ *is a minimal sGLSS realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *in innovation form.*

3. *There exists a minimal sGLSS realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *in innovation form.*

4. *Any two minimal sGLSS of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *in innovation form are isomorphic.*

The proof is presented in Appendix 5.8.3, it relies on the fact that any dLSS can be transformed to a minimal one, that minimal dLSSs with the same input-output map are isomorphic and on Lemma 5.1-5.2.

Note that observability and reachability of the associated dLSS $\mathscr{S}_{\mathcal{S}}$ of the form 5.4 can be checked using simple rank conditions of suitable matrices [81]. That is, our conditions for minimality can be checked numerically as follows:

1. Compute the the i-step extended reachability matrices $\mathcal{R}_{\mathscr{S}_{\mathcal{S}},i}$ and the i-step observability matrices $\mathcal{O}_{\mathscr{S}_{\mathcal{S}},i}$ of $\mathscr{S}_{\mathcal{S}}$:

$$\begin{aligned} \mathcal{R}_{\mathscr{S}_{\mathcal{S}},i} &= \begin{bmatrix} \mathcal{R}_{\mathscr{S}_{\mathcal{S}},0} & \sqrt{p_1}A_1\mathcal{R}_{\mathscr{S}_{\mathcal{S}},i-1} & \dots & \sqrt{p_{n_\mu}}A_{n_\mu}\mathcal{R}_{\mathscr{S}_{\mathcal{S}},i-1} \end{bmatrix}, \\ \mathcal{O}_{\mathscr{S}_{\mathcal{S}},i} &= \begin{bmatrix} \mathcal{O}^T_{\mathscr{S}_{\mathcal{S}},0} & \sqrt{p_1}A_1^T\mathcal{O}^T_{\mathscr{S}_{\mathcal{S}},i-1} & \dots & \sqrt{p_{n_\mu}}A_{n_\mu}^T\mathcal{O}^T_{\mathscr{S}_{\mathcal{S}},i-1} \end{bmatrix}^T, \end{aligned} \tag{5.13}$$

where $\mathcal{R}_{\mathscr{S}_{\mathcal{S}},0} = \begin{bmatrix} \begin{bmatrix}\sqrt{p_1}B_1 & G_1\end{bmatrix} & \begin{bmatrix}\sqrt{p_2}B_2 & G_2\end{bmatrix} & \dots & \begin{bmatrix}\sqrt{p_{n_\mu}}B_{n_\mu} & G_{n_\mu}\end{bmatrix} \end{bmatrix}$, $\mathcal{O}_{\mathscr{S}_{\mathcal{S}},0} = C^T$, and for all $i > 1$ and for all $\sigma \in \Sigma$:

$$G_\sigma = \sqrt{p_\sigma}(A_\sigma T^{\mathbf{x}^s,\mathbf{x}^s}_{\sigma,\sigma}C^T + K_\sigma T^{\mathbf{v},\mathbf{v}}_{\sigma,\sigma}F^T).$$

2. Verify if $\mathrm{rank}\{\mathcal{R}_{\mathscr{S},n_x-1}\} = n_x$, and $\mathrm{rank}\{\mathcal{O}_{\mathscr{S},n_x-1}\} = n_x$.

Moreover, when dealing with non-minimal sGLSS $\mathcal{S}$, the latter can be transformed effectively to a minimal one in innovation form as follows: first we compute the associated dLSS $\mathscr{S}_{\mathcal{S}}$, we minimize it using Algorithm 2 (see Chapter 2) to obtain a minimal dLSS $\mathscr{S}_m$ realization of $M_{\mathbf{y},\mathbf{u}}$, and then we compute the sGLSS $\mathcal{S}_{\mathscr{S}_m}$ associated with the latter minimal dLSS.

Remark 5.3 (Improvement over Chapter 3)**.** *For a sGLSS* $\mathcal{S}$ *of the form* (2.1)*, we now refine the minimality condition presented in Chapter 3 (Corollary 3.2).*
In Chapter 3 the minimality condition was stated in terms of the reachability and observability of a dLSS that captures only the stochastic component $\mathcal{S}_s$ *of* $\mathcal{S}$*. This system, denoted by* (2.19)*, differs from the associated dLSS defined in equation* (5.4)*.*
To clarify, the dLSS in (5.4) *is given by:*

$$\mathscr{S}_{\mathcal{S}} = (\{\sqrt{p_\sigma}A_\sigma, \begin{bmatrix}\sqrt{p_\sigma}B_\sigma & G_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, C, \begin{bmatrix}D & I_{n_y}\end{bmatrix}).$$

The dLSS associated to the stochastic-only component $\mathcal{S}_s$ from (2.19) *is:*

$$\mathscr{S}_{\mathcal{S}_s} = (\{\sqrt{p_\sigma}A_\sigma, G_\sigma\}_{\sigma=1}^{n_\mu}, C, I_{n_y}),$$

and it is obtained from (5.4) *by deleting the columns of $\sqrt{p_\sigma}B_\sigma$. The minimality of latter is sufficient, but not necessary for the reachability and observability of the dLSS associated with $\mathcal{S}$.*
Likewise, in Chapter 3 we showed isomorphism between sGLSSs in innovation form which satisfy the above sufficient condition for minimality, and for which the image of B_σ belongs to the image of K_σ. In contrast, Theorem 5.1 establishes isomorphism for a much wider class of sGLSSs in innovation form. This marks a significant generalization and improvement over the minimality results of Chapter 3.

5.4 Minimial covariance realization algorithm

In this section we present a Ho-Kalman-like algorithm for computing a minimal sGLSS realization innovation form from $M_{\mathbf{y},\mathbf{u}}$. To this end, we first recall the Ho-Kalman realization algorithm for dLSS.

5.4.1 Reduced basis Ho-Kalman algorithm for dLSS

Below we recall from [19] an adaptation of the Ho-Kalman-like algorithm for dLSS from [81]. Define a $(n, n_\mathrm{y}, n_\mathrm{u})$-*selection (selection for short)* as a pair of word-index sets (α, β) such that,

$$\alpha = \{(u_i, k_i)\}_{i=1}^n,\ \beta = \{(\sigma_j, v_j, l_j)\}_{j=1}^n, \tag{5.14}$$

for some $u_i \in \Sigma^+$, $|u_i| \leq n$, $k_i \in \{1, 2, \cdots, n_\mathrm{y}\}$, $\sigma_j \in \Sigma$, $v_j \in \Sigma^+$, $|v_j| \leq n$, $l_j \in \{1, 2, \cdots, n_\mathrm{u}\}$, $i, j \in \{1, \ldots, n\}$.

Intuitively, the components of α, β which belong to $\Sigma^+ \cup \Sigma$ determine the choice of Markov parameters, while the other components determine a choice of row and column indices of the chosen Markov parameters, see [72] for an example.

Formally, let $M : \Sigma^* \to \mathbb{R}^{n_\mathrm{y} \times n_\mathrm{u}}$ be a Markov function. Define the Hankel matrix $\mathcal{H}^M_{\alpha,\beta} \in \mathbb{R}^{n \times n}$ and the matrices $\mathcal{H}^M_{\sigma,\alpha,\beta} \in \mathbb{R}^{n \times n}$, $\mathcal{H}^M_{\alpha,\sigma} \in \mathbb{R}^{n \times n_\mathrm{u}}$ and $\mathcal{H}^M_{\beta} \in \mathbb{R}^{n_\mathrm{y} \times n}$ as follows as follows:

$$\left[\mathcal{H}^M_{\alpha,\beta}\right]_{i,j} = \left[M(\sigma_j v_j u_i)\right]_{k_i,l_j}, \quad i, j = 1, \ldots, n \tag{5.15}$$

$$\left[\mathcal{H}^M_{\sigma,\alpha,\beta}\right]_{i,j} = \left[M(\sigma_j v_j \sigma u_i)\right]_{k_i,l_j},\ i, j = 1, \ldots, n \tag{5.16}$$

$$\left[\mathcal{H}^M_{\alpha,\sigma}\right]_{i,j} = \left[M(\sigma u_i)\right]_{k_i,j},\ j = 1, \ldots, n_\mathrm{u}, i = 1, \ldots, n \tag{5.17}$$

$$\left[\mathcal{H}^M_{\beta}\right]_{i,j} = \left[M(\sigma_j v_j)\right]_{i,l_j},\ i = 1, \ldots, n_\mathrm{y}, j = 1, \ldots n \tag{5.18}$$

where $(u_i, k_i) \in \alpha, (\sigma_j, v_j, l_j) \in \beta$ are as in the ordering in (5.14). Finally, let $\mathcal{L}_{\alpha,\beta}$ be the set of all words $w \in \Sigma^*$ such that $M(w)$ occurs in one of the matrices, (5.15) – (5.16), i.e.,

$$\mathcal{L}_{\alpha,\beta} = \{u_i, \sigma_j v_j, \sigma_j v_j u_i, \sigma_j v_j \sigma u_i, \sigma u_i \mid \sigma \in \Sigma\}_{i,j=1}^n$$

The promised algorithm is presented in Algorithm 5, which is an extension of the linear Ho-Kalman [56].

Algorithm 5: Ho-Kalman for dLSSs,
Input: $(n, n_{\mathrm{y}}, n_{\mathrm{u}})$-selection (α, β), $\{M(w)\}_{w \in \mathcal{L}_{\alpha,\beta}}$, $M(\epsilon)$.
Output: dLSS $(\{\hat{A}_\sigma, \hat{B}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, M(\epsilon))$

1: Construct the matrices $\mathcal{H}^M_{\alpha,\beta}$, $\mathcal{H}^M_{\sigma,\alpha,\beta}$ $\mathcal{H}^M_{\alpha,\sigma}$ and $\mathcal{H}^M_{\beta}$, and set

$$\hat{A}_\sigma = (\mathcal{H}^M_{\alpha,\beta})^{-1}\mathcal{H}^M_{\sigma,\alpha,\beta},\ \hat{B}_\sigma = (\mathcal{H}^M_{\alpha,\beta})^{-1}\mathcal{H}^M_{\alpha,\sigma},\ \hat{C} = \mathcal{H}^M_{\beta}.$$

Lemma 5.3 ([19])**.** *Assume that there exists a minimal dLSS realization of M of dimension n. Then there exists an $(n, n_{\mathrm{y}}, n_{\mathrm{u}})$-selection (α, β) such that rank$(\mathcal{H}^M_{\alpha,\beta}) = n$. For any $(n, n_{\mathrm{y}}, n_{\mathrm{u}})$-selection (α, β) for which rank$(\mathcal{H}^M_{\alpha,\beta}) = n$, Algorithm 5 returns a minimal dLSS realization of M.*

5.4.2 Covariance realization algorithm

Theorem 5.1, together with Algorithm 5 and Lemma 5.3 suggest that in order to compute a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, one could apply Algorithm 5 to Hankel-matrices corresponding to $M_{\mathbf{y},\mathbf{u}}$. However, the definition of $M_{\mathbf{y},\mathbf{u}}$ uses the covariances of the processes $\mathbf{y}^d$ and $\mathbf{y}^s$, which are not directly available. Below we show how to compute $M_{\mathbf{y},\mathbf{u}}$ using only the covariances of $\mathbf{y}$ and $\mathbf{u}$. To this end, note that by Lemma 5.1 the dLSS $(\{\sqrt{p_\sigma}A_\sigma, \sqrt{p_\sigma}B_\sigma\}_{\sigma=1}^{n_\mu}, C, D)$ is a realization of $\Psi_{\mathbf{u},\mathbf{y}}$.

Lemma 5.4. *If there exists a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, then*

$$\begin{aligned} &\Lambda^{\mathbf{y}^d,\mathbf{u}}_w = \Lambda^{\mathbf{y},\mathbf{u}}_w, \quad \Lambda^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma w} = \Lambda^{\mathbf{y},\mathbf{y}}_{\sigma w} - \Lambda^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma w} \\ &T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma} = T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma} - T^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma,\sigma} \end{aligned} \tag{5.19}$$

for all $w \in \Sigma^*$, $\sigma \in \Sigma$. Moreover, if $(\{\tilde{A}^d_\sigma, \tilde{B}^d_\sigma\}_{\sigma=1}^{n_\mu}, \tilde{C}^d, \tilde{D}^d)$ is a minimal dLSS realization of $\Psi_{\mathbf{u},\mathbf{y}}$, then

$$\begin{aligned} \Lambda^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma w} &= \frac{1}{p_\sigma}\tilde{C}^d\tilde{A}^d_w(\tilde{A}^d_\sigma\tilde{P}_\sigma(\tilde{C}^d)^T + p_\sigma\tilde{B}^d_\sigma Q_{\mathbf{u}}(\tilde{D}^d)^T) \\ T^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma,\sigma} &= \frac{1}{p_\sigma}(\tilde{C}^d\tilde{P}_\sigma(\tilde{C}^d)^T + p_\sigma\tilde{D}^d Q_{\mathbf{u}}(\tilde{D}^d)^T) \end{aligned} \tag{5.20}$$

where

$$\tilde{P}_\sigma = \lim_{\mathcal{J}\to\infty} \tilde{P}^{\mathcal{J}}_\sigma$$

$$\tilde{P}^0_\sigma = 0,$$

and

$$\tilde{P}_{\sigma}^{\mathcal{J}+1} = p_{\sigma} \sum_{\sigma_1 \in \Sigma} \left(\frac{1}{p_{\sigma_1}} \tilde{A}_{\sigma_1}^d \tilde{P}_{\sigma_1}^{\mathcal{J}} (\tilde{A}_{\sigma_1}^d)^T + \tilde{B}_{\sigma_1}^d Q_{\mathbf{u}} (\tilde{B}_{\sigma_1}^d)^T \right).$$

In particular, if we define the matrices $\Lambda_{\sigma w}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}$, $T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}$, $\Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}}$, $T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}}$ $w \in \Sigma^*$, $\sigma \in \Sigma$ *such that*

$$\begin{aligned}
\Lambda_{\sigma w}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}} &= \Lambda_{\sigma w}^{\mathbf{y},\mathbf{y}} - \Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}}, \quad T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}} = T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}} - T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}} \\
\Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}} &= \frac{1}{p_{\sigma}} \tilde{C}^d \tilde{A}_w^d (\tilde{A}_{\sigma}^d \tilde{P}_{\sigma}^{\mathcal{J}} (\tilde{C}^d)^T + p_{\sigma} \tilde{B}_{\sigma}^d Q_{\mathbf{u}} (\tilde{D}^d)^T), \\
T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}} &= \frac{1}{p_{\sigma}} (\tilde{C}^d \tilde{P}_{\sigma}^{\mathcal{J}} (\tilde{C}^d)^T + p_{\sigma} \tilde{D}^d Q_{\mathbf{u}} (\tilde{D}^d)^T)
\end{aligned} \tag{5.21}$$

then

$$\lim_{\mathcal{J} \to \infty} \Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}} = \Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d}, \quad \lim_{\mathcal{J} \to \infty} T_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}} = T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d} \tag{5.22}$$

and hence

$$\lim_{\mathcal{J} \to \infty} \Lambda_{\sigma w}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}} = \Lambda_{\sigma w}^{\mathbf{y}^s,\mathbf{y}^s}, \quad \lim_{\mathcal{J} \to \infty} T_{\sigma w}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}} = T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s} \tag{5.23}$$

The proof of Lemma 5.4 is presented in Appendix 5.4. The first part of the lemma relies on orthogonality of $\mathbf{y}^s(t)$ and $\{\mathbf{z}_w^{\mathbf{u}}(t)\}_{w \in \Sigma^*}$ established in Chapter 3. The second part follows by identifying dLSS realizations of $\Psi_{\mathbf{u},\mathbf{y}}$ with asGLSS realizations of $\mathbf{y}^d$ whose noise is $\mathbf{u}$, and using the equations in [91, Lemma 4], or alternatively, Section 2.7.3, for the covariances of $\mathbf{y}^d$.
We can then use Lemma 5.4 to compute the necessary values of $M_{\mathbf{y},\mathbf{u}}$ as follows. First, we compute a minimal dLSS realization of $\Psi_{\mathbf{u},\mathbf{y}}$ using Algorithm 5, and then use (5.19) – (5.20) to compute those values of $\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s}$, and hence of $M_{\mathbf{y},\mathbf{u}}$, which are necessary for applying Algorithm 5 to $M_{\mathbf{y},\mathbf{u}}$. This idea is formalized in Algorithm 6. For Algorithm 6 to be well-posed, Algorithm 5 should be applicable in Step 2 and Step 5. In turn, for the latter

$$\operatorname{rank}(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}}) = n_x, \quad \operatorname{rank}(\mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{u},\mathbf{y}}}) = \bar{n}.$$

should hold. Moreover, in order to apply (5.9) in Step 6, each of the matrices $\{{}^{\mathcal{J}}\hat{Q}_{\sigma}^{\mathcal{I}}\}_{\mathcal{I}=0}^{\bar{\mathcal{I}}}$ should be invertible for all $\sigma \in \Sigma$. We show that for large enough $\mathcal{J}$ Algorithm 6 is well-posed and that as $\bar{\mathcal{I}}, \mathcal{J}$ tend to ∞, Algorithm 6 returns a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form.

Corollary 5.1. *Assume that* $\mathbf{y}$ *is full rank and there exists a minimal sGLSS realization of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *of dimension* n_x, *and a minimal dLSS realization of* $\Psi_{\mathbf{y},\mathbf{u}}$ *of dimension* $\bar{n}$. *Assume that the selections* (α, β) *and* $(\bar{\alpha}, \bar{\beta})$ *satisfy*

$$\begin{aligned}
&\operatorname{rank}(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u}}}) = n_x, \\
&\operatorname{rank}(\mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{u},\mathbf{y}}}) = \bar{n}.
\end{aligned} \tag{5.25}$$

Then there exists an integer $\mathcal{J}_\star > 0$ *such that for all* $\mathcal{J} > \mathcal{J}_\star$,

$$\operatorname{rank}(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}}) = n_x, \quad \textit{and} \quad \forall \mathcal{I} = 0, \dots, \bar{\mathcal{I}}, \forall \sigma \in \Sigma : \ {}^{\mathcal{J}}\hat{Q}_{\sigma}^{\mathcal{I}} \textit{ is invertible} \tag{5.26}$$

Algorithm 6: Minimal covariance realization algorithm,
Input: $(n_x, n_y, n_u + n_y)$-selection (α, β); $(\bar{n}, n_y, n_u)$-selection $(\bar{\alpha}, \bar{\beta})$; covariances $\{\Lambda_w^{\mathbf{y},\mathbf{u}}\}_{w \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}}}$, $\{\Lambda_w^{\mathbf{y},\mathbf{y}}\}_{w \in \mathcal{L}_{\alpha,\beta}}$, $\Lambda_\epsilon^{\mathbf{y},\mathbf{u}}$, $\{T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}}\}_{\sigma \in \Sigma}$; integers $\bar{\mathcal{I}}, \mathcal{J} > 0$.
Output: Matrices $\{\frac{{}^{\mathcal{J}}\hat{A}_\sigma}{\sqrt{p_\sigma}}, \frac{{}^{\mathcal{J}}\hat{B}_\sigma}{\sqrt{p_\sigma}}, {}^{\mathcal{J}}\hat{K}_\sigma^{\bar{\mathcal{I}}}, {}^{\mathcal{J}}\hat{Q}_\sigma^{\bar{\mathcal{I}}}\}_{\sigma=1}^{n_\mu}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D}$.

1: Use (5.19) to compute $\{\Psi_{\mathbf{u},\mathbf{y}}(w)\}_{w \in \mathcal{L}_{\bar{\alpha},\bar{\beta}}}$ and $\Psi_{\mathbf{u},\mathbf{y}}(\epsilon)$.
2: Run Algorithm 5 with $M = \Psi_{\mathbf{u},\mathbf{y}}$ and selection $(\bar{\alpha}, \bar{\beta})$ and denote the result by $\mathscr{S}_d = (\{\tilde{A}_i^d, \tilde{B}_i^d\}_{i=1}^{n_\mu}, \tilde{C}^d, \tilde{D}^d)$.
3: Compute $\{\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}, \Lambda_w^{\mathbf{y}^d,\mathbf{u}}\}_{w \in \mathcal{L}_{\alpha,\beta}}$, using (5.21), (5.19) and $\{\Lambda_w^{\mathbf{y},\mathbf{u}}, \Lambda_w^{\mathbf{y},\mathbf{y}}\}_{w \in \mathcal{L}_{\alpha,\beta}}$.
4: Define the Markov function $M_{\mathbf{y},\mathbf{u},\mathcal{J}}$ in the same way as $M_{\mathbf{y},\mathbf{u}}$ but with $\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s}$ replaced by $\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}$, i.e.

$$M_{\mathbf{y},\mathbf{u},\mathcal{J}}(w) = \begin{cases} \begin{bmatrix} \Lambda_w^{\mathbf{y}^d,\mathbf{u}} Q_{\mathbf{u}}^{-1}, & \Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}} \end{bmatrix} & w \neq \epsilon, \\ \begin{bmatrix} \Lambda_\epsilon^{\mathbf{y}^d,\mathbf{u}} Q_{\mathbf{u}}^{-1}, & I_{n_y} \end{bmatrix} & w = \epsilon. \end{cases} \tag{5.24}$$

Compute $M_{\mathbf{y},\mathbf{u},\mathcal{J}}(w)$ from $\{\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}, \Lambda_w^{\mathbf{y}^d,\mathbf{u}}\}$ for $w \in \mathcal{L}_{\alpha,\beta}$.
5: Run Algorithm 5 with $M = M_{\mathbf{y},\mathbf{u},\mathcal{J}}$ and selection (α, β), and denote by $\mathscr{S} = (\{{}^{\mathcal{J}}\hat{A}_i, \begin{bmatrix} {}^{\mathcal{J}}\hat{B}_i, & {}^{\mathcal{J}}\hat{G}_i \end{bmatrix}\}_{i=1}^{n_\mu}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D})$ the result.
6: From $\{T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}}\}_{\sigma \in \Sigma}$ compute $\{T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}\}_{\sigma \in \Sigma}$ using (5.21) and the dLSS $\mathscr{S}_d$ from Step 2. Define the matrices ${}^{\mathcal{J}}\hat{K}_\sigma^{\mathcal{I}} := \hat{K}_\sigma^{\mathcal{I}}$, ${}^{\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}} := \hat{Q}_\sigma^{\mathcal{I}}$, where $\hat{K}_\sigma^{\mathcal{I}}$ and $\hat{Q}_\sigma^{\mathcal{I}}$ are as in (5.9) for $\mathcal{I} = 0, \ldots, \bar{\mathcal{I}}$, with $\hat{A}_\sigma, \hat{G}_\sigma, \hat{C}, \hat{D}, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s}$ replaced by ${}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D}, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}$.

and hence Algorithm 5 in Step 2 and Step 5 can be applied and (5.9) *in Step 6 is well-posed, and Algorithm 6 returns matrices such that the following limits exist*

$$\lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{A}_\sigma = \hat{A}_\sigma, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{B}_\sigma = \hat{B}_\sigma \textit{ for } \sigma \in \Sigma, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{C} = \hat{C}, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{D} = \hat{D} \tag{5.27}$$

$$\lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{K}_\sigma^{\mathcal{I}} = \hat{\hat{K}}_\sigma^{\mathcal{I}}, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}} = \hat{\hat{Q}}_\sigma^{\mathcal{I}}, \quad \textit{for } \sigma \in \Sigma, \ \mathcal{I} = 0, 1, \ldots, \bar{\mathcal{I}} \tag{5.28}$$

$$\hat{K}_\sigma = \lim_{\bar{\mathcal{I}}\to\infty} \hat{\hat{K}}_\sigma^{\bar{\mathcal{I}}} = \lim_{\bar{\mathcal{I}}\to\infty} \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{K}_\sigma^{\bar{\mathcal{I}}}, \quad \hat{\hat{Q}}_\sigma = \lim_{\bar{\mathcal{I}}\to\infty} \hat{\hat{Q}}_\sigma^{\bar{\mathcal{I}}} = \lim_{\bar{\mathcal{I}}\to\infty} \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{Q}_\sigma^{\bar{\mathcal{I}}}, \ \textit{for } \ \sigma \in \Sigma. \tag{5.29}$$

and the tuple

$$\tilde{S} = \left(\left\{ \frac{\hat{A}_\sigma}{\sqrt{p_\sigma}}, \frac{\hat{B}_\sigma}{\sqrt{p_\sigma}}, \hat{K}_\sigma \right\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, \mathbf{e}^s \right) \tag{5.30}$$

defines a sGLSS which is a minimal realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ *in innovation form, and*

$$T_{\sigma,\sigma}^{\boldsymbol{e}^s,\boldsymbol{e}^s} p_\sigma = \hat{Q}_\sigma \tag{5.31}$$

for all $\sigma \in \Sigma$. *Moreover, there exist an* $(n_x, n_y, n_u + n_y)$*-selection* (α, β) *and* $(\bar{n}, n_y, n_u)$*-selection* $(\bar{\alpha}, \bar{\beta})$ *which satisfy* (5.25).

The proof of Corollary 5.1 is presented in Appendix 5.8.5, it uses Lemma 5.3, Theorem 5.1 and the notion of sGLSSs associated with dLSSs.
That is, Algorithm 6 returns a minimal sGLSS in innovation form based on the input and output covariances, if the selections (α, β), $(\bar{\alpha}, \bar{\beta})$ give Hankel-matrices of a correct rank, and there always exist such selections.
Since minimal sGLSS realizations of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form are isomorphic, it follows that Algorithm 6 finds a sGLSS which has the same deterministic behavior as the true system, i.e. the same output response as the true system for any inputs, noise and switching signals, if the true system is assumed to be minimal in innovation form.

5.5 Identification algorithm

We formulate an identification algorithm based on Algorithm 6. The formulated algorithm uses empirical covariances instead of the true ones $\Lambda_w^{\mathbf{y},\mathbf{u}}$, $\Lambda_w^{\mathbf{y},\mathbf{y}}$, $T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}}$ and applies to them Algorithm 6. To this end, we make the following assumptions.

Assumption 5.4. *(1) The $(n_x, n_y, n_u + n_y)$-selection (α, β) and the $(\bar{n}, n_y, n_u)$-selection $(\bar{\alpha}, \bar{\beta})$ satisfy* (5.25)*, $\mathbf{y}$ is full rank, and n_x and $\bar{n}$ satisfy the hypothesis of Corollary 5.1.*

(2) The process $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ is ergodic, and the observed sample paths $y : \mathbb{Z} \to \mathbb{R}^{n_y}$, $u : \mathbb{Z} \to \mathbb{R}^{n_u}$ and $q : \mathbb{Z} \to \Sigma$ of $\mathbf{y}$, $\mathbf{u}$ and $\boldsymbol{\mu}$ respectively satisfy the following: For all $w, v \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}} \cup \Sigma \cup \{\epsilon\}$ define the empirical covariances

$$T_{v,w,N}^{\mathbf{y},\mathbf{b}} = \frac{\sum_{t=N_0}^{N} z_v^y(t)(z_w^b(t))^T}{N - N_0}, \quad where \;\; b = \begin{cases} y & if \;\; \mathbf{b} = \mathbf{y} \\ u & if \;\; \mathbf{b} = \mathbf{u} \end{cases}, \tag{5.32}$$

and for all $\mu_\sigma(t) = \chi(q(t) = \sigma)$, $\sigma \in \Sigma$, and for all $w = \sigma_1\sigma_2 \ldots \sigma_r \in \Sigma^+$, $r > 0$, $\sigma_1, \ldots, \sigma_r \in \Sigma$, $b \in \{y, u\}$

$$\mu_w(t) = \mu_{\sigma_1}(t - k + 1)\mu_{\sigma_2}(t - k + 2) \cdots \mu_{\sigma_r}(t)$$
$$z_w^b(t) = b(t - r)\mu_w(t - 1)\frac{1}{\sqrt{p_w}}, \quad z_\epsilon^b(t) = b(t),$$

and N_0 is an upper bound on the length of words in $\mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}} \cup \Sigma$. Then we assume that for all $\mathbf{b} \in \{\mathbf{y}, \mathbf{u}\}$,

$$\Lambda_w^{\mathbf{y},\mathbf{u}} = \lim_{N\to\infty} T_{\epsilon,w,N}^{\mathbf{y},\mathbf{u}}, \quad T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}} = \lim_{N\to\infty} T_{\sigma,\sigma,N}^{\mathbf{y},\mathbf{y}} \tag{5.33}$$

The first assumption is not restrictive, such selections always exist, and they can be found by an exhaustive search through all the possible selections. The assumption (5.33) says that the observed sample (y, u, q) satisfies the law of large numbers; ergodicity, which is a standard assumption in identification, implies that almost any sample satisfies (5.33). In order to be able to apply Algorithm

Algorithm 7: sGLSS identification algorithm,
Input: Data $\{y(t), u(t), q(t)\}_{t=1}^{N}$, $(n_x, n_y, n_u + n_y)$-selection (α, β) and $(\bar{n}, n_y, n_u)$-selection $(\bar{\alpha}, \bar{\beta})$; integers $\bar{\mathcal{I}}, \mathcal{J} > 0$.
Output: The matrices $\{\tilde{A}_{\sigma}^{N,\mathcal{J}}, \tilde{B}_{\sigma}^{N,\mathcal{J}}, \tilde{K}_{\sigma}^{N,\bar{\mathcal{I}},\mathcal{J}}, \tilde{Q}_{\sigma}^{N,\bar{\mathcal{I}},\mathcal{J}}\}_{\sigma=1}^{n_{\mu}}, \tilde{C}^{N,\mathcal{J}} \tilde{D}^{N,\mathcal{J}}$ returned by Algorithm 6.

1: Run Algorithm 6 with the covariances $\Lambda_{w}^{\mathbf{y},\mathbf{y}}$, $\Lambda_{w}^{\mathbf{y},\mathbf{u}}$, $\Lambda_{\epsilon}^{\mathbf{y},\mathbf{u}}$ and $T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}}$ being replaced by the empirical covariances $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{y}}$, $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{u}}$, $T_{\epsilon,\epsilon,N}^{\mathbf{y},\mathbf{u}}$ and $T_{\sigma,\sigma,N}^{\mathbf{y},\mathbf{y}}$ respectively for $w \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}}$, $\sigma \in \Sigma$ and with the integers $\bar{\mathcal{I}}, \mathcal{J}$.

6, the latter should be well-posed for the data at hand, i.e., the second equation of (5.25) and both equations in (5.26) should hold, if the covariances used to construct $\Psi_{\mathbf{y},\mathbf{u}}$ and $M_{\mathbf{y},\mathbf{u},\mathcal{J}}$ are replaced by the empirical covariances described in Algorithm 7. If this is the case, the we say that Algorithm 7 is *well-posed* for the given choice of $(\mathcal{J}, N)$. In Lemma 5.5 we will show that for large enough N and $\mathcal{J}$, Algorithm 7 is well-posed, and as $N, \mathcal{J}, \bar{\mathcal{I}}$ tend to ∞, the matrices of Algorithm 7 will converge to the matrices of a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form. Moreover, in the proof of Lemma 5.5 we will spell out formally the conditions under which Algorithm 7 is well-posed.

Lemma 5.5 (Consistency)**.** *Under Assumption 5.4, there exist an integer $\mathcal{J}_{\star} > 0$ such that for any $\mathcal{J} > \mathcal{J}_{\star}$ there exists an integer $N_{\star}(\mathcal{J})$ such that for all $N > N_{\star}(\mathcal{J})$, Algorithm 7 is well-posed for this choice of $(\mathcal{J}, N)$, and the matrices returned by Algorithm 7 satisfy the following:*

$$\tilde{K}_{\sigma} = \lim_{\bar{\mathcal{I}}\to\infty} \lim_{\mathcal{J}\to\infty} \lim_{N\to\infty} \tilde{K}_{\sigma}^{N,\bar{\mathcal{I}},\mathcal{J}}, \;\; p_{\sigma} T_{\sigma,\sigma}^{\mathbf{e}^{s},\mathbf{e}^{s}} = \lim_{\bar{\mathcal{I}}\to\infty} \lim_{\mathcal{J}\to\infty} \lim_{N\to\infty} \tilde{Q}_{\sigma}^{N,\bar{\mathcal{I}},\mathcal{J}}$$
$$[\tilde{A}_{\sigma}, \tilde{B}_{\sigma}] = \lim_{\mathcal{J}\to\infty} \lim_{N\to\infty} [\tilde{A}_{\sigma}^{N,\mathcal{J}}, \tilde{B}_{\sigma}^{N,\mathcal{J}}], \quad [\tilde{C}, \tilde{D}] = \lim_{\mathcal{J}\to\infty} \lim_{N\to\infty} [\tilde{C}^{N,\mathcal{J}}, \tilde{D}^{N,\mathcal{J}}]$$

and $\mathcal{S} = (\{\tilde{A}_{\sigma}, \tilde{B}_{\sigma}, \tilde{K}_{\sigma}, \}_{\sigma=1}^{n_{\mu}}, \tilde{C}, \tilde{D}, \mathbf{e}^{s})$ is a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form.

That is, by Lemma 5.5, Algorithm 7 is a statistically consistent. Moreover, if the true system is a minimal sGLSS in innovation form, which, by Theorem 5.1 we can assume w.l.g, then the sGLSS returned by Algorithm 7 will be isomorphic to the true system, as $N \to \infty$. In particular, in the limit, the identified system has *the same deterministic behavior* as the true system. That is, while the data used for identification had to be sampled from certain distributions, the identified model recreates the behavior of the true one for any switching signal and input.

Remark 5.4 (Computing empirical covariances)**.** *Algorithm 7 can be improved by using* (5.32) *directly for computing empirical covariances $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{b}}$, $\mathbf{b} \in \{\mathbf{y}, \mathbf{u}\}$ results in slow convergence. Instead, we propose to compute the empirical covariances by solving a linear regression problem. More precisely, assume that $\mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}} = \{w_1, \ldots, w_k\}$ and $\mathcal{L}_{\alpha,\beta} = \{w_1, \ldots, w_r\}$ for some*

$k, r > 0$. *Let us then define the following matrices*

$$R = \begin{bmatrix} y(N_0) \\ \vdots \\ y(N) \end{bmatrix}, \quad \Phi_{\mathbf{b}} = \begin{bmatrix} z^{\mathbf{b}}_{w_1}(N_0) & \cdots & z^{\mathbf{b}}_{w_{l(\mathbf{b})}}(N_0) \\ \vdots & & \vdots \\ z^{\mathbf{b}}_{w_1}(N) & \cdots & z^{\mathbf{b}}_{w_{l(\mathbf{b})}}(N) \end{bmatrix}$$

where for $\mathbf{b} = \mathbf{y}$, $l(\mathbf{b}) = r$ *and* $b = y$, *and for* $\mathbf{b} = \mathbf{u}$, $l(\mathbf{b}) = k$ *and* $b = u$. *By the well-known formula,*

$$\hat{\Theta}^{\mathbf{b}} = (\Phi_{\mathbf{b}}^T \Phi_{\mathbf{b}})^{-1} \Phi_{\mathbf{b}}^T R$$

is the least squares solution of the equation $R = \Phi_{\mathbf{b}} \hat{\Theta}^{\mathbf{b}}$. *Hence,*

$$\frac{1}{N - N_0} (\Phi_{\mathbf{b}}^T \Phi_{\mathbf{b}}) \hat{\Theta}^{\mathbf{b}} = \frac{1}{N - N_0} \Phi_{\mathbf{b}}^T R,$$

and the latter contains all covariances $T^{\mathbf{y},\mathbf{b}}_{\epsilon,w,N}$, $\mathbf{b} \in \{\mathbf{u}, \mathbf{y}\}$ *required for Algorithm 7. In turn,* $\hat{\Theta}^{\mathbf{b}}$ *can be computed using standard linear regression tools.*

Building on this approach, once the sGLSS parameters (the model matrices) are estimated using Algorithm 7, on can consider them as initial estimation and further refine them through a GB optimization procedure. This leads to the following refinement step:

Algorithm 8: Refined sGLSS Estimation via Gradient Descent,
Input: Dataset $\mathcal{D} = \{(u(t), y(t), q(t))\}_{t=1}^{T}$.
Output: Return the refined parameters $\Theta^\star$ obtained after convergence of the gradient descent.

1: **Initial Estimation**
Run Algorithm 7 to obtain initial sGLSS parameters
$\Theta^{(0)} = \left[\{\tilde{A}_\sigma^{(0)}, \tilde{B}_\sigma^{(0)}, \tilde{K}_\sigma^{(0)}, \tilde{Q}_\sigma^{(0)}\}_{\sigma=1}^{n_\mu}, \tilde{C}^{(0)}, \tilde{D}^{(0)} \right]$
2: **Gradient-Based Refinement**
Apply the Gradient-Based algorithm of [20, Algorithm 7.1] using $\Theta^{(0)}$ as initialization.
Note: Unlike [20], where zero matrices are used to initialize the noise gain matrices K_σ, here we use the non-zero estimates $K_\sigma^{(0)}$ from Algorithm 7

5.6 Numerical example

In this section, we illustrate Algorithm 7 on a numerical example. All computations are carried out on an i7 2.11-GHz Intel core processor with 32GB of RAM running MATLAB R2023b. For data generation we use the following randomly generatated sGLSS (2.1) with $\Sigma = \{1, 2\}$

$$\begin{aligned}
A_1 &= \begin{bmatrix} 0.1039 & 0.0255 & 0.5598 \\ 0.4338 & 0.0067 & 0.0078 \\ 0.3435 & 0.0412 & 0.0776 \end{bmatrix}, B_1 = \begin{bmatrix} 1.6143 \\ 5.9383 \\ 7.3671 \end{bmatrix}, \\
A_2 &= \begin{bmatrix} 0.1834 & 0.2456 & 0.0511 \\ 0.0572 & 0.2445 & 0.0642 \\ 0.1395 & 0.6413 & 0.5598 \end{bmatrix}, B_2 = \begin{bmatrix} 6.0624 \\ 4.9800 \\ 3.1372 \end{bmatrix}, \\
K_1 &= [0.4942\ 0.2827\ 0.8098]^T, \\
K_2 &= [0.6215\ 0.1561\ 0.7780]^T \\
C &= \begin{bmatrix} 0.1144 & 0.7623 & 0.0020 \end{bmatrix}, D = 1, F = 1.
\end{aligned}$$

Training and validation data of length N and N_{val}, respectively, are generated using the data generator sGLSS with white-noise input process $\mathbf{u}$ with uniform distribution $\mathcal{U}(-1, 1)$, an i.i.d. process $\boldsymbol{\theta}$ sucht that $\mathcal{P}(\boldsymbol{\theta}(t) = q) = 0.5$, $q = 1, 2$, and a Gaussian white noise $\mathbf{v}$ with variance $\sigma_{\mathbf{v}}^2$. The corresponding *Signal-to-Noise Ratio* SNR are shown in Table 5.1. We run the version of Algorithm 7 using the Least-Square method from Remark 5.4, with $(\alpha, \beta) = (\bar{\alpha}, \bar{\beta})$,

$$\alpha = \{(11, 1), (1, 1), (\epsilon, 1)\}, \ \ \beta = \{(2, \epsilon, 1), (1, 2, 1), (1, 1, 1)\},$$

and the GB refinement step from Algorithm 8. For the latter we used the LPVcore toolbox [24].

For validation, the true output minus the measurement noise $Fv(t)$, denoted by $y(t)$, was compared with the output $\hat{y}(t)$ predicted by the estimated model in innovation form using past inputs and outputs, as explained in Section 5.3.2, (5.11). The quality of the match is evaluated via *Best Fit Rate* criterion

$$\text{BFR} = \max\left\{1 - \sqrt{\frac{\sum_{t=1}^{N_{\text{val}}} (y(t) - \hat{y}(t))^2}{\sum_{t=1}^{N_{\text{val}}} (y(t) - \bar{y})^2}}, 0\right\} \times 100\%,$$

where $\bar{y}$ denotes the sample mean of the output in the validation set.

Table 5.1: BFR on a noise-free validation data ($N_{\text{val}} = 500$)

Method	$N = 5000$		$N = 10000$	
	SNR $= 6.1$ dB, $\sigma_{\mathbf{v}} = 1.5$			
	BFR[%]	time[s]	BFR[%]	time[s]
GB + zero noise gain	83.13	197	82.42	361
Algorithm 7	85.67	2.60	90.86	2.03
Algorithm 7 + GB	90.5	256	91.71	323
	SNR $= 0.5$ dB, $\sigma_{\mathbf{v}} = 3$			
GB + zero noise gain	63.25	188	66.11	316
Algorithm 7	76.97	1.94	87.29	1.90
Algorithm 7 + GB	84.17	236	88.03	376

The results are reported in Table 5.1. We compare Algorithm 7 with using only the GB search algorithm from [20] where the initial value of the noise gain is zero, while the other matrices are initialized by the outcome of Algorithm 7. This is then equivalent with combining the CRA algorithm of [19] with GB. Algorithm 7 appears to perform better than using only GB search, especially for noisy data and a larger number of data points. Moreover, Algorithm 7 is faster than GB search. If combined with GB search, the performance of Algorithm 7 improves further, but at the expense of computational time.

5.7 Conclusion

In this chapter, we show a characterization of minimality and uniqueness of LSSs in innovation form, and we presented a realization algorithm for computing minimal LSS in innovation form. Using this realization algorithm, we formulated a system identification algorithm proven to be statistically consistent. The latter algorithm was evaluated on a numerical example and demonstrated promising performance, both in terms of runtime complexity and estimation accuracy.

5.8 Appendix

5.8.1 Proof of Lemma 5.1

In order to present the proof we need some preliminaries definitions, notations and lemmas. Assume that $\mathcal{S}$ of the form (2.1) is a sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. Recall from Notation 2.1 the Hilbert-space $\mathcal{H}_1$. Following the notation of Appendix of Chapter 3, for a ZMWSSI process $\mathbf{r}$ denote by $\mathcal{H}^{\mathbf{r}}_{t,+}$ and $\mathcal{H}^{\mathbf{r}}_{t}$ the closed-subspace of $\mathcal{H}_1$ generated by the components of random variables $\{\mathbf{z}^{\mathbf{r}}_w(t)\}_{w\in\Sigma^+} \cup \{\mathbf{r}(t)\}$ and $\{\mathbf{z}^{\mathbf{r}}_w(t)\}_{w\in\Sigma^+}$ respectively.

Lemma 5.6. *Assume that there exists an sGLSS of* $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. *Then,* $(\{A_\sigma, B_\sigma\}, C, D, \mathbf{u})$ *is an asGLSS of* $(\mathbf{y}^d, \boldsymbol{\mu})$*, if and only if the dLSS* $(\{\sqrt{p_\sigma}A_\sigma, \sqrt{p_\sigma}B_\sigma\}_{\sigma\in\Sigma}, C, D)$ *is a realization of* $\Psi_{\mathbf{u},\mathbf{y}}$ *and* $\sum_{i=1}^{n_\mu} p_i A_i \otimes A_i$ *is Schur.*

Proof of Lemma 5.6. • **Part(1) (only if direction ⇒):** Assume that $(\{A_\sigma, B_\sigma\}_{\sigma\in\Sigma}, C, D, \mathbf{u})$ is an asGLSS of $(\mathbf{y}^d, \boldsymbol{\mu})$. Then $\sum_{i=1}^{n_\mu} p_i A_i \otimes A_i$ is Schur by definition of asGLSS, and using Lemma 2.2 it follows that $\mathbf{y}^d(t) = \sum_{N=0}^{\infty}\sum_{s\in\Sigma^*,|s|=N}\sum_{\sigma\in\Sigma}\sqrt{p_{\sigma s}}CA_sB_\sigma\mathbf{z}^{\mathbf{u}}_{\sigma s}(t) + D\mathbf{u}(t)$.

$$\begin{aligned}
E[\mathbf{y}^d(t)(\mathbf{z}^{\mathbf{u}}_w(t))^T] &= \sum_{N=0}^{\infty}\sum_{s\in\Sigma^*,|s|=N}\sum_{\sigma\in\Sigma}\sqrt{p_{\sigma s}}CA_sB_\sigma E[\mathbf{z}^{\mathbf{u}}_{\sigma s}(t)(\mathbf{z}^{\mathbf{u}}_w(t))^T] + DE[\mathbf{u}(t)(\mathbf{z}^{\mathbf{u}}_w(t))^T] \\
&= \sqrt{p_{\sigma s}}CA_sB_\sigma Q_{\mathbf{u}} = \Psi_{\mathbf{y},\mathbf{u}}(\sigma s)
\end{aligned}$$

This equation follows from the fact that $\mathbf{u}$ is white noise process, and therefore

$$E[\mathbf{u}(t)(\mathbf{z}^{\mathbf{u}}_w(t))^T] = 0$$

and

$$E[\mathbf{z}^{\mathbf{u}}_{\sigma s}(t)(\mathbf{z}^{\mathbf{u}}_{w}(t))^T] = \begin{cases} Q_{\mathbf{u}}, & \text{if } \sigma s = w, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly,

$$E[\mathbf{y}^d(t)\mathbf{u}(t)] = DQ_{\mathbf{u}} = \Psi_{\mathbf{y},\mathbf{u}}(\epsilon)Q_{\mathbf{u}}.$$

Hence, $\Psi_{\mathbf{y},\mathbf{u}}(\sigma s) = \sqrt{p_{\sigma s}}CA_sB_\sigma$ and $\Psi_{\mathbf{y},\mathbf{u}}(\epsilon) = D$ i.e., $(\{\sqrt{p_\sigma}A_\sigma, \sqrt{p_\sigma}B_\sigma\}_{\sigma\in\Sigma}, C, D)$ is a realization of $\Psi_{\mathbf{y},\mathbf{u}}$.

- **Part(2) (if direction $\Leftarrow$):** Assume that $\{(\sqrt{p_\sigma}A_\sigma, \sqrt{p_\sigma}B_\sigma\}_{\sigma\in\Sigma}, C, D)$ is a dLSS realization of $\Psi_{\mathbf{y},\mathbf{u}}$ and $\sum_{i=1}^{n_\mu} p_i A_i \otimes A_i$ is Schur. Consider the asGLSS $\mathcal{S} = (\{A_\sigma, B_\sigma\}_{\sigma\in\Sigma}, C, D, \mathbf{u})$. Note that $\mathcal{S}$ is a well-defined asGLSS, as $\mathbf{u}$ is a white noise w.r.t. $\boldsymbol{\mu}$.

Let $\tilde{\mathbf{x}}$ be the unique trajectory of $\mathcal{S}$, and let $\tilde{\mathbf{y}}$ be the unique output of $\mathcal{S}$. From Lemma 2.2 and $\tilde{\mathbf{y}}(t) = C\tilde{\mathbf{x}}(t) + D\mathbf{u}(t)$ it follows that

$$\tilde{\mathbf{y}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\sigma\in\Sigma} \sqrt{p_{\sigma w}} CA_wB_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t) + D\mathbf{u}(t),$$

where the right-hand side converges in the mean-square sense. Using $\Psi_{\mathbf{y},\mathbf{u}}(\sigma s) = CA_sB_\sigma$, it follows that

$$\tilde{\mathbf{y}}(t) = \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\sigma\in\Sigma} \Psi_{\mathbf{y},\mathbf{u}}(\sigma w)\mathbf{z}^{\mathbf{u}}_{\sigma w}(t) + \Psi_{\mathbf{y},\mathbf{u}}(\epsilon)\mathbf{u}(t) \tag{5.34}$$

By assumption, there exists a sGLSS

$$\mathcal{S} = (\{A'_\sigma, B'_\sigma, K'_\sigma\}_{\sigma\in\Sigma}, C', D', F', \mathbf{v}),$$

where $\mathcal{S}$ is a realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. By the decomposition theorem, Theorem 3.2, of Chapter 3,

$$\mathcal{S}_d = (\{A'_\sigma, B'_\sigma\}_{\sigma\in\Sigma}, C', D', \mathbf{u})$$

is an asGLSS realization of $(\mathbf{y}^d, \boldsymbol{\mu})$. Hence, by **Part(1)** of this lemma, $(\{A'_\sigma, B'_\sigma\}_{\sigma\in\Sigma}, C', D')$ is a dLSS realization of $\Psi_{\mathbf{y},\mathbf{u}}$, where $\Psi_{\mathbf{y},\mathbf{u}}(\sigma s) = \sqrt{p_{\sigma s}}C'A'_sB'_\sigma$ and $\Psi_{\mathbf{y},\mathbf{u}}(\epsilon) = D'$.

But from [91, Lemma 1], it follows that

$$\begin{aligned} \mathbf{y}^d(t) &= \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\sigma\in\Sigma} \sqrt{p_{\sigma w}} C'A'_wB'_\sigma \mathbf{z}^{\mathbf{u}}_{\sigma w}(t) + D'\mathbf{u}(t) \\ &= \sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \sum_{\sigma\in\Sigma} \Psi_{\mathbf{y},\mathbf{u}}(\sigma w)\mathbf{z}^{\mathbf{u}}_{\sigma w}(t) + \Psi_{\mathbf{y},\mathbf{u}}(\epsilon)\mathbf{u}(t) \end{aligned} \tag{5.35}$$

Hence, by combining (5.34) and (5.35) it follows that $\tilde{\mathbf{y}}(t) = \mathbf{y}^d(t)$. More precisely, $\mathcal{S} = (\{A_\sigma, B_\sigma\}_{\sigma\in\Sigma}, C, D, \mathbf{u})$ is a realization of $(\mathbf{y}, \boldsymbol{\mu})$. $\square$

Proof of Lemma 5.1. Assume the sGLSS of the form (2.1) is a realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. From the decomposition theorem, Theorem 3.1 of Chapter 3, it follows that $\mathcal{S}_s$ defined in (3.6) is an asGLSS realization of $(\mathbf{y}^s, \boldsymbol{\mu})$, and $\mathcal{S}_d$ is an asGLSS realization of $(\mathbf{y}^d, \boldsymbol{\mu})$ as defined in (3.5). Consider the dLSS $\mathscr{S}_{\mathcal{S}_s}$ as defined in (2.19).

From Theorem 2.3 it follows that $\mathscr{S}_{\mathcal{S}}$ is a realization of $\Psi_{\mathbf{y}^s}$, i.e.,

$$\Psi_{\mathbf{y}^s}(\sigma s) = \sqrt{p_w} C A_s G_\sigma \tag{5.36}$$

where $w = \sigma s \in \Sigma^+$, and $\Psi_{\mathbf{y}^s}(w) = I$ for $w = \epsilon$. From Lemma 5.6 it follows that $(\{\sqrt{p_\sigma} A_\sigma, \sqrt{p_\sigma} B_\sigma\}_{\sigma=1}^{n_\mu}, C, D)$ is a dLSS realization of $\Psi_{\mathbf{y},\mathbf{u}}$. Hence

$$\Psi_{\mathbf{y},\mathbf{u}}(w) = \begin{cases} \sqrt{p_{\sigma s}} C A_s B_\sigma & w = \sigma s,\ \sigma \in \Sigma,\ s \in \Sigma^* \\ D & w = \epsilon \end{cases} \tag{5.37}$$

Combining (5.36) and (5.37) results in having

$$\begin{aligned} M_{\mathbf{y},\mathbf{u}}(\sigma s) &= \begin{bmatrix} \Psi_{\mathbf{y},\mathbf{u}}(\sigma s) & \Psi_{\mathbf{y}^s}(\sigma s) \end{bmatrix} \\ &= \sqrt{p_s} C A_s \begin{bmatrix} \sqrt{p_\sigma} B_\sigma & G_\sigma \end{bmatrix} \\ M_{\mathbf{y},\mathbf{u}}(\epsilon) &= \begin{bmatrix} \Psi_{\mathbf{y},\mathbf{u}}(\epsilon) & \Psi_{\mathbf{y}^s}(\epsilon) \end{bmatrix} \\ &= \begin{bmatrix} D & I \end{bmatrix} \end{aligned} \tag{5.38}$$

That is, $\mathscr{S}_s$ is a realization of $M_{\mathbf{y},\mathbf{u}}$. □

5.8.2 Proof of Lemma 5.2

Using the terminology of [91, Appendix C, Subsection 1)], consider the recognizable representation

$$\mathcal{R} = (n_x, \{A_\sigma\}_{\sigma \in \Sigma}, \begin{bmatrix} G_1 \dots G_{n_\mu} \end{bmatrix}, C),$$

and consider the formal power series the formal power series

$$\Sigma^* \ni s \mapsto \begin{bmatrix} \Psi_{\mathbf{y}^s}(1s), & \dots, & \Psi_{\mathbf{y}^s}(n_\mu s). \end{bmatrix} \tag{5.39}$$

From [91, Theorem 4] it follows that $\mathcal{R}$ is the representation of the formal power series defined in (5.39). Moreover, $\mathcal{R}$ is stable and observable in the terminology of [91, Appendix C]. The latter follows from the observation that observability of the representation $\mathcal{R}$ is equivalent to the observability of the dLSS $(\{A_i, \begin{bmatrix} B_i & G_i \end{bmatrix}\}_{i=1}^{n_\mu}, C, D)$, see [79, Theorem 5.1]. The the well-posedness of the recursion (5.9) (invertibility of the matrices $\hat{Q}_\sigma^{\mathcal{I}}$), the existence of the limit $K_\sigma = \lim_{I \to \infty} K_\sigma^I$ and the equations (5.9) and (5.10) follow from [91, Lemma 21], by noticing that the proof of [91, Lemma 21] requires only the observability of the representation $\mathcal{R}$. Moreover, from [91, Lemma 21], it follows that the asGLSS $(\{A_i, K_i\}_{i=1}^{n_\mu}, C, I, \mathbf{e}^s)$ is a realization of $(\mathbf{y}^s, \boldsymbol{\mu})$. From Lemma 5.6 it follows that the asGLSS $(\{A_i, B_i\}_{i=1}^{n_\mu}, C, D, \mathbf{u})$ is a realization of $(\mathbf{y}^d, \boldsymbol{\mu})$. Hence by Theorem 3.2 it follows that the sGLSS $(\{A_i, B_i, K_i\}_{i=1}^{n_\mu}, C, D, I, \mathbf{e}^s)$ is a realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.

5.8.3 Proof of Theorem 5.1

Before presenting the proof of Theorem 5.1, we need the following lemma.

Lemma 5.7. *If $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ has an sGLSS realization, and $\mathscr{S} = (\{\hat{A}_\sigma, \hat{B}_\sigma\}_{\sigma\in\Sigma}, \hat{C}, \hat{D})$ is a minimal dLSS realization of $M_{\mathbf{y},\mathbf{u}}$ then $\mathscr{S}$ is observable and $\sum_{i=1}^{n_\mu} p_i \hat{A}_i \otimes \hat{A}_i$ is Schur.*

Proof of Lemma 5.7. Let $\mathscr{S} = (\{\hat{A}_\sigma, \hat{B}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, \mathbf{u})$ be a dLSS of dimension $\hat{n_x}$ We will use the terminology of [91, Appendix C] on formal power series and their recognizable representations. Note that by [79, Section 5] there exists a one-to-one correspondence between recognizable representations of formal power series and dLSS realizations of Markov functions. Consider the formal series

$$\Psi : \Sigma^* \ni s \mapsto \begin{bmatrix} M_{\mathbf{y},\mathbf{u}}(1s) & \dots & M_{\mathbf{y},\mathbf{u}}(n_\mu s) \end{bmatrix} \tag{5.40}$$

Then following [79, Construction 5.1] we can associate with the dLSS $\mathscr{S}$ the recognizable representation

$$R_{\mathscr{S}} = (\hat{n_x}, \{\hat{A}_\sigma\}_{\sigma\in\Sigma}, \begin{bmatrix} \begin{bmatrix} \hat{B}_1 & \hat{G}_1 \end{bmatrix} & \dots & \begin{bmatrix} \hat{B}_{n_\mu} & \hat{G}_{n_\mu} \end{bmatrix} \end{bmatrix}, \hat{C}) \tag{5.41}$$

From [79, Theorem 5.1] it follows that $R_{\mathscr{S}}$ is a recognizable representation of Ψ if and only if $\mathscr{S}$ is a realization of $M_{\mathbf{y},\mathbf{u}}$. It is easy to see that by [79, Lemma 5.3] $R_{\mathscr{S}}$ is reachable (observable), if and only if $\mathscr{S}$ satisfies the reachability and observability rank conditions of [86]. Hence, as it was pointed out in [79, Corollary 5.2], the dLSS $\mathscr{S}$ is minimal, if and only if $R_{\mathscr{S}}$ is minimal.
We will proceed to show that $R_{\mathscr{S}}$ is stable in the sense of [91], i.e.,

$$\sum_{\sigma=1}^{n_\mu} p_\sigma \hat{A}_\sigma \otimes \hat{A}_\sigma \text{ is Schur.}$$

To this end, let $\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v})$ be an sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ of the form of (2.1). Then by Definition 2.6, $\sum_{\sigma=1}^{n_\mu} p_\sigma A_\sigma \otimes A_\sigma$ is Schur. Consider the dLSS $\mathscr{S}_{\mathcal{S}}$ associated with $\mathcal{S}$. By Lemma 5.1, $\mathscr{S}_{\mathcal{S}}$ is a realization of $M_{\mathbf{y},\mathbf{u}}$. Let us consider the formal series Ψ defined in (5.40) above and let us consider the recognizable representation

$$R_{\mathscr{S}_{\mathcal{S}}} = (\{\sqrt{p_\sigma} A_\sigma\}_{\sigma\in\Sigma}, \tilde{G}, C)$$

associated with the dLSS $\mathscr{S}_{\mathcal{S}}$, where

$$\tilde{G} = \begin{bmatrix} \begin{bmatrix} \sqrt{p_i} B_1 & G_1 \end{bmatrix} & \dots & \begin{bmatrix} \sqrt{p_{n_\mu}} B_{n_\mu} & G_{n_\mu} \end{bmatrix} \end{bmatrix}.$$

With the same argument as above, namely, by using the fact that $\mathscr{S}_{\mathcal{S}}$ is a realization of $M_{\mathbf{y},\mathbf{u}}$ and [79, Theorem 5.1], it is easy to see that $R_{\mathscr{S}_{\mathcal{S}}}$ is a representation of Ψ. Since $R_{\mathscr{S}_{\mathcal{S}}}$ is stable in the terminology of [91, Appendix C] (as $\sum_{i=1}^{n_\mu} \sqrt{p_\sigma} A_\sigma \otimes \sqrt{p_\sigma} A_\sigma = \sum_{i=1}^{n_\mu} p_\sigma A_\sigma \otimes A_\sigma$ is Schur), then by [91, Theorem 6], Ψ is square summable, i.e.,

$$\sum_{N=0}^{\infty} \sum_{\substack{|w| = N \\ w \in \Sigma^*}} \|\Psi(w)\|_2^2 < +\infty$$

From [91, Theorem 6] it follows that any minimal representation of Ψ is stable. In particular, if $\mathscr{S}$ is minimal, then $R_{\mathscr{S}_\mathcal{S}}$ is minimal, therefore, $R_{\mathscr{S}_\mathcal{A}}$ is stable, i.e., $\sum_{\sigma=1}^{n_\mu} p_\sigma \hat{A}_\sigma \otimes \hat{A}_\sigma$ is Schur. The observability of $\mathscr{S}$ follows from its minimality. $\square$

Proof of Theorem 5.1. • **Part 1: Observability and reachability $\implies$ minimality**.
Assume that $\mathcal{S}$ is reachable and observable sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. Then by definition the dLSS $\mathscr{S}_\mathcal{S}$ is also reachable and observable, and hence it is a minimal realization of $M_{\mathbf{y};\mathbf{u}}$ by [86]. Let $\mathcal{S}'$ be another sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. Then by Lemma 5.1 the dLSS $\mathscr{S}_{\mathcal{S}'}$ associated with $\mathcal{S}'$ has the same Markov parameters as $\mathscr{S}_\mathcal{S}$. Hence, since $\mathscr{S}_\mathcal{S}$ is minimal dimensional, the dimension of $\mathscr{S}_\mathcal{S}$ cannot exceed that of $\mathscr{S}_{\mathcal{S}'}$. Since the dimension of a sGLSS is the same as that of the dLSS associated with it, it follows that the dimension of $\mathcal{S}$ cannot exceed that of $\mathcal{S}'$.

• **Part 2:** Consider the following minimal dLSS:

$$\mathscr{S}_m = (\{\hat{A}_\sigma, \begin{bmatrix}\hat{B}_\sigma & \hat{G}_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, \hat{C}, \begin{bmatrix}\hat{D} & I_{n_y}\end{bmatrix}),$$

then its associated sGLSS can be presented as:

$$\mathcal{S}_{\mathscr{S}_m} = (\{\frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma, \frac{1}{\sqrt{p_\sigma}}\hat{B}_\sigma, \hat{K}_\sigma\}_{\sigma=1}^{n_\mu}, \hat{C}, \hat{D}, I_{n_y}, \mathbf{e}^s),$$

where $\hat{K}_\sigma = \lim_{\mathcal{I}\to\infty} \hat{K}_\sigma^\mathcal{I}$, and $\{\hat{K}_\sigma^\mathcal{I}\}_{\sigma\in\Sigma,\mathcal{I}\in\mathbb{N}}$ satisfies

$$\hat{K}_\sigma^\mathcal{I} = \left(\hat{G}_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma\hat{P}_\sigma^\mathcal{I}(\hat{C})^T\right)\left(\hat{Q}_\sigma^\mathcal{I}\right)^{-1}$$

as in (5.9). First, we claim that the dLSS associated with $S_{\mathscr{S}_m}$ equals $\mathscr{S}_m$. Indeed, the dLSS associated with $\mathcal{S}_{\mathscr{S}_m}$ can be computed from (5.4) as follows:

$$\begin{aligned}\mathscr{S}_{\mathcal{S}_{\mathscr{S}_m}} &= (\{\sqrt{p_\sigma}\frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma, \begin{bmatrix}\sqrt{p_\sigma}\frac{1}{\sqrt{p_\sigma}}B_\sigma & \tilde{G}_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, \hat{C}, \begin{bmatrix}\hat{D} & I_{n_y}\end{bmatrix}) \\ &= (\{\hat{A}_\sigma, \begin{bmatrix}\hat{B}_\sigma & \tilde{G}_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, \hat{C}, \begin{bmatrix}\hat{D} & I_{n_y}\end{bmatrix})\end{aligned}$$

where $\tilde{G}_\sigma$ have the following expression as in (5.3):

$$\tilde{G}_\sigma = \sqrt{p_\sigma}(\frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma T_{\sigma,\sigma}^{\mathbf{x}^s,\mathbf{x}^s}C^T + \hat{K}_\sigma T_{\sigma,\sigma}^{\mathbf{e}^s,\mathbf{e}^s}).$$

Note that from (5.10), we have:

$$p_\sigma T_{\sigma,\sigma}^{\mathbf{e}^s,\mathbf{e}^s} = \lim_{\mathcal{I}\to\infty}\hat{Q}_\sigma^\mathcal{I} = \hat{Q}_\sigma, \quad p_\sigma T_{\sigma,\sigma}^{\hat{\mathbf{x}}^s,\hat{\mathbf{x}}^s} = \lim_{\mathcal{I}\to\infty}\hat{P}_\sigma^\mathcal{I} = \hat{P}_\sigma,$$

then, we get the following:

$$\begin{aligned}\tilde{G}_\sigma &= \sqrt{p_\sigma}(\frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma\frac{1}{p_\sigma}\hat{P}_\sigma C^T + \hat{K}_\sigma\frac{1}{p_\sigma}\hat{Q}_\sigma) \\ &= \frac{1}{p_\sigma}\hat{A}_\sigma\hat{P}_\sigma C^T + \frac{1}{\sqrt{p_\sigma}}\left(\hat{G}_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}\hat{A}_\sigma\hat{P}_\sigma(\hat{C})^T\right)\left(\hat{Q}_\sigma\right)^{-1}\hat{Q}_\sigma \\ &= \frac{1}{p_\sigma}\hat{A}_\sigma\hat{P}_\sigma C^T + \hat{G}_\sigma - \frac{1}{p_\sigma}\hat{A}_\sigma\hat{P}_\sigma(\hat{C})^T = \hat{G}_\sigma.\end{aligned}$$

Then $\mathscr{S}_m = \mathscr{S}_{\mathcal{S}_{\mathscr{S}_m}} = (\{\hat{A}_\sigma, \begin{bmatrix}\hat{B}_\sigma & \hat{G}_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, \hat{C}, \begin{bmatrix}\hat{D} & I_{n_y}\end{bmatrix})$. Since $\mathscr{S}_m$ is minimal, it is reachable and observable. Hence, by the if direction of Part 1 proven above, $\mathcal{S}_{\mathscr{S}_m}$ is reachable and observable, and hence minimal. The rest of the statement follows from Lemma 5.2, i.e., $\mathcal{S}_{\mathscr{S}_m}$ is a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.

- **Part 1: Minimality $\implies$ observability and reachability**.

Assume that $\mathcal{S}$ is a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. Then by Lemma 5.1, the dLSS $\mathscr{S}_\mathcal{S}$ is a realization of $M_{\mathbf{y},\mathbf{u}}$. We argue that $\mathscr{S}_\mathcal{S}$ is also a minimal dLSS realization of $M_{\mathbf{y},\mathbf{u}}$. If the latter is true, then $\mathscr{S}_\mathcal{S}$ is observable and reachable, and hence by definition of observability and span-reachability of sGLSSs, so is $\mathcal{S}$.
It remains to show that the associated dLSS $\mathscr{S}_\mathcal{S}$ is indeed minimal. Assume that $\mathscr{S}_\mathcal{S}$ is not minimal. Consider a minimal dLSS realization $\mathscr{S}_m$ of $M_{\mathbf{y},\mathbf{u}}$. It then follows that the dimension of $\mathscr{S}_m$ is strictly smaller than the dimension of $\mathscr{S}_{\mathcal{S}_m}$, and the latter is equals to $\dim\ \mathcal{S}$. From Part 2 it follows that the sGLSS $\mathcal{S}_{\mathscr{S}_m}$ is a minimal realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$, and by construction, the dimension of $\mathcal{S}_{\mathscr{S}_m}$ equals the dimension of $\mathscr{S}_m$. The latter is strictly smaller than $\dim \mathcal{S}$, i.e., $\dim \mathcal{S}_{\mathscr{S}_m} < \dim \mathcal{S}$. The latter contradicts to minimality of $\mathcal{S}$.

- **Part 3:** Let

$$\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v})$$

be a sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$. From Lemma 5.1, we define

$$\mathscr{S}_\mathcal{S} = (\{\sqrt{p_\sigma}A_\sigma, \begin{bmatrix}\sqrt{p_\sigma}B_\sigma & G_\sigma\end{bmatrix}\}_{\sigma=1}^{n_\mu}, C, \begin{bmatrix}D & I_{n_y}\end{bmatrix}),$$

the associated dLSS, which is a realization of $M_{\mathbf{y},\mathbf{u}}$. From $\mathscr{S}_\mathcal{S}$ as an input of Algorithm 5 and from Lemma 5.3, we can get a minimal dLSS

$$\hat{\mathscr{S}} = (\{\hat{A}_\sigma, \hat{B}_\sigma\}_{\sigma\in\Sigma}, \hat{C}, \hat{D})$$

realization $M_{\mathbf{y},\mathbf{u}}$. We know, from Lemma 5.7, that $\sum_{i=1}^{n_\mu} p_i\hat{A}_i \otimes \hat{A}_i$ is Schur. It follows, from **part 2** of Theorem 5.1, that the associated sGLSS $\hat{\mathcal{S}}_\mathscr{S}$ is in innovation form and it is a minimal realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$.
Therefore, if $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ has a sGLSS realization, then there exists a minimal sGLSS realization in innovation form.

- **Part 4:** Let the two systems

$$\mathcal{S} = (\{A_\sigma, B_\sigma, K_\sigma\}_{\sigma=1}^{n_\mu}, C, D, F, \mathbf{v})$$
$$\widetilde{\mathcal{S}} = (\{\tilde{A}_\sigma, \tilde{B}_\sigma, \tilde{K}_\sigma\}_{\sigma=1}^{n_\mu}, \tilde{C}, \tilde{D}, \tilde{F}, \mathbf{v})$$

be two sGLSSs, and assume that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are minimal sGLSS of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form, i.e., $F = \tilde{F} = I_{n_y}$, and $\mathbf{v} = \mathbf{e}^s$. Let, also, the two systems

$$\mathscr{S}_{\mathcal{S}} = (\{\sqrt{p_i}A_i, \begin{bmatrix} G_i, \sqrt{p_i}B_i \end{bmatrix}\}_{i=0}^{n_\mu}, C, D)$$
$$\mathscr{S}_{\widetilde{\mathcal{S}}} = (\{\sqrt{p_i}\tilde{A}_i, \begin{bmatrix} \tilde{G}_i & \sqrt{p_i}\tilde{B}_i \end{bmatrix}\}_{i=0}^{n_\mu}, \tilde{C}, \tilde{D})$$

be the corresponding dLSSs. By Lemma 5.1, $\mathscr{S}_{\mathcal{S}}$ and $\mathscr{S}_{\widetilde{\mathcal{S}}}$ are realizations of the same Markov function which is $M_{\mathbf{y},\mathbf{u}}$, and they are both reachable and observable, hence they minimal. Thus, $\tilde{D} = D$, and by minimality of $\mathscr{S}_{\mathcal{S}}$ and $\mathscr{S}_{\widetilde{\mathcal{S}}}$, they are isomorphic, i.e., there exists a non-singular matrix T such that for $i = 1, \dots, n_\mu$ the following equations hold:

$$\begin{aligned} \tilde{A}_i = TA_iT^{-1}, \quad \tilde{B}_i = TB_i, \quad \tilde{G}_i = TG_i, \\ \tilde{C} = CT^{-1}, \qquad D = \tilde{D}. \end{aligned}$$

In order to show that $\tilde{K}_i = TK_i$, notice that $\tilde{G}_i = TG_i$, $i = 1, \dots, n_\mu$ holds, and use the the proof of [91, Part (ii), Theorem 2], by applying it to the sGLSS $\mathcal{S}_s = (\{A_i, K_i\}_{i=1}^{n_\mu}, C, I, \mathbf{e}^s)$ and $\widetilde{\mathcal{S}}_s = (\{\tilde{A}_i, \tilde{K}_i, \}_{i=1}^{n_\mu}, \tilde{C}, I, \mathbf{e}^s)$. More precisely, T satisfies the properties of the matrix T defined in the proof of [91, Part (ii), Theorem 2]. Then, the rest of the proof of [91, Part (ii), Theorem 2] can be repeated verbatim. □

5.8.4 Proof of Lemma 5.4

Proof. By definition of decomposition presented in Chapter 3, we know that

$$\mathbf{y}(t) = \mathbf{y}^s(t) + \mathbf{y}^d(t).$$

From that, we have

$$\mathbf{z}_w^{\mathbf{y}}(t) = \mathbf{z}_w^{\mathbf{y}^s}(t) + \mathbf{z}_w^{\mathbf{y}^d}(t).$$

From [91, Lemma 11], we know that the entries of $\mathbf{y}^d(t)$ and the entries of $\mathbf{z}_w^{\mathbf{y}^d}(t)$ are in $\mathcal{H}_{t,+}^{\mathbf{u}}$. From Lemma 3.5, it follows that $\mathbf{z}_w^{\mathbf{y}^s}(t)$ is orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$ and hence to $\mathbf{z}_w^{\mathbf{y}^d}(t)$. That is,

$$T_{\sigma,\sigma}^{\mathbf{y},\mathbf{y}} = E\left[\mathbf{y}(t)\mathbf{z}_w^{\mathbf{y}}(t)\right] = E\left[\mathbf{y}^s(t)\mathbf{z}_w^{\mathbf{y}^s}(t)\right] + E\left[\mathbf{y}^s(t)\mathbf{z}_w^{\mathbf{y}^d}(t)\right] + E\left[\mathbf{y}^d(t)\mathbf{z}_w^{\mathbf{y}^s}(t)\right] + E\left[\mathbf{y}^d(t)\mathbf{z}_w^{\mathbf{y}^d}(t)\right].$$

And since $\mathbf{y}^s(t)$ is orthogonal to $\mathcal{H}_{t,+}^{\mathbf{u}}$ and the entries of $\mathbf{z}_w^{\mathbf{y}^d}(t)$ belong to $\mathcal{H}_{t,+}^{\mathbf{u}}$, then

$$E\left[\mathbf{y}^s(t)\mathbf{z}_w^{\mathbf{y}^d}(t)\right] = 0.$$

Similarly, since the components of $\mathbf{y}^d(t)$ are in $\mathcal{H}^{\mathbf{u}}_{t,+}$ and $\mathbf{z}^{\mathbf{y}^s}_w(t)$ is orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$ by Lemma 3.5, then $E\left[\mathbf{y}^d(t)\mathbf{z}^{\mathbf{y}^s}_w(t)\right] = 0$. Then

$$\Lambda^{\mathbf{y},\mathbf{y}}_w = E\left[\mathbf{y}(t)\mathbf{z}^{\mathbf{y}}_w(t)\right] = E\left[\mathbf{y}^s(t)\mathbf{z}^{\mathbf{y}^s}_w(t)\right] + E\left[\mathbf{y}^d(t)\mathbf{z}^{\mathbf{y}^d}_w(t)\right] = \Lambda^{\mathbf{y}^s,\mathbf{y}^s}_w + \Lambda^{\mathbf{y}^d,\mathbf{y}^d}_w.$$

Also, note that

$$\begin{aligned}\Lambda^{\mathbf{y}^d,\mathbf{u}}_w &= E\left[\mathbf{y}^d(t)\mathbf{z}^{\mathbf{u}}_w(t)\right] = E\left[(\mathbf{y}(t) - \mathbf{y}^s(t))\mathbf{z}^{\mathbf{u}}_w(t)\right] \\ &= E\left[\mathbf{y}(t)\mathbf{z}^{\mathbf{u}}_w(t)\right] - E\left[\mathbf{y}^s(t)\mathbf{z}^{\mathbf{u}}_w(t)\right] = E\left[\mathbf{y}(t)\mathbf{z}^{\mathbf{u}}_w(t)\right]\end{aligned} \tag{5.42}$$

since $\mathbf{y}^s(t)$ is orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$ and hence to $\mathbf{z}^{\mathbf{u}}_w(t)$, thus

$$E\left[\mathbf{y}^s(t)(\mathbf{z}^{\mathbf{u}^s}_w(t))^T\right] = 0.$$

Then $\Lambda^{\mathbf{y}^d,\mathbf{u}}_w = \Lambda^{\mathbf{y},\mathbf{u}}_w$. Finally, note that

$$\begin{aligned}T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma} = E\left[\mathbf{z}^{\mathbf{y}}_\sigma(t)(\mathbf{z}^{\mathbf{y}}_\sigma(t))^T\right] &= E\left[\mathbf{z}^{\mathbf{y}^s}_\sigma(t)(\mathbf{z}^{\mathbf{y}^s}_\sigma(t))^T\right] + E\left[\mathbf{z}^{\mathbf{y}^s}_\sigma(t)(\mathbf{z}^{\mathbf{y}^d}_\sigma(t))^T\right] \\ &+ E\left[\mathbf{z}^{\mathbf{y}^d}_\sigma(t)(\mathbf{z}^{\mathbf{y}^s}_\sigma(t))^T\right] + E\left[\mathbf{z}^{\mathbf{y}^d}_\sigma(t)(\mathbf{z}^{\mathbf{y}^d}_\sigma(t))^T\right].\end{aligned} \tag{5.43}$$

By definition, the components of $\mathbf{y}^d(t)$ belong $\mathcal{H}^{\mathbf{u}}_{t,+}$, and hence by [91, Lemma 11], the components of $\mathbf{z}^{\mathbf{y}^d}_\sigma(t)$ belong to $\mathcal{H}^{\mathbf{u}}_{t,+}$. By Lemma 3.5, $\mathbf{z}^{\mathbf{y}^s}_\sigma(t)$ is orthogonal to $\mathcal{H}^{\mathbf{u}}_{t,+}$ hence

$$E\left[\mathbf{z}^{\mathbf{y}^s}_\sigma(t)(\mathbf{z}^{\mathbf{y}^d}_\sigma(t))^T\right] = E\left[\mathbf{z}^{\mathbf{y}^d}_\sigma(t)(\mathbf{z}^{\mathbf{y}^s}_\sigma(t))^T\right] = 0.$$

Subsequently, $T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma} = T^{\mathbf{y}^d,\mathbf{y}^d}_{\sigma,\sigma} + T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}$. Then equation (5.19) of Lemma 5.4 holds. Equation (5.20) of Lemma 5.4 follows from [91, Lemma 4], which is recalled in Section 2.7.4. □

5.8.5 Proof of Corollary 5.1

Proof. Let us recall the equation (5.2) and Definition 5.2.

$$M_{\mathbf{y},\mathbf{u}}(w) = \begin{cases} \left[\Lambda^{\mathbf{y}^d,\mathbf{u}}_w Q^{-1}_{\mathbf{u}}, \quad \Lambda^{\mathbf{y}^s,\mathbf{y}^s}_w\right] = \left[\Psi_{\mathbf{u},\mathbf{y}}(w), \quad \Psi_{\mathbf{y}^s}(w)\right] & w \neq \epsilon, \\ \left[\Lambda^{\mathbf{y}^d,\mathbf{u}}_\epsilon Q^{-1}_{\mathbf{u}}, \quad I_{n_y}\right] = \left[\Psi_{\mathbf{u},\mathbf{y}}(w), \quad I_{n_y}\right] & w = \epsilon. \end{cases}$$

From Lemma 5.3, it follows that

$$\mathscr{S}_d = (\{\tilde{A}^d_i, \tilde{B}^d_i\}_{i=1}^{n_\mu}, \tilde{C}^d, \tilde{D}^d)$$

is a minimal dLSS realization of the covariance sequence $\Psi_{\mathbf{u},\mathbf{y}}(w)$. Then from Lemma 5.4, it follows that Step 3 and Step 4 of Algorithm 6 result in computing $\{M_{\mathbf{y},\mathbf{u},\mathcal{J}}(w)\}_{w \in \mathcal{L}_{\alpha,\beta}}$, and (5.23) holds, hence for all $w \in \Sigma^*$

$$\lim_{\mathcal{J}\to\infty} M_{\mathbf{y},\mathbf{u},\mathcal{J}}(w) = M_{\mathbf{y},\mathbf{u}}(w) \tag{5.44}$$

In particular, $\lim_{\mathcal{J}\to\infty} \mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}} = \mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u}}}$ and $\lim_{\mathcal{J}\to\infty} \det(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}}) = \det(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u}}})$ where $\det(X)$ denotes the determinant of a matrix X. Since $\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u}}}$ is full rank, $\det(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u}}}) \neq 0$ and hence there exists an integer $\mathcal{J}_{\star,1} > 0$ such that for all $\mathcal{J} > \mathcal{J}_{\star,1}$ it holds that $\det(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}}) \neq 0$, and hence $\mathrm{rank}\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}} = n_x$, i.e., the first part of (5.26) holds. Moreover, it is easy to see that when Algorithm 5 is applied to a Markov functions M such that $\mathcal{H}_{\alpha,\beta}^{M}$ is full rank, the resulting matrices are continuous functions of the finite collection $\{M(w)\}_{w\in\mathcal{L}_{\alpha,\beta}}$. Hence, if

$$\mathscr{S}_d = (\{\hat{A}_i, \begin{bmatrix}\hat{B}_i, & \hat{G}_i\end{bmatrix}\}_{i=1}^{n_\mu}, \hat{C}, \hat{D})$$

be the dLSS returned by Algorithm 5 applied to $M_{\mathbf{y},\mathbf{u}}$, then from (5.44) and continuity of Algorithm 5 it follows that

$$\begin{aligned}
&\lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{A}_\sigma = \hat{A}_\sigma, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{B}_\sigma = \hat{B}_\sigma, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{G}_\sigma = \hat{G}_\sigma, \quad \sigma \in \Sigma \\
&\lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{C} = \hat{C}, \quad \lim_{\mathcal{J}\to\infty} {}^{\mathcal{J}}\hat{D} = \hat{D}
\end{aligned} \tag{5.45}$$

In particular (5.27) holds. By Lemma 5.3, $\mathscr{S}_d$ is a minimal dLSS realization of $M_{\mathbf{y},\mathbf{u}}$. Let $\hat{K}_\sigma^{\mathcal{I}}$, $\hat{Q}_\sigma^{\mathcal{I}}$ as in (5.9) for $\mathcal{I} = 0,\ldots,\bar{\mathcal{I}}$. From Lemma 5.2 it follows that the limits

$$\hat{Q}_\sigma := p_\sigma T_{\sigma,\sigma}^{\mathbf{e}^s,\mathbf{e}^s} = \lim_{\mathcal{I}\to\infty} \hat{Q}_\sigma^{\mathcal{I}}, \quad \hat{K}_\sigma := \lim_{\mathcal{I}\to\infty} \hat{K}_\sigma^{\mathcal{I}} \tag{5.46}$$

exist, and $\hat{Q}_\sigma^{\mathcal{I}}$ is invertible for all $\sigma \in \Sigma$. By induction on $\mathcal{I}$ it is easy to see that $\hat{Q}_\sigma^{\mathcal{I}}$ and $\hat{K}_\sigma^{\mathcal{I}}$ are continuous functions of $\hat{A}_\sigma, \hat{G}_\sigma, \hat{C}, \hat{D}$, $T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s}$ and the latter function defined on an open domain.

More precisely, for matrices $\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \bar{C}, \bar{D}$, define the tuple of matrices

$$\begin{aligned}
&\{\bar{K}_\sigma^{\mathcal{I}}, \bar{Q}_\sigma^{\mathcal{I}}\}_{\sigma\in\Sigma,\mathcal{I}=0,\ldots,\bar{\mathcal{I}}} = g_{\bar{\mathcal{I}}}(\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \bar{C}, \bar{D}) \\
&\bar{P}_\sigma^{\mathcal{I}+1} = p_\sigma \sum_{\sigma_1\in\Sigma} \frac{1}{p_{\sigma_1}} \left(\bar{A}_{\sigma_1}\bar{P}_{\sigma_1}^{\mathcal{I}}\bar{A}_{\sigma_1}^T + \bar{K}_{\sigma_1}^{\mathcal{I}}\bar{Q}_{\sigma_1}^{\mathcal{I}}(\bar{K}_{\sigma_1}^{\mathcal{I}})^T\right) \\
&\bar{Q}_\sigma^{\mathcal{I}} = p_\sigma \bar{T}_{\sigma,\sigma} - \bar{C}\bar{P}_\sigma^{\mathcal{I}}(\bar{C})^T \\
&\bar{K}_\sigma^{\mathcal{I}} = \left(\bar{G}_\sigma\sqrt{p_\sigma} - \frac{1}{\sqrt{p_\sigma}}\bar{A}_\sigma\bar{P}_\sigma^{\mathcal{I}}(\bar{C})^T\right)\left(\bar{Q}_\sigma^{\mathcal{I}}\right)^{-1} \\
&\bar{P}_\sigma^0 = 0,\ \mathcal{I} = 0,\ldots,\bar{\mathcal{I}}, \sigma \in \Sigma
\end{aligned} \tag{5.47}$$

Define the function

$$\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}\}_\sigma, \bar{C}, \bar{D} \mapsto g_{\bar{\mathcal{I}}}(\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \bar{C}, \bar{D}) \tag{5.48}$$

on the domain which consists of matrices $\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}\}_\sigma, \bar{C}, \bar{D}$ such that $\sum_{\sigma\in\Sigma} \bar{A}_\sigma \otimes \bar{A}_\sigma$ is Schur, $\bar{T}_{\sigma,\sigma}$ is positive definite for all $\sigma \in \Sigma$, and the recursion in (5.47) is well-posed, $\bar{Q}_\sigma^{\mathcal{I}}$ are invertible for all $\sigma \in \Sigma$. Notice that if $\bar{T}_{\sigma,\sigma}$ are positive definite, then $\bar{Q}_\sigma^0$ is well defined and invertible, so for

$\bar{\mathcal{I}} = 0$ the domain of definition of $g_{\bar{\mathcal{I}}}$ is not empty. Note that for any $\bar{\mathcal{I}}$, $\{\hat{A}_\sigma, \hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \hat{C}, \hat{D}$ is in the domain of definition of $g_{\bar{\mathcal{I}}}$ for all $\bar{\mathcal{I}}$, and in fact,

$$g_{\bar{\mathcal{I}}}(\{\hat{A}_\sigma, \hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \hat{C}, \hat{D}) = \{\hat{K}^{\mathcal{I}}_\sigma, \hat{Q}^{\mathcal{I}}_\sigma\}_{\sigma\in\Sigma,\mathcal{I}=0,\ldots,\bar{\mathcal{I}}} \tag{5.49}$$

That is, intuitively the function $g_{\bar{\mathcal{I}}}$ represents the application of (5.9) to matrices which correspond to an estimate of a dLSS realization of $M_{\mathbf{y},\mathbf{u}}$ and of the covariances $T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}$, $\sigma \in \Sigma$.

Since $\bar{P}^{\mathcal{I}+1}_\sigma$ is a polynomial function of the entries of $\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}, \bar{Q}^{\mathcal{I}}_\sigma, \bar{K}^{\mathcal{I}}_\sigma, \bar{P}^{\mathcal{I}}_\sigma\}_{\sigma\in\Sigma}$, and $\bar{Q}^{\mathcal{I}}_\sigma, \bar{K}^{\mathcal{I}}_\sigma$ are rational functions of $\{\bar{A}_\sigma, \bar{G}_\sigma, \bar{T}_{\sigma,\sigma}, \bar{P}^{\mathcal{I}}_\sigma\}_{\sigma\in\Sigma}, \bar{C}, \bar{D}$, by induction on $\mathcal{I}$ it follows that the domain of definition of $g_{\bar{\mathcal{I}}}$ is an open set (if we identify the tuples of matrices which serve as its argument with a subspace of an Euclidean space), and $g_{\bar{\mathcal{I}}}$ is continuous on its domain of definition.

From (5.45) and (5.23) it then follows that the tuple $(\{{}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D})$ converges to $(\{\hat{A}_\sigma, \hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \hat{C}, \hat{D})$ as $\mathcal{J} \to \infty$. The latter tuple is in the domain of definition $g_{\bar{\mathcal{I}}}$. Since the domain of definition $g_{\bar{\mathcal{I}}}$ is open, it follows that there exists an integer $\mathcal{J}_{\star,2} > 0$ such that for all $\mathcal{J} > \mathcal{J}_{\star,2}$, the tuple $(\{{}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D})$ belongs to the domain of definition of $g_{\bar{\mathcal{I}}}$. This then implies that

$$g_{\bar{\mathcal{I}}}(\{{}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D}) = \{{}^{\mathcal{J}}\hat{Q}^{\mathcal{I}}_\sigma, {}^{\mathcal{J}}\hat{K}^{\mathcal{I}}_\sigma\}_{\sigma\in\Sigma,\mathcal{I}=0,\ldots,\bar{\mathcal{I}}}$$

is well-defined and the matrices ${}^{\mathcal{J}}\hat{Q}^{\mathcal{I}}_\sigma$ are invertible for all $\mathcal{I} = 0, \ldots, \bar{\mathcal{I}}$, $\sigma \in \Sigma$. Finally, from continuity of $g_{\bar{\mathcal{I}}}$, from (5.49)and the fact that the tuple $(\{{}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D})$ converges to $(\{\hat{A}_\sigma, \hat{G}_\sigma, T^{\mathbf{y}^s,\mathbf{y}^s}_{\sigma,\sigma}\}_{\sigma\in\Sigma}, \hat{C}, \hat{D})$ as $\mathcal{J} \to \infty$ it follows that (5.28) holds with $\hat{\hat{K}}^{\mathcal{I}}_\sigma := \hat{K}^{\mathcal{I}}_\sigma$, $\hat{\hat{Q}}^{\mathcal{I}}_\sigma := \hat{Q}^{\mathcal{I}}_\sigma$. That is, if we define $\mathcal{J}_\star = \max\{\mathcal{J}_{\star,1}, \mathcal{J}_{\star,2}\}$, then for all $\mathcal{J} > \mathcal{J}_\star$, (5.26) holds and we have shown (5.27)-(5.28). Finally, (5.29) follows by combining (5.46) and (5.28).

From **part 2** of Theorem 5.1, it follows that the sGLSS $\tilde{\mathcal{S}}$ from (5.30) is a minimal sGLSS realization of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\mu})$ in innovation form. Finally, (5.31) is a direct consequence of (5.46). The last statement of the Corollary 5.1 follows from Lemma 5.3. □

5.8.6 Proof of Lemma 5.5

Proof. The main idea of the proof is to show that the result of Algorithm 7 converges to the result of the realization algorithm Algorithm 6.

To facilitate the discussion, we will use the notation of Algorithm 6 to denote the matrices, dLSSs and covariances computed in Step 1-Step 6, Algorithm 6, when the latter is applied to the *true covariance* $\Lambda^{\mathbf{y},\mathbf{y}}_w$, $\Lambda^{\mathbf{y},\mathbf{u}}_w$, $\Lambda^{\mathbf{y},\mathbf{u}}_\epsilon$, and $T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma}$ $w \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}}$, $\sigma \in \Sigma$.

In contrast, we will add an index N to the left of the index $\mathcal{J}$ or to the subscript/superscript to denote the dLSSs and covariances computed by Step 1-Step 6, Algorithm 6, if the latter is applied to data where the true covariances $\Lambda^{\mathbf{y},\mathbf{y}}_w$, $\Lambda^{\mathbf{y},\mathbf{u}}_w$, $\Lambda^{\mathbf{y},\mathbf{u}}_\epsilon$, and $T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma}$ are replaced by the *empirical covariances* $T^{\mathbf{y},\mathbf{y}}_{\epsilon,w,N}$, $T^{\mathbf{y},\mathbf{u}}_{\epsilon,w,N}$, $T^{\mathbf{y},\mathbf{u}}_{\epsilon,\epsilon,N}$ and $T^{\mathbf{y},\mathbf{y}}_{\sigma,\sigma,N}$ respectively for $w \in \mathcal{L}_{\alpha,\beta} \cup \mathcal{L}_{\bar{\alpha},\bar{\beta}}$, $\sigma \in \Sigma$.

In particular, for Step 1, for every $w \in \Sigma^*$, $\sigma \in \Sigma$, define the quantities $\Lambda^{\mathbf{y}^d,\mathbf{u},N}_w$ as

$$\Lambda^{\mathbf{y}^d,\mathbf{u},N}_w = T^{\mathbf{y},\mathbf{u}}_{\epsilon,w,N}, \tag{5.50}$$

and define the Markov function

$$\Psi_{\mathbf{y},\mathbf{u},N} : \Sigma^* \ni w \mapsto \Lambda_w^{\mathbf{y}^d,\mathbf{u},N} Q_{\mathbf{u}}^{-1} \tag{5.51}$$

For Step 2 of Algorithm 6 applied to empirical covariances, denote by

$$\mathscr{S}_{d,N} = (\{\tilde{A}_i^{d,N}, \tilde{B}_i^{d,N}\}_{i=1}^{n_\mu}, \tilde{C}^{d,N}, \tilde{D}^{d,N})$$

the result of applying Algorithm 5 to the Markov function $\Psi_{\mathbf{y},\mathbf{u},N}$ and selections $(\bar{\alpha},\bar{\beta})$.

For Step 3 of Algorithm 6 applied to empirical covariances, denote by $\Lambda_{\sigma w}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}}$, $\Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,N,\mathcal{J}}$, $T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,N,\mathcal{J}}$ the quantities obtained by using (5.21) but with $\Lambda_{\sigma w}^{\mathbf{y},\mathbf{y}}$ replaced by $T_{\epsilon,\sigma w,N}^{\mathbf{y},\mathbf{y}}$, and $\Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}}$, $T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,\mathcal{J}}$ replaced by the right hand sides of (5.21) where instead of

$$\tilde{A}_\sigma^d, \tilde{B}_\sigma^d, \tilde{C}^d, \tilde{D}^d$$

we use

$$\tilde{A}_\sigma^{d,N}, \tilde{B}_\sigma^{d,N}, \tilde{C}^{d,N}, \tilde{D}^{d,N}$$

for all $\sigma \in \Sigma$, i.e., for all $\sigma \in \Sigma$, $w \in \Sigma^*$

$$\begin{aligned}
\Lambda_{\sigma w,N}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}} &= T_{\epsilon,\sigma w,N}^{\mathbf{y},\mathbf{y}} - \Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,N,\mathcal{J}}, \quad T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}} = T_{\sigma,\sigma,N}^{\mathbf{y},\mathbf{y}} - T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,N,\mathcal{J}} \\
\Lambda_{\sigma w}^{\mathbf{y}^d,\mathbf{y}^d,N,\mathcal{J}} &= \frac{1}{p_\sigma}\tilde{C}^{d,N}\tilde{A}_w^{d,N}(\tilde{A}_\sigma^{d,N}\tilde{P}_\sigma^{N,\mathcal{J}}(\tilde{C}^{d,N})^T + p_\sigma \tilde{B}_\sigma^{d,N} Q_{\mathbf{u}} (\tilde{D}^{d,N})^T), \\
T_{\sigma,\sigma}^{\mathbf{y}^d,\mathbf{y}^d,N,\mathcal{J}} &= \frac{1}{p_\sigma}(\tilde{C}^{d,N}\tilde{P}_\sigma^{N,\mathcal{J}}(\tilde{C}^{d,N})^T + p_\sigma \tilde{D}^{d,N} Q_{\mathbf{u}} (\tilde{D}^{d,N})^T) \\
\tilde{P}_\sigma^{N,\mathcal{J}+1} &= p_\sigma \sum_{\sigma_1\in\Sigma}\left(\frac{1}{p_{\sigma_1}}\tilde{A}_{\sigma_1}^{d,N}\tilde{P}_{\sigma_1}^{N,\mathcal{J}}(\tilde{A}_{\sigma_1}^{d,N})^T + \tilde{B}_{\sigma_1}^{d,N} Q_{\mathbf{u}} (\tilde{B}_{\sigma_1}^{d,N})^T\right), \quad \tilde{P}_\sigma^{N,0} = 0.
\end{aligned} \tag{5.52}$$

For Step 4 of Algorithm 6 applied to empirical covariances, we define the Markov function $M_{\mathbf{y},\mathbf{u},N,\mathcal{J}}$ as the Markov function $M_{\mathbf{y},\mathbf{u}}$, but with $\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}}$ replaced by $\Lambda_{w,N}^{\mathbf{y}^s,\mathbf{y}^s}$ and $\Lambda_w^{\mathbf{y}^d,\mathbf{u}}$ replaced by $T_{\epsilon,w,N}^{\mathbf{y},\mathbf{u}}$, i.e.,

$$M_{\mathbf{y},\mathbf{u},N,\mathcal{J}}(w) = \begin{cases} \begin{bmatrix} T_{\epsilon,w,N}^{\mathbf{y},\mathbf{u}} Q_{\mathbf{u}}^{-1}, & \Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}} \end{bmatrix} & w \neq \epsilon, \\ \begin{bmatrix} T_{\epsilon,\epsilon,N}^{\mathbf{y}^d,\mathbf{u}} Q_{\mathbf{u}}^{-1}, & I_{n_y} \end{bmatrix} & w = \epsilon. \end{cases} \tag{5.53}$$

For Step 5 of Algorithm 6 applied to empirical covariances, denote by

$$\mathscr{S} = (\{{}^{N,\mathcal{J}}\hat{A}_\sigma, \begin{bmatrix} {}^{N,\mathcal{J}}\hat{B}_\sigma, & {}^{N,\mathcal{J}}\hat{G}_\sigma \end{bmatrix}\}_{\sigma=1}^{n_\mu}, {}^{N,\mathcal{J}}\hat{C}, {}^{N,\mathcal{J}}\hat{D})$$

For Step 6 of Algorithm 6, when applied to empirical covariances, denote by ${}^{N,\mathcal{J}}\hat{K}^{\mathcal{I}} := \hat{K}_\sigma^{\mathcal{I}}$, ${}^{N,\mathcal{J}}\hat{Q}^{\mathcal{I}} := \hat{Q}_\sigma^{\mathcal{I}}$ the matrices as in (5.9) for $\mathcal{I} = 0,\ldots,\bar{\mathcal{I}}$, with $\hat{A}_\sigma, \hat{G}_\sigma, \hat{C}, \hat{D}$, $T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s}$ replaced by ${}^{N,\mathcal{J}}\hat{A}_\sigma$, ${}^{N,\mathcal{J}}\hat{G}_\sigma$, ${}^{N,\mathcal{J}}\hat{C}$, ${}^{N,\mathcal{J}}\hat{D}$, $T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,N\mathcal{J}}$.

Then the result of Algorithm 7 is the tuple

$$(\{\frac{{}^{N,\mathcal{J}}\hat{A}_\sigma}{\sqrt{p_\sigma}}, \frac{{}^{N,\mathcal{J}}\hat{B}_\sigma}{\sqrt{p_\sigma}}, {}^{N,\mathcal{J}}\hat{K}_\sigma^{\bar{\mathcal{I}}}, {}^{N,\mathcal{J}}\hat{Q}_\sigma^{\bar{\mathcal{I}}}\}_{\sigma=1}^{n_\mu}, {}^{N,\mathcal{J}}\hat{C}, {}^{\mathcal{J},N}\hat{D}) =:$$
$$(\{\tilde{A}_\sigma^{N,\mathcal{J}}, \tilde{B}_\sigma^{N,\mathcal{J}}, \tilde{K}_\sigma^{N,\bar{\mathcal{I}},\mathcal{J}}, \tilde{Q}_\sigma^{N,\bar{\mathcal{I}},\mathcal{J}}\}_{\sigma=1}^{n_\mu}, \tilde{C}^{N,\mathcal{J}} \tilde{D}^{N,\mathcal{J}})$$

From the discussion above it is clear that for Algorithm 7 to be well-posed, the following conditions should hold

$$\begin{aligned}&\mathrm{rank}(\mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{y},\mathbf{u},N}}) = \bar{n}\\ &\mathrm{rank}(\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},N\mathcal{J}}}) = n_x,\\ &{}^{N,\mathcal{J}}\hat{Q}^{\mathcal{I}} \text{ is invertable for all } \quad \mathcal{I}=0,\dots,\bar{\mathcal{I}},\ \sigma\in\Sigma\end{aligned} \tag{5.54}$$

We will show that there exists an integer $\mathcal{J}_\star > 0$ such that for every $\mathcal{J} > \mathcal{J}_\star$ there exists an in integer $N_\star(\mathcal{J})$ such that for all $N > N_\star(\mathcal{J})$ the conditions in (5.54) hold.

To this end, we will use the following observation: if $X_{n,l}$ is a sequence of square matrices for every $l = 1,\dots,L$ such that $\lim_{n\to\infty} X_{n,l} = X_l$ and X_l is invertible for $l = 1,\dots,L$, then there exists an integer $n_\star > 0$, such that for all $n > n_\star$, $X_{n,l}$ is invertible for all $l = 1,\dots,L$. This can be seen by noticing that the determinants satisfy $\lim_{n\to\infty}\det(X_{n,l}) = \det(X_l)$ and as $\det(X_l) \neq 0$ it follows that there exists $n_\star$ such that for all $n > n_\star$, for all $l = 1,\dots,L$, $\det(X_{n,l}) \neq 0$.

Note that $\lim_{N\to\infty}\Psi_{\mathbf{y},\mathbf{u},N}(w) = \Psi_{\mathbf{y},\mathbf{u}}(w)$ by (5.33) for all $w\in\Sigma^*$, and hence $\lim_{N\to\infty}\mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{y},\mathbf{u},N}} = \mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{y},\mathbf{u}}}$. Since by assumption $\mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{y},\mathbf{u}}}$ is full rank, there exists $N_{\star,0}$ such that for all $N \geq N_{\star,0}$, $\mathrm{rank}(\mathcal{H}_{\bar{\alpha},\bar{\beta}}^{\Psi_{\mathbf{y},\mathbf{u},N}}) = \bar{n}$, i.e., the first condition of (5.54) holds.

Moreover since the matrices returned by Algorithm 5 are continuous in M, it follows that the matrices $\mathscr{S}_{d,N}$ converge to that of $\mathscr{S}_d$, i.e. ,

$$\lim_{N\to\infty}\tilde{A}_\sigma^{d,N} = \hat{A}_\sigma^d,\quad \lim_{N\to\infty}\tilde{B}_\sigma^{d,N} = \tilde{B}_\sigma^d,\ \lim_{N\to\infty}\tilde{C}^{d,N} = \hat{C}^d,\ \lim_{N\to\infty}\tilde{D}^{d,N} = \tilde{D}^d \tag{5.55}$$

Next, we explore the second equation of (5.54) To this end, consider the Hankel-matrix $\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},N,\mathcal{J}}}$. Since the right-hand side of (5.21) is continuous in the matrices of $\mathcal{S}_d$ from (5.55) it follows that

$$\lim_{N\to\infty}\Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}} = \Lambda_w^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}},\quad \lim_{N\to\infty}T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}} = T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}} \tag{5.56}$$

Since $\lim_{N\to\infty}T_{\epsilon,w,N}^{\mathbf{y},\mathbf{u}} = \Lambda_w^{\mathbf{y},\mathbf{u}}$ and $\lim_{N\to\infty}T_{\epsilon,\epsilon,N}^{\mathbf{y},\mathbf{u}} = \Lambda_\epsilon^{\mathbf{y},\mathbf{u}}$ and by (5.19) $\Lambda_w^{\mathbf{y}^d,\mathbf{u}} = \Lambda_w^{\mathbf{y},\mathbf{u}}$, hence

$$\lim_{N\to\infty}M_{\mathbf{y},\mathbf{u},N,\mathcal{J}}(w) = M_{\mathbf{y},\mathbf{u},\mathcal{J}}(w) \tag{5.57}$$

for all $w\in\Sigma^*$. From Corollary 5.1 it follows that there exists an integer $\mathcal{J}_\star > 0$ such that for all $\mathcal{J} > \mathcal{J}_\star$ equation (5.26) holds.

Since (5.57) implies that

$$\lim_{N\to\infty}\mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},N,\mathcal{J}}} = \mathcal{H}_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},\mathcal{J}}},$$

it follows that for any $\mathcal{J} > \mathcal{J}_\star$ there exists an integer $N_{\star,1}(\mathcal{J}) > 0$ such that for all $N > N_{\star,1}(\mathcal{J})$, it holds that $\mathrm{rank} H_{\alpha,\beta}^{M_{\mathbf{y},\mathbf{u},N,\mathcal{J}}} = n_x$. That is, for $N > \max\{N_{\star,0}, N_{\star,1}(\mathcal{J})\}$ the first to conditions of (5.26) holds.

Moreover, since the matrices returned by Algorithm 5 are continuous in the input Markov-function, it follows

$$\begin{aligned}&\lim_{N\to\infty} {}^{N,\mathcal{J}}\hat{A}_\sigma = {}^{\mathcal{J}}\hat{A}_\sigma, \lim_{N\to\infty} {}^{N,\mathcal{J}}\hat{G}_\sigma = {}^{\mathcal{J}}\hat{G}_\sigma, \quad \lim_{N\to\infty} {}^{N,\mathcal{J}}\hat{B}_\sigma = {}^{\mathcal{J}}\hat{B}_\sigma, \quad \sigma \in \Sigma, \\ &\lim_{N\to\infty} [{}^{N,\mathcal{J}}\hat{C}, {}^{N,\mathcal{J}}\hat{D}] = [{}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D}]\end{aligned} \tag{5.58}$$

By Corollary 5.1, for any $\mathcal{J} > \mathcal{J}_\star$, the matrices ${}^{\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}}$ are invertible for all $\sigma \in \Sigma$ and $\mathcal{I} = 0, \ldots, \bar{\mathcal{I}}$. Recall from (5.48) in the proof of Corollary 5.1 the definition of the function $g_{\bar{\mathcal{I}}}$, and recall that this function is continuous with an open domain of definition. It then follows that

$$g_{\bar{\mathcal{I}}}(\{{}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}\}_{\sigma\in\Sigma}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D}) = \{{}^{\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}}, {}^{\mathcal{J}}\hat{K}_\sigma^{\mathcal{I}}\}_{\sigma\in\Sigma,\mathcal{I}=0,\ldots,\bar{\mathcal{I}}} \tag{5.59}$$

Moreover, since (5.58) and (5.56) hold, it follows that

$$\begin{aligned}\lim_{N\to\infty} (\{{}^{N,\mathcal{J}}\hat{A}_\sigma, {}^{N,\mathcal{J}}\hat{G}_\sigma, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}}\}_{\sigma\in\Sigma}, {}^{N,\mathcal{J}}\hat{C}, {}^{N,\mathcal{J}}\hat{D}) = \\ (\{{}^{\mathcal{J}}\hat{A}_\sigma, {}^{\mathcal{J}}\hat{G}_\sigma, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,\mathcal{J}}\}_{\sigma\in\Sigma}, {}^{\mathcal{J}}\hat{C}, {}^{\mathcal{J}}\hat{D})\end{aligned} \tag{5.60}$$

and as the latter tuple is in the domain of definition of $g_{\bar{\mathcal{I}}}$, which is open, it follows that there exists an integer $N_{\star,2}(\mathcal{J}) > 0$ such that for any $N > N_{\star,2}(\mathcal{J})$, the tuple

$$(\{{}^{N,\mathcal{J}}\hat{A}_\sigma, {}^{N,\mathcal{J}}\hat{G}_\sigma, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}}\}_{\sigma\in\Sigma}, {}^{N,\mathcal{J}}\hat{C}, {}^{N,\mathcal{J}}\hat{D})$$

is in the domain of definition of $g_{\bar{\mathcal{I}}}$ and

$$g_{\bar{\mathcal{I}}}(\{{}^{N,\mathcal{J}}\hat{A}_\sigma, {}^{N,\mathcal{J}}\hat{G}_\sigma, T_{\sigma,\sigma}^{\mathbf{y}^s,\mathbf{y}^s,N,\mathcal{J}}\}_{\sigma\in\Sigma}, {}^{N,\mathcal{J}}\hat{C}, {}^{N,\mathcal{J}}\hat{D}) = \{{}^{N,\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}}, {}^{N,\mathcal{J}}\hat{K}_\sigma^{\mathcal{I}}\}_{\sigma\in\Sigma,\mathcal{I}=0,\ldots,\bar{\mathcal{I}}}$$

and the matrix ${}^{N,\mathcal{J}}\tilde{Q}_\sigma^{\mathcal{I}}$ is invertible for all $\sigma \in \Sigma$, $\mathcal{I} = 0, \ldots, \bar{\mathcal{I}}$. Hence, the last equation of (5.55) holds for $\mathcal{J} > \mathcal{J}_\star$ $N > N_\star(\mathcal{J}) = \max\{N_{\star,0}(\mathcal{J}), N_{\star,1}(\mathcal{J}), N_{\star,2}(\mathcal{J})\}$. Moreover, by continuity of $g_{\bar{\mathcal{I}}}$, (5.59) and (5.60),

$$\lim_{N\to\infty} {}^{N,\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}} = {}^{\mathcal{J}}\hat{Q}_\sigma^{\mathcal{I}}, \quad \lim_{N\to\infty} {}^{N,\mathcal{J}}\hat{K}_\sigma^{\mathcal{I}} = {}^{\mathcal{J}}\hat{K}_\sigma^{\mathcal{I}}, \quad \sigma \in \Sigma \tag{5.61}$$

Combining Corollary 5.1 with (5.58) - (5.61) yields the statement of the theorem. □

Chapter 6

Conclusion

6.1 Summary

This thesis is organized into six chapters, beginning with the introduction in **Chapter 1** and concluding in **Chapter 6**.

Chapter 2 lays the groundwork by presenting the necessary preliminaries essential for the following chapters. In **Chapter 2**, we formally introduce the general model class and its subclasses, along with the technical definitions and properties. These necessary properties characterize the processes involved. In particular, we introduce the notion of:

- Admissible switching processes.
- Zero mean wide sense stationary process with respect to the switching process.
- White noise process with respect to the switching process.
- Square integrable process with respect to the switching process.

All these definitions are necessary in order to define formally the class of stationary Generalized Linear Switched System (GLSS) model. This class of systems is the main focus of the thesis.

Additionally, **Chapter 2** explains the motivation for stochastic realization theory of GLSS, then provides an overview of stochastic realization theory for autonomous stationary GLSS systems, as understanding the autonomous case is crucial for progressing to stochastic realization theory of such systems with inputs.

This overview discusses the existence and minimality of autonomous stationary GLSS in innovation form. First, it presents, formally, a definition for deterministic LSS and the realization theory contributions and algorithms from the literature associated to such systems. Second, it explains the relationship between the stochastic autonomous stationary GLSS and deterministic LSS by recalling a minimization algorithm from previous results from the litrature on realization theory. Each of the following chapters presents a distinct contribution to the realization theory of the systems under study.

In **Chapter 3**, the first theoretical contribution to the GLSS class is introduced. It demonstrates that the output of these systems can be decomposed into deterministic and stochastic components. Using this decomposition, the existence of a GLSS state-space representation in innovation form is proven. Moreover, sufficient conditions are provided to ensure that such representations are minimal and unique up to isomorphism.

Isomorphism implies that if two minimal GLSS models generate the same output for the data used in the identification experiment, they will generate the same outputs for any input and switching signal, including those that do not satisfy the assumptions of the identification experiment. Intuitively, this property should also hold when the two GLSS models produce approximately the same outputs, meaning they do not need to match exactly.

This contribution marks a step forward in stochastic realization theory, with the ultimate goal of developing efficient system identification algorithms. However, to achieve this goal, the autonomous stochastic case—specifically, the stochastic component of the decomposed GLSS—must first be studied. This is the focus of **Chapter 4**.

Chapter 4 delves into the autonomous stochastic GLSS, offering a new algebraic characterization for these systems to be minimal and in innovation form. This characterization relies solely on the system's representation matrices, making it a key contribution to the realization theory of GLSS with inputs and the challenges of system identification. The uniqueness of minimal GLSS in innovation form is crucial, as it enables the proposal of identifiable parameterizations for system identification and guarantees the consistency of system identification algorithms.

It is desirable to determine, based on the system matrices, whether a given GLSS is in innovation form. Furthermore, to ensure that the identified model closely approximates the input-output behavior of the true system, the true system itself must be minimal in innovation form. However, while certain assumptions about the system's matrices may sometimes be available, the statistical properties of the noise are almost always unknown. Unfortunately, the definition of innovation form involves a condition on the noise, adding an additional layer of challenges. This emphasizes even more how important is the algebraic characterization presented in Chapter 4 since it relies solely on the system's representation matrices, bypassing the need for explicit noise assumptions.

Building on the previous chapters, **Chapter 5** presents the final contribution of the thesis, introducing necessary condition for GLSS to be minimal in innovation form (Note that, in contrast to this chapter, **Chapter 3** introduces sufficient conditions for a GLSS to be minimal in innovation form). In addition, it presents a new realization theory algorithm and, subsequently, a new statistically consistent subspace identification algorithm for GLSS models, based on the former algorithm. In addition, Gradient-based search is added to the newly developed identification algorithm. A comparison was made with the latest algorithm presented in the literature in order to study the efficiency in terms of precision and time consumption.

6.2 Future works

In conclusion, this thesis has made a contribution in the stochastic realization theory of generalized linear switched systems and the system identification of such systems. However, it is important to note that there are still many open problems worth exploring. We list some of the possible directions that can be further studied in the future.

6.2.1 Extension to general switching

In control theory, when dealing with subclasses of the studied systems in this thesis, the switching process is, often, controllable. For these cases, the covariance realization and system identification algorithms can be applied. However, in other cases, this is not possible: the switching signal cannot be chosen at will or it may even be non-measurable. This calls for extensions of the results of this thesis to more general switching scenarios. To this end, as the first step, we should extend the results of the thesis to switching processes which are observed but not admissible according to Definition 2.2. One could try applying a change of probability measure to make such switching signals i.i.d. and solve the realization and identification problems over the new probability measure. A similar approach was already used for filtering stochastic systems [28]. We believe that such manipulation can work and we already started investigating the deterministic case. Nevertheless, further theoretical work and practical simulations ought to be done in order to reach the stochastic case.

6.2.2 Extension to time series

The subspace identification algorithm presented in Chapter 5 can learn a GLSS model from a single input-output dataset. However, in practice, one is often bound to work with multiple time series. Treating these time series as a single large dataset when applying the algorithm can be risky, as the different time series are typically uncorrelated—due to discontinuities in data observation, for example—potentially leading to inaccurate output covariance calculations. To address this, we average empirical covariance estimates from individual time series. However, high accuracy demands large datasets, and theoretical guarantees are needed. Therefore, further adaptations should be explored to improve the algorithm's efficiency in estimating models from time series data.

Bibliography

[1] Z. Alkhoury, M. Petreczky, and G. Mercère. Structural properties of affine LPV to LFR transformation: minimality, input-output behavior and identifiability. In *Proceedings of the IEEE Conference on Decision and Control*, Las Vegas, USA, 2016.

[2] Z. Alkhoury, M. Petreczky, and G. Mercère. Identifiability of affine linear parameter-varying models. *Automatica*, 80:62 – 74, 2017.

[3] Ziad Alkhoury. *Minimality, input-output equivalence and identifiability of LPV systems in State-Space and Linear Fractional Representations*. PhD thesis, University of Poitiers, 11 2017.

[4] A. R. P. Andriën and D. J. Antunes. Near-optimal map estimation for markov jump linear systems using relaxed dynamic programming. *IEEE L-CSS*, 4(4):815–820, 2020.

[5] L. Bako, K. Boukharouba, E. Duviella, and S. Lecoeuche. A recursive identification algorithm for switched linear/affine models. *Nonlinear Analysis: Hybrid Systems*, 5(2):242 – 253, 2011.

[6] L. Bako, G. Mercère, and S. Lecoeuche. On-line structured subspace identification with application to switched linear systems. *International Journal of Control*, 82(8):1496–1515, 2009.

[7] L. Bako, G. Mercère, R. Vidal, and S. Lecoeuche. Identification of switched linear state space models without minimum dwell time. In *IFAC Symposium on System Identification, Saint Malo, France*, 2009.

[8] M.P. Balenzuela, A.G Wills, C. Renton, and B. Ninness. Parameter estimation for jump markov linear systems. *Automatica*, 135:109949, 2022.

[9] B. Bamieh and L. Giarré. Identification of linear parameter-varying models. *International Journal of Robust Nonlinear Control*, 12(9):841–853, 2002.

[10] M. Baştuğ, M. Petreczky, R. Tóth, R. Wisniewski, J. Leth, and D. Efimov. Moment matching based model reduction for LPV state-space models. In *Proc. 54th IEEE Conference on Decision and Control*, pages 5334–5339, Osaka, Japan, 2015.

[11] P. Bilingsley. *Probability and measure*. Wiley, 1986.

[12] J. Borges, V. Verdult, M. Verhaegen, and M. Ayala Botto. A switching detection method based on projected subspace classification. In *Conference on Decision and Control-European Control Conference, Seville, Spain*, 2005.

[13] S.P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

[14] P.E. Caines. *Linear Stochastic Systems*. John Wiley and Sons, New-York, 1998.

[15] Alessandro Chiuso. On the relation between cca and predictor-based subspace identification. *IEEE Transactions on Automatic Control*, 52(10):1795–1812, 2007.

[16] E. Cinquemani, R. Porreca, G. Ferrari-Trecate, and J. Lygeros. A general framework for the identification of jump markov linear systems. In *IEEE CDC*, 2007.

[17] M. V. Cook. *Flight dynamics principles: a linear systems approach to aircraft stability and control*. Butterworth-Heinemann, 2012.

[18] O. L. V. Costa, M. D. Fragoso, and R. P. Marques. *Discrete-time Markov jump linear systems*. Springer Science & Business Media, 2006.

[19] P. Cox, M. Petreczky, and R. Tóth. Towards efficient maximum likelihood estimation of LPV-SS models. *Automatica*, 97(9):392–403, 2018.

[20] P. B. Cox. *Towards efficient identification of linear parameter-varying state-space models*. PhD thesis, Eindhoven Univ. of Technology, 2018.

[21] P. B. Cox and R. Tóth. Linear parameter-varying subspace identification: A unified framework. *Automatica*, 123:109296, 2021.

[22] Willem L De Koning. Digital optimal reduced-order control of pulse-width-modulated switched linear systems. *Automatica*, 39(11):1997–2003, 2003.

[23] B. De Schutter. Minimal state-space realization in linear system theory: an overview. *Journal of Computational and Applied Mathematics*, 121(1 -2):331–354, 2000.

[24] P. den Boef, P. B. Cox, and R. Tóth. LPVcore: MATLAB toolbox for LPV modelling, identification and control. *IFAC SYSID*, 2021.

[25] Mohamed Djemai and Michael Defoort. Hybrid dynamical systems. *Lecture notes in Control and information sciences*, 457, 2015.

[26] P. L. dos Santos, J. A. Ramos, and J. L. M. de Carvalho. Identification of LPV systems using successive approximations. In *Proc. of 47th IEEE Conference on Decision and Control*, pages 4509–4515, 2008.

[27] P.L. dos Santos, J.A. Ramos, and J.L.M. de Carvalho. Identification of bilinear systems with white noise inputs: An iterative deterministic-stochastic subspace approach. *IEEE Transactions on Control Systems Technology*, 17(5):1145–1153, Sept 2009.

[28] Robert J Elliott, Lakhdar Aggoun, and John B Moore. *Hidden Markov models: estimation and control*, volume 29. Springer Science & Business Media, 2008.

[29] W. Favoreel, B. De Moor, and P. Van Overschee. Subspace identification of bilinear systems subject to white inputs. *IEEE Transactions on Automatic Control*, 44(6):1157–1165, Jun 1999.

[30] F. Felici, J. W. Van Wingerden, and M. Verhaegen. Subspace identification of MIMO LPV systems using a periodic scheduling sequence. *Automatica*, 43:1684–1697, 2007.

[31] Ian Fialho and Gary J Balas. Road adaptive active suspension design using linear parameter-varying gain-scheduling. *IEEE transactions on control systems technology*, 10(1):43–54, 2002.

[32] M. Fliess. Realizations of nonlinear systems and abstract transitive lie-algebras. *Bulletin of Americal Mathematical Society*, 2(3), 1980.

[33] A. Garulli, S. Paoletti, and A. Vicino. A survey on switched and piecewise affine system identification. In *Symposuim on System Identification, SYSID*, pages 344–355, 2012.

[34] F. Gécseg and I Peák. *Algebraic theory of automata*. Akadémiai Kiadó, Budapest, 1972.

[35] Zoubin Ghahramani and Geoffrey E Hinton. Parameter estimation for linear dynamical systems. *A Technical Report*, 1996.

[36] L. Giarré, D. Bauso, P. Falugi, and B. Bamieh. Lpv model identification for gain scheduling control: An application to rotating stall and surge control problem. *Control Engineering Practice*, 14(4):351–361, 2006.

[37] Stuart Gibson and Brett Ninness. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41(10):1667–1682, 2005.

[38] K. Glover and J. C. Willems. Parametrizations of linear dynamical systems: canonical forms and identifiability. *IEEE Transactions on Automatic Control*, 19:640–646, 1974.

[39] R.L. Grossman and R.G. Larson. An algebraic approach to hybrid systems. *Theoretical Computer Science*, 138:101–112, 1995.

[40] B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems*. Stichting Mathematische Centrum, Amsterdam, 1989.

[41] Christopher Heil. *A Basis Theory Primer: Expanded Edition*. Birkhäuser Boston, 2011.

[42] A. Isidori. Direct construction of minimal bilinear realizations from nonlinear input-output maps. *IEEE Transactions on Automatic Control*, pages 626–631, 1973.

[43] A. Isidori. *Nonlinear Control Systems*. Springer Verlag, 1989.

[44] A. Isidori, P. D'Alessandro, and A. Ruberti. Realization and structure theory of bilinear dynamical systems. *SIAM J. Control*, 12(3), 1974.

[45] B. Jakubczyk. Existence and uniqueness of realizations of nonlinear systems. *SIAM J. Control and Optimization*, 18(4):455 – 471, 1980.

[46] B. Jakubczyk. Construction of formal and analytic realizations of nonlinear systems. In Isodori and Hinrichsen, editors, *Feedback Control of Linear and Nonlinear Systems*, volume 39 of *Lecture Notes in Control and Information Sciences*, pages 147 – 156, 1982.

[47] B. Jakubczyk. Realization theory for nonlinear systems, three approaches. In *Algeb. & Geom. Methods in Nonlin. Contr. Theory*, 1986.

[48] Magnus Jansson and Bo Wahlberg. A linear regression approach to state-space subspace system identification. *Signal Processing*, 52(2):103–129, 1996. Subspace Methods, Part II: System Identification.

[49] Xing Jin, Biao Huang, and David S Shook. Multiple model lpv approach to nonlinear process identification with em algorithm. *Journal of Process Control*, 21(1):182–193, 2011.

[50] Mikael K-J Johansson. *Piecewise linear control systems: a computational approach*, volume 284. Springer, 2003.

[51] W. P. BLAIR Jr. and D. D. SWORDER. Feedback control of a class of linear discrete systems with jump parameters and quadratic cost criteria †. *International Journal of Control*, 21(5):833–841, 1975.

[52] T. Kailath. *Linear Systems*. Prentice Hall, Engelwood Cliffs, 1980.

[53] R. E. Kalman. Advanced theory of linear systems. In *Topics in Mathematical System Theory*, pages 237–339. McGraw-Hill, New York, 1969.

[54] R. E. Kalman. Introduction to the algebraic theory of linear dynamical systems. In *Mathematical systems theory and economics, I, II (Proc. Internat. Summer School, Varenna, 1967)*, pages 41–65. Lecture Notes in Oper. Res. and Math. Econom., Vols. 11, 12. Springer, Berlin, 1969.

[55] R. E. Kalman, P. L. Falb, and M. A. Arbib. *Topics in mathematical system theory*. McGraw-Hill Book Co., New York, 1969.

[56] T. Katayama. *Subspace Methods for System Identification*. Springer-Verlag, 2005.

[57] Johan Kon, Jeroen van de Wijdeven, Dennis Bruijnen, Roland Tóth, Marcel Heertjes, and Tom Oomen. Direct learning for parameter-varying feedforward control: A neural-network approach. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 3720–3725, 2023.

[58] N. Kulcsár and Tóth R. On the similarity state transformation for linear parameter-varying systems. In *Proc. 18th IFAC World Congress*, 2011.

[59] W. E. Larimore. System identification, reduced-order filtering and modeling via canonical variate analysis. In *American Control Conference*, 1983.

[60] Wallace E Larimore. Identification of nonlinear parameter-varying systems via canonical variate analysis. In *2013 american control conference*, pages 2247–2262. IEEE, 2013.

[61] W.E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *29th IEEE Conference on Decision and Control*, pages 596–604 vol.2, 1990.

[62] F. Lauer and G. Bloch. *Hybrid System Identification: Theory and Algorithms for Learning Switching Models*, volume 478 of *Lecture Notes in Control and Information Sciences*. Springer International Publishing, 2019.

[63] Lawton H Lee and Kameshwar Poolla. Identifiability issues for parameter-varying and multidimensional linear systems. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80401, page V01AT03A012. American Society of Mechanical Engineers, 1997.

[64] Lawton H. Lee and Kameshwar Poolla. Identification of Linear Parameter-Varying Systems Using Nonlinear Programming. *Journal of Dynamic Systems, Measurement, and Control*, 121(1):71–78, 03 1999.

[65] A. Lindquist and G. Picci. *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*. Springe Berlin, 2015.

[66] L. Ljung. *System Identification: Theory for the user (2nd Ed.)*. PTR Prentice Hall., Upper Saddle River, USA, 1999.

[67] Y. Ma and R. Vidal. A closed form solution to the identification of hybrid ARX models via the identification of algebraic varieties. In *Hybrid Systems: Computation and Control*, pages 449–465, 2005.

[68] Yi Ma and R. Vidal. Identification of deterministic switched arx systems via identification of algebraic varieties. In *Hybrid Systems: Computation and Control*, volume 3414 of *Lecture Notes in Computer Science*, pages 449 – 465, 2005.

[69] A. Marcos and G.J. Balas. Development of linear-parameter-varying models for aircraft. *Journal of Guidance Control and Dynamics*, 27(2):218–228, 2004.

[70] T. McKelvey, A. Helmersson, and T. Ribarits. Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica*, 40:1629–1635, 2004.

[71] Tomas McKelvey, Anders Helmersson, and Thomas Ribarits. Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica*, 40(9):1629–1635, 2004.

[72] M. Mejari and M. Petreczky. Realization and identification algorithm for stochastic LPV state-space models with exogenous inputs. In *3rd IFAC Workshop on Linear Parameter-Varying Systems*, Eindhoven, Netherlands, 2019.

[73] M. Mejari and M. Petreczky. Realization and identification algorithm for stochastic LPV state-space models with exogenous inputs. *arXiv 1905.10113*, 2019.

[74] Manas Mejari and Mihaly Petreczky. Consistent and computationally efficient estimation for stochastic lpv state-space models: realization based approach. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3805–3810. IEEE, 2019.

[75] Javad Mohammadpour and Carsten W Scherer. *Control of linear parameter varying systems with applications*. Springer Science & Business Media, 2012.

[76] G. Papageorgiou, K. Glover, G. D'Mello, and Y. Patel. Taking robust LPV control into flight on the VAAC harrier. In *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187)*, volume 5, pages 4558–4564 vol.5, 2000.

[77] R. L. M. Peeters. *System Identification Based on Riemannian Geometry: Theory and Algorithms*. PhD thesis, Free University, Amsterdam, 1994.

[78] M. Petreczky. *Realization Theory of Hybrid Systems*. PhD thesis, Vrije Universiteit, Amsterdam, 2006.

[79] M. Petreczky. Realization theory of linear and bilinear switched systems: A formal power series approach: Part i. *ESAIM Control, Optimization and Calculus of Variations*, 17:410–445, 2011.

[80] M. Petreczky. Realization theory of linear and bilinear switched systems: A formal power series approach: Part ii. *ESAIM Control, Optimization and Calculus of Variations*, 17:446–471, 2011.

[81] M. Petreczky. Realization theory of linear hybrid systems. In *Hybrid Dynamical Systems: Observation and control*, Lecture Notes in Control and Information Sciences. Springer-Verlag, 2015.

[82] M. Petreczky and L. Bako. On the notion of persistence of excitation for linear switched systems. *Nonlinear Analysis: Hybrid Systems*, 48:101308, 2023.

[83] M. Petreczky, L. Bako, and van J.H. Schuppen. Realization theory for discrete-time linear switched systems. *Automatica*, 49(11):3337–3344, 2013.

[84] M. Petreczky, L. Bako, and J.H. van Schuppen. Identifiability of discrete-time linear switched systems. In *Hybrid Systems: Computation and Control*, pages 141–150. ACM, 2010.

[85] M. Petreczky and G. Mercere. Affine LPV systems: Realization theory, input-output equations and relationship with linear switched systems. In *51th IEEE Conference on Decision and Control (CDC),*, pages 4511–4516, 2012.

[86] M. Petreczky, R. Tóth, and G. Mercére. Realization theory for LPV state-space representations with affine dependence. *IEEE Transactions on Automatic Control*, 62(9):4667–4674, 2017.

[87] M. Petreczky and J. H. Van Schuppen. Partial-realization of linear switched systems: A formal power series approach. *Automatica*, 47(10):2177–2184, 2011.

[88] M. Petreczky and J.H. van Schuppen. Realization theory for linear hybrid systems. *IEEE Trans. on Automatic Control*, 55:2282 – 2297, 2010.

[89] M. Petreczky and J.H. van Schuppen. Span-reachability and observability of bilinear hybrid systems. *Automatica*, 46:501–509, 2010.

[90] M. Petreczky and R. Vidal. Realization theory of stochastic jump-markov linear systems. In *Proc. 46th IEEE Conference on Decision and Control*, 2007.

[91] M. Petreczky and R. Vidal. Realization theory for a class of stochastic bilinear systems. *IEEE Transactions on Automatic Control*, 63(1):69–84, 2018.

[92] M. Petreczky, R. Wisniewski, and J. Leth. Balanced truncation for linear switched systems. *Nonlinear Analysis: Hybrid Systems*, 10:4–20, November 2013.

[93] Mihály Petreczky, Roland Tóth, and Guillaume Mercère. LPV-ARX representations of LPV state-space models with affine dependence. *Systems & Control Letters*, 173:105459, 2023.

[94] Giorgio Picci and Tohru Katayama. Stochastic realization with exogenous inputs and 'subspace-methods' identification. *Signal Processing*, 52(2):145–160, 1996.

[95] D. Piga, P. Cox, R. Tóth, and V. Laurain. LPV system identification under noise corrupted scheduling and output signal observations. *Automatica*, 53:329–338, 2015.

[96] T. Ribarits. *The role of parametrizations in identification of linear dynamic systems*. PhD thesis, Vienna University of Technology, Vienna, Austria, 2002.

[97] Syed Z. Rizvi, Javad Mohammadpour, Roland Tóth, and Nader Meskin. An iv-svm-based approach for identification of state-space lpv models under generic noise conditions. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 7380–7385, 2015.

[98] Syed Zeeshan Rizvi, Javad Mohammadpour Velni, Farshid Abbasi, Roland Tóth, and Nader Meskin. State-space lpv model identification using kernelized machine learning. *Automatica*, 88:38–47, 2018.

[99] E. Rouphael, M. Petreczky, and L. Belkoura. On minimal LPV state-space represenations in innovation form: an algebraic characterization. *the 61st IEEE Conference on Decision and Control*, 2022.

[100] Elie Rouphael, Manas Mejari, Mihaly Petreczky, and Lotfi Belkoura. Minimal covariance realization and system identification algorithm for a class of stochastic linear switched systems with i.i.d. switching. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pages 6355–6362, 2024.

[101] Elie Rouphael, Manas Mejari, Mihaly Petreczky, and Lotfi Belkoura. Toward stochastic realization theory for generalized linear switched systems with inputs: Decomposition into stochastic and deterministic components and existence and uniqueness of innovation form. *IEEE Control Systems Letters*, 8:1841–1846, 2024.

[102] T. Sarkar, A. Rakhlin, and M. Dahleh. Nonparametric system identification of stochastic switched linear systems. In *IEEE CDC*, 2019.

[103] J. S. Shamma and M. Athans. Analysis of gain scheduled control for nonlinear plants. *IEEE Transactions on Automatic Control*, 35(8):898–907, 1990.

[104] Jeff S Shamma. *Analysis and design of gain scheduled control systems*. PhD thesis, Massachusetts Institute of Technology, 1988.

[105] Jeff S Shamma and Michael Athans. Guaranteed properties of gain scheduled control for linear parameter-varying plants. *Automatica*, 27(3):559–564, 1991.

[106] J.S. Shamma and M. Athans. Gain scheduling: potential hazards and possible remedies. *IEEE Control Systems Magazine*, 12(3):101–107, 1992.

[107] E. D. Sontag. *Polynomial Response Maps*, volume 13 of *Lecture Notes in Control and Information Sciences*. Springer Verlag, 1979.

[108] Eduardo D. Sontag. Realization theory of discrete-time nonlinear systems: Part I – the bounded case. *IEEE Transactions on Circuits and Systems*, 26(5):342–356, 1979.

[109] Brian L Stevens, Frank L Lewis, and Eric N Johnson. *Aircraft control and simulation: dynamics, controls design, and autonomous systems*. John Wiley & Sons, 2015.

[110] Zh. Sun and Sh. S. Ge. *Switched linear systems : control and design*. Springer, London, 2005.

[111] H. Sussman. Existence and uniqueness of minimal realizations of nonlinear systems. *Mathematical Systems Theory*, 10:263–284, 1977.

[112] D. Sworder and R. Rogers. An lq-solution to a control problem associated with a solar thermal central receiver. *IEEE Transactions on Automatic Control*, 28(10):971–978, 1983.

[113] R. Tóth. Modeling and identification of linear parameter-varying systems. In *Lecture Notes in Control and Information Sciences, Vol. 403*, Heidelberg, 2010. Springer.

[114] R. Tóth, H.S. Abbas, and H. Werner. On the state-space realization of LPV input-output models: Practical approaches. *IEEE Transactions on Control Systems Technology*, 20(1), 2012.

[115] R. Tóth, J. C. Willems, P. S. C. Heuberger, and P. M. J. Van den Hof. The behavioral approach to linear parameter-varying systems. *IEEE Transactions on Automatic Control*, 56(7):2499–2514, 2011.

[116] J. van den Hof. Structural identifiability of linear compartmental systems. *IEEE Transactions on Automatic Control*, 43:800–818, 1998.

[117] A. van der Schaft and H. Schumacher. *An Introduction to Hybrid Dynamical Systems*. Springer-Verlag London, 2000.

[118] P. van Overschee and B. De Moor. *Subspace Identification for Linear Systems*. Kluwer Academic Publishers, 1996.

[119] Peter Van Overschee and Bart De Moor. N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994. Special issue on statistical signal processing and control.

[120] C. van Putten and J. H. van Schuppen. Invariance properties of the conditional independence relation. *Ann. Probab.*, 13(3):934–945, 08 1985.

[121] Jan H van Schuppen. *Control and system theory of discrete-time stochastic systems*. Springer, 2021.

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

[123] Jan-Willem van Wingerden, Federico Felici, and Michel Verhaegen. Subspace identification of mimo lpv systems using a piecewise constant scheduling sequence with hard/soft switching. In *2007 European Control Conference (ECC)*, pages 927–934. IEEE, 2007.

[124] Bálint Vanek, Gary J Balas, and Roger EA Arndt. Linear, parameter-varying control of a supercavitating vehicle. *Control Engineering Practice*, 18(9):1003–1012, 2010.

[125] V. Verdult and M. Verhaegen. Subspace identification of multivariable linear parameter-varying systems. *Automatica*, 38(5):805–814, 2002.

[126] V. Verdult and M. Verhaegen. Subspace identification of piecewise linear systems. In *43rd IEEE Conf. on Decision and Control*, 2004.

[127] V. Verdult and M. Verhaegen. Kernel methods for subspace identification of multivariable LPV and bilinear systems. *Automatica*, 41:1557–1565, 2005.

[128] Vincent Verdult, Niek Bergboer, and Michel Verhaegen. Identification of fully parameterized linear and nonlinear state-space systems by projected gradient search. *IFAC Proceedings Volumes*, 36(16):711–716, 2003.

[129] Michel Verhaegen and Patrick Dewilde. Subspace model identification part 1. the output-error state-space model identification class of algorithms. *International Journal of Control*, 56(5):1187–1210, 1992.

[130] R. Vidal. Recursive identification of switched ARX systems. *Automatica*, 44(9):2274 – 2287, 2008.

[131] Mark W. Watson and Robert F. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400, 1983.

[132] Xiukun Wei. *Advanced LPV Techniques for Diesel Engines*. PhD thesis, Universität Linz, 2006.

[133] Xiukun Wei. On persistent excitation for parameter estimation of quasi-lpv systems and its application in modeli. *IFAC Proceedings Volumes*, 39:517–522, 03 2006.

[134] S. Weiland, A. Lj. Juloski, and B. Vet. On the equivalence of switched affine models and switched arx models. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 2614–2618, 2006.

[135] J. C. Willems. On interconnections, control, and feedback. *IEEE Transactions on Automatic Control*, 42(3):326 – 339, 1997.

[136] J. C. Willems. The behavioral approach to open and interconnected systems. *IEEE Control Systems Magazine*, 27:46–99, 2007.

[137] J.C. Willems and J.W. Polderman. *An Introduction to Mathematical Systems Theory: A Behavioral Approach*. Springer Verlag, New York, 1998.

[138] A. Wills and B. Ninness. *System identification of linear parameter varying state-space models*, pages 295–315. World Scientific, 2011.

[139] Adrian Wills and Brett Ninness. On gradient-based search for multivariable system estimates. *IEEE Transactions on Automatic Control*, 53(1):298–306, 2008.

[140] Xudong Zhao, Yonggui Kao, Ben Niu, and Tingting Wu. *Control synthesis of switched systems*. Springer, 2017.