

# Optimal Sampling and Interpolation



**Hanumant Singh Shekhawat**

# **OPTIMAL SAMPLING AND INTERPOLATION**

**Hanumant Singh Shekhawat**

## Composition of the Graduation Committee:

Chairman and Secretary:

Prof. Dr. Ir. A.J. Mouthaan    University of Twente, EWI

Promotor and Assistant Promotor:

Prof. Dr. A.A. Stoorvogel    University of Twente, EWI

Dr. Ir. G. Meinsma            University of Twente, EWI

Members:

Prof. Dr. A. Bagchi            University of Twente, EWI

Dr. Ir. R.N.J. Veldhuis        University of Twente, EWI

Prof. Dr. S. Weiland         Eindhoven University of Technology

Dr. L. Mirkin                  Technion IIT

Prof. Dr. H.J. Zwart          Eindhoven University of Technology

**UNIVERSITEIT TWENTE.**

The research described in this thesis was performed at the Hybrid Systems group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands.

**disc**

This dissertation has been completed in partial fulfillment of the requirements of the Dutch Institute of Systems and Control (disc) for graduate study.

Cover Image: Flag positions in the Semaphore flag signaling system.

The image has been reproduced with permission of Jim Croft, Campbell, Australia.

Title:        Optimal Sampling and Interpolation

Author:     Hanumant Singh Shekhawat

ISBN:       978-90-365-3473-4

DOI:        <http://dx.doi.org/10.3990/1.9789036534734>

Copyright © 2012 by Hanumant Singh Shekhawat, Enschede, The Netherlands.  
All rights reserved. No part of this publication may be reproduced by print, photocopy or any other means without the prior written permission from the copyright owner.

Printed by Ipskamp Drukkers, Enschede, The Netherlands.

# **OPTIMAL SAMPLING AND INTERPOLATION**

## **DISSERTATION**

to obtain  
the degree of doctor at the University of Twente,  
on the authority of the rector magnificus,  
Prof. Dr. H. Brinksma,  
on account of the decision of the graduation committee,  
to be publicly defended on  
Thursday 29th of November 2012 at 14:45 hours

by

**Hanumant Singh Shekhawat**

born on 6 December 1979

in Kanpur, India.

This thesis has been approved by:

Promotor: Prof. Dr. A.A. Stoorvogel

Assistant Promotor: Dr. Ir. G. Meinsma

*This dissertation is dedicated to my family*



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Problem formulation . . . . .	4
1.3	Overview of the thesis . . . . .	6
1.4	Overview of contributions . . . . .	7
<b>2</b>	<b>Sampled-data system theory</b>	<b>9</b>
2.1	Notation . . . . .	9
2.2	Sampled-data system . . . . .	11
2.2.1	Signal Generator . . . . .	11
2.2.2	Sampler . . . . .	12
2.2.3	Hold . . . . .	13
2.2.4	Discrete system . . . . .	13
2.3	Lifting . . . . .	14
2.3.1	Lifting in time domain . . . . .	14
2.3.2	Lifting in frequency domain . . . . .	17
2.4	Lifted domain spaces and norms . . . . .	21
2.4.1	Lifted signal space and norm . . . . .	21
2.4.2	Adjoint systems and conjugate transfer function . . . . .	24
2.4.3	System norms . . . . .	25
2.5	Causality . . . . .	27
2.6	Stability . . . . .	33
2.6.1	Stability with causality . . . . .	34
2.7	Concluding remarks . . . . .	34
2.A	Proofs . . . . .	34
<b>3</b>	<b>Frequency truncated norms</b>	<b>39</b>
3.1	Introduction . . . . .	39
3.2	A generic version . . . . .	42
3.2.1	Matrix functions and principal logarithm . . . . .	43
3.2.2	Integration of a transfer function . . . . .	47
3.2.3	Finite integral for proper $K(s)$ . . . . .	49
3.3	Computation of frequency truncated $L^2$ -norm . . . . .	50



3.3.1	Stable and causal $G(s)$ . . . . .	50
3.3.2	Finite integral for $G(s)$ . . . . .	52
3.4	Conclusions . . . . .	53
3.A	Proofs . . . . .	53
<b>4</b>	<b>Non-causal downsampling</b> . . . . .	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Notations . . . . .	59
4.3	Sampled data setup for downsampling . . . . .	60
4.4	Discrete lifting in time domain . . . . .	62
4.5	Discrete lifting in frequency domain . . . . .	65
4.5.1	Transfer function in lifted frequency domain . . . . .	68
4.6	Signal and System norms . . . . .	69
4.6.1	System norms . . . . .	70
4.7	Rank theorem . . . . .	70
4.8	Non-causal downsampling problem . . . . .	72
4.9	$L^2$ optimal downsampling . . . . .	74
4.9.1	Preliminaries . . . . .	75
4.9.2	$L^2$ optimal downsampling solution . . . . .	77
4.9.3	$L^2$ Error Norm . . . . .	82
4.10	$L^\infty$ optimal downsampling . . . . .	83
4.10.1	Preliminaries . . . . .	83
4.10.2	$L^\infty$ optimal downsampling solution . . . . .	87
4.11	Downsampling in the presence of noise . . . . .	91
4.11.1	$L^2$ optimal downsampling in the presence of noise . . . . .	94
4.11.2	$L^\infty$ optimal downsampling in the presence of noise . . . . .	97
4.12	Concluding Remarks . . . . .	100
4.A	Proofs of the results in sections 4.5, 4.7 and 4.8 . . . . .	100
4.B	Proofs of the results in Section 4.9 . . . . .	104
4.C	Proofs of the results in Section 4.10 . . . . .	107
4.D	Proofs of the results in Section 4.11 . . . . .	113
<b>5</b>	<b>Relaxed causal sampling</b> . . . . .	<b>117</b>
5.1	Introduction . . . . .	117
5.2	Problem formulation and solution . . . . .	118
5.3	State-space of linear $h$ -time shift invariant systems . . . . .	121
5.3.1	State-space with two point boundary condition (STPBC) . . . . .	124
5.3.2	Systems in STPBC . . . . .	126
5.3.3	Stability and causality of systems given in STPBC . . . . .	133
5.3.4	$H^2$ norm of systems given in STPBC . . . . .	135
5.3.5	Computations . . . . .	138
5.4	STPBC solution . . . . .	140
5.4.1	Left coprime-factorization of $\check{G}$ . . . . .	142
5.4.2	Simplification of Assumption $\mathcal{A}_3$ . . . . .	143
5.4.3	Inner-outer factorization of hold . . . . .	144

5.4.4	The condition $(I - \hat{H}_i \hat{H}_i^\sim) \check{G}_v \in L^\infty$ . . . . .	146
5.4.5	The condition $(I - \hat{H}_i \hat{H}_i^\sim) \check{G}_v \in L^2$ . . . . .	155
5.4.6	Obtaining $\check{V}$ . . . . .	155
5.4.7	Optimal relaxed causal sampler . . . . .	158
5.4.8	Optimal $H^2$ norm . . . . .	158
5.4.9	Example . . . . .	163
5.5	Conclusions . . . . .	166
5.A	Proofs of the results in Section 5.2 . . . . .	167
5.B	Proofs of the results in Section 5.3 . . . . .	170
5.C	Proofs of the results in Section 5.4 . . . . .	176
<b>6</b>	<b>Conclusions and Recommendations</b> . . . . .	<b>195</b>
6.1	Recommendations . . . . .	196
<b>A</b>	<b>State-space representation</b> . . . . .	<b>197</b>
A.1	Continuous time system . . . . .	197
A.2	Discrete time system . . . . .	198
	<b>List of notations</b> . . . . .	<b>199</b>
	<b>Bibliography</b> . . . . .	<b>203</b>
	<b>Summary</b> . . . . .	<b>209</b>
	<b>Samenvatting</b> . . . . .	<b>211</b>
	<b>Acknowledgments</b> . . . . .	<b>213</b>
	<b>Index</b> . . . . .	<b>215</b>



# Chapter 1

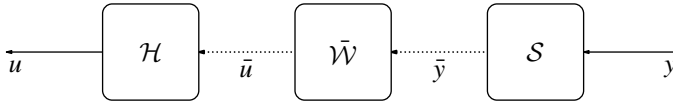
## Introduction

This thesis is a system theoretical exploration of optimal samplers, downsamplers and interpolators (holds).

### 1.1 Motivation

Message or signal communication, storage and analysis are some of the oldest needs of a society. A signal or message is the information that needs to be processed (i.e transmitted or stored etc.) over time or space. If a signal varies continuously with time/space then such a signal is called an *analog (or continuous time) signal*. Most of the signals (e.g. voice, seismic data) are analog in nature. On the other hand if information is just coming at discrete time/space instants then such a signal is called a *discrete signal*. If a discrete signal can have values from a finite set, then the discrete signal is called a *digital signal*. A *system* is a fundamental part of the signal processing. It is a device that processes a signal to give a desired output. If a system processes an analog signal entirely in the analog domain then the system is an *analog system (or continuous-time system)* and the whole process is called *analog signal processing*. On the other hand if a system processes a digital (or discrete) signal entirely in the digital (or discrete) domain then the system is a *discrete system* and the whole process is known as *digital (or discrete) signal processing*. If a system processes signal in a mixture of digital/discrete and analog domain then it is a *hybrid system*.

The use of electrical signals for message transmission in the nineteenth century increased the efficiency of the signal processing. At that time most of the systems were analog. However, in the mid twentieth century with the advent of modern integrated circuits the cost of digital signal processing (DSP) reduced significantly. Even though we lose information in the analog to digital conversion, DSP provides better quality, ease of implementation, reconfigurability, ease of storage, lower cost, etc. As a result, DSP started to replace analog signal processing in most of the applications. This is the trend till now.



**Figure 1.1:** Signal reconstruction setup

The fundamental problem in DSP is the signal reconstruction problem. Here the main aim is to recover an analog signal from its samples with minimal error. The signal reconstruction setup is shown in Figure 1.1. Here a sampler  $\mathcal{S}$  samples an analog signal  $y$  to produce the discrete signal  $\bar{y}$ . The discrete signal  $\bar{y}$  is processed by a discrete system  $\bar{\mathcal{W}}$ . The discrete system can do various jobs on the discrete signals like filtering noise etc. The output  $\bar{u}$  of the discrete system is converted back to an analog signal by a hold (interpolator or  $D/A$  converter)  $\mathcal{H}$ . The main aim in the signal reconstruction problem is to make the reconstructed signal  $u$  as close as possible to the analog signal  $y$ . Normally digital signals are quantized after sampling. However, for simplicity the quantization errors are not taken into account in the signal reconstruction problem that we consider in this thesis.

A sampling operation typically means loss of information. Therefore, an interesting question is whether it is possible to reconstruct the original analog signal  $y$  exactly from its samples by a suitable choice of hold and the discrete system. One famous answer to the above question is Shannon's theorem and it depends upon the sampling period and the bandwidth of the original analog signal  $y$ :

**Theorem 1.1.1** (see [49, 60, 42]). *Let  $y(t)$  be a signal whose Fourier transform  $Y(j\omega)$  exists. If  $y(t)$  is bandlimited to  $\omega_B$  rad/sec, i.e.  $Y(j\omega) = 0 \forall |\omega| \geq \omega_B$  then*

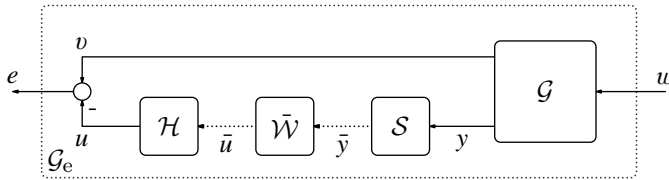
$$y(t) = \sum_{k \in \mathbb{Z}} y\left(k \frac{\pi}{\omega_B}\right) \operatorname{sinc}\left(\frac{\omega_B}{\pi} t - k\right)$$

where  $\operatorname{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ .

Assume that in the signal reconstruction problem our analog signal  $y$  is ideally sampled with sampling period  $h$  i.e.  $\bar{y}[k] = y(kh)$  and that  $y$  is bandlimited to  $\omega_N := \frac{\pi}{h}$ . The frequency  $\omega_N$  is known as the *Nyquist frequency* associated with the sampling period  $h$ . Using Shannon's theorem it is straightforward to prove that the discrete system  $\bar{\mathcal{W}} = I$  and the following hold

$$u = \mathcal{H}\bar{u} : \quad u(t) = \sum_{k \in \mathbb{Z}} \bar{u}(kh) \operatorname{sinc}\left(\frac{t}{h} - k\right) \quad (1.1)$$

leads to perfect reconstruction i.e.  $u = y$ . The selection of the discrete system  $\bar{\mathcal{W}} = I$  is arbitrary because it can be shown that the discrete filter can be embedded in either the hold or sampler (see [31] or Section 2.2.4). Therefore, in the rest of the chapter we skip this discrete system.



**Figure 1.2:** Sampled-data setup

The Shannon theorem works perfectly for bandlimited signals but most of the signals in nature are not bandlimited. Sampling of these type of signals often results in *aliasing*. A signal is aliased to another signal if ideal sampling of these two signals result in indistinguishable discrete signals. To circumvent this problem analog signals before sampling are passed through a low-pass filter that is bandlimited to the Nyquist frequency. This will lead to exact reconstruction, however at the cost of throwing away a lot of information available in the frequency bands that are filtered away. This leads to the following interesting questions.

- Is the use of a low-pass filter bandlimited to the Nyquist frequency optimal?
- If not, then what is the optimal way to do sampling and interpolation?
- How much of the information is lost in sampling and interpolation?
- What is the theoretical minimum of information lost in the sampling operation?

To answer these questions in general, researchers started looking at these problems as mathematical optimization problems from a system theoretical viewpoint (see Sun et al. [55] and Unser [59]). The Sampled data system theory is a system theoretical method that treats discrete and analog signals in a common framework.

This theory was first applied in the signal reconstruction problem in 1996 by Khargonekar and Yamamoto [23] (in 1995, Chen and Francis [56] applied the sampled-data system theory to the signal reconstruction problem entirely in the discrete domain). Instead of aiming at exact reconstruction as in the Shannon case, minimization of the error without throwing away any frequencies is the main criterion in the signal reconstruction using sampled-data system theory. Khargonekar and Yamamoto [23] used a sampled-data setup similar to the setup shown below. The distinctive feature of the sampled-data setup shown in Figure 1.2 is that it optimizes the analog performance. This setup is much closer to reality as most of the signals we use are analog in nature and utilized in the analog domain.

In the sampled-data setup shown in Figure 1.2, a *signal model* or *signal generator*  $\mathcal{G}$  is used to represent the information that we know about our analog signal  $y$  and  $v$ . This is another distinctive feature of sampled-data system theory. For example, if we are processing audible signals then we know that the spectrum of these signals lies in between 20 Hertz to 20 kilo Hertz. Then  $\mathcal{G}$  can be a band-pass filter with passband 20 Hertz to 20 kilo Hertz. Moreover, if we know that

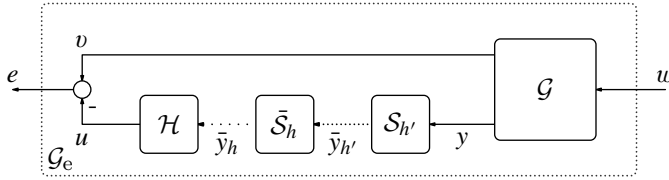
the signals we are going to process is human speech then 250 Hertz to 4000 Hertz bandpass filter is sufficient. Key point here is that the reconstruction performance increases if  $\mathcal{G}$  contains more information about the signals  $y$  and  $v$  (see [7] for more detail). In the sampled-data setup, we sample our original signal  $y$  using a sampler  $\mathcal{S}$ . The resulting signal after it is processed by discrete system  $\mathcal{W}$  is reconstructed back to the analog domain by an hold  $\mathcal{H}$ . This reconstructed signal  $u$  is compared with signal  $v$  to compute error  $e := v - u$ . In most of the applications  $v$  is the same as  $y$  but for a generic treatment of signal reconstruction problem  $v$  may be taken different from  $y$ . Throughout this thesis we assume that the signal model  $\mathcal{G}$  is given. The Shannon case is a special case of the sampled-data signal reconstruction problem shown in Figure 1.2 where  $\mathcal{G}$  is fixed as an ideal low pass filter bandlimited to the Nyquist frequency.

Any system whose present output depends upon the future inputs is known as non-causal. In contrary, a causal system does not have access to future. Any ideal low pass filter is not practically realizable because its present output depends upon *all* of the future inputs. This is an another limitation of the Shannon's theorem. Starting from [23] in 1996, sampled data system theory is applied to several signal processing applications using different error criteria with or without causality constraints. For example downsampling with causality constraints (using fast sampler/fast hold approximation) is treated in [20, 43, 41], audio compression in [1], image application in [22] etc. For a complete list of applications see the review paper by Yamamoto et al. [66].

We know that if the analog input signal is bandlimited and we are free to choose the sampler and hold then by Shannon's sampling theorem we have zero reconstruction error if our sampler is the ideal sampler and our hold is given by (1.1). Therefore, they are optimal in this case. However, if the input analog signal  $y$  is not bandlimited then passing it through an ideal low-pass filter (bandlimited to the Nyquist frequency) before ideally sampling and using the hold in (1.1), is also optimal (in some norm sense). This is proved in [59, 58, 31] where [31] used sampled-data system theory. Meinsma and Mirkin [31] also applied sampled-data system theory to the cases where a non-causal sampler (or hold) is fixed and we have to design hold (or sampler) [31]. They also designed relaxed causal (i.e. with limited access to future) hold given a sampler using sampled data system theory [29].

## 1.2 Problem formulation

The main objective in this research is to solve some of the signal processing problems using the sampled data system theory. Generally the design of filters in signal processing is done either in discrete or in analog domain. However, in most of the signal processing applications (e.g. audio processing) nowadays the primary information is in analog format and utilized in analog domain at the end but the information is processed digitally. The situation is depicted in Figure 1.2. In these signal processing applications, we are interested in minimizing the reconstruction error  $e$  (which is analog) but internally the information is processed digitally. Such



**Figure 1.3:** Downsampling in sampled-data setting

a system can be analyzed and designed using Shannon's theorem however then we have to restrict our input signals  $y$  to just bandlimited signals (i.e. the signal model  $\mathcal{G}$  is a bandlimited system). System theoretic methods are useful to generalize analysis of such a systems with a generic class of input signals  $y$ . In these methods, sampler, hold and the signal models are treated as operators or systems operating on the signals in fairly large class. System theoretic method such as sample data system theory enables us to analyze the signal processing setup shown Figure 1.2 containing samplers, holds and signal models in a common framework. This approach also helps us in the analysis and design in a unified manner for both stochastic and deterministic signals [3, 2]. Another advantage of using sampled-data system theory in the design process is that we can calculate the reconstruction error without any practical implementation.

The reconstruction error gives us a criteria to measure the performance of our design. It is shown in [31] that frequency truncated norms naturally arise in signal processing via sample-data system theory. Direct integration for these type of norms is often time consuming, therefore it is preferred to have a closed form expressions for the frequency truncated norms. *The first aim of this research is to obtain a closed form expressions for the frequency truncated norm.*

Sample data system theory has been used in solving several problems in signal processing in a generic way. Some of these problems are already answered in [29, 23, 66] and the references therein. In this thesis, we use sample-data system theory to obtain a generic solution of downsampling and optimal relaxed causal sampling problems.

Downsampling of the sampled signal is required in several signal processing applications. Downsampling can be achieved by a use of downsampler  $\bar{S}_h$  in between sampler and hold which reduces the sampling rate of its input discrete signal by some integer factor (see Figure 1.3). The downsampling problem we consider is to design the downsampler and the hold given the sampler and the signal model. Earlier most of the approaches were somehow based on Shannon's sampling theorem (i.e. by bandlimiting the signal model  $\mathcal{G}$ ). As an alternative, [20, 41, 42] used sampled-data system theory to solve downsampling problems using fast-sample/fast-hold approximation. Meinsma and Mirkin [31] solved the downsampling problem in a generic sample-data framework however for a limited class of signal models. To move further, we consider the downsampling problem with all linear continuous time invariant (LCTI) signal models. Hence, *the second*



*aim of this research is to obtain a solution for the downsampling problem with linear continuous time invariant signal models.*

In the signal reconstruction problem, non-causal samplers given a hold and signal model can be designed by the method explained in [31]. Even though non-causal solutions are important in obtaining limiting behavior of our systems, it is not always practical. In practice, our system must be causal or relaxed causal. The design of an optimal causal sampler using sampled-data system theory is discussed in [37, 38, 43]. Relaxed causality means that we have some limited access to the future inputs. The constraint of relaxed causality makes our problem quite a bit more difficult, but interesting also. *The third aim of this research is to provide a frequency domain abstract and state space solution to optimal sampler design problem with relaxed causality constraint.*

## 1.3 Overview of the thesis

This sections contains an overview of the chapters in this thesis.

### **Chapter 2: Sampled-data system theory**

This chapter mainly contains the fundamentals of the sampled-data system theory. Most of the content in this chapter is from [30]. However, this chapter also takes inspiration from important work like [3], [2], [66], [57]. Signal generators or signal models, samplers, holds are important components in the sampled data system theory. Details of these components are discussed in this chapter. This chapter also contains the concept of lifting, lifting transforms, signal and system norms which serves as a foundation to the later chapters.

### **Chapter 3: Frequency truncated norms**

This chapter contains a method to express frequency truncated norms in terms of the matrix logarithm. The results in this chapter can be applied in other areas of system theory like model reduction.

### **Chapter 4: Non-causal downsampling**

This chapter concentrates on the downsampling problem using sampled data system theory. It contains a general formulation and solution of optimal downsampling in the sampled-data setup for all linear continuous time invariant signal models. Here we allow non-causal solutions. The effect of noise on the downsampling is also discussed in this chapter.

### **Chapter 5: Relaxed causal sampling**

This chapter concentrates on the optimal relaxed causal sampler design. In this chapter, we provide a frequency domain abstract solution to optimal sampler de-

sign with relaxed causality constraint. Mirkin [35, 36] introduced STPBC (state-space with two-point boundary condition) representations for linear  $h$ -time shift invariant system. We also discuss this representation in detail. This representation is useful in obtaining a closed form solution of the optimal sampler with relaxed causality constraint. We also give an expression for the minimal error norm for the optimal sampler.

## **Chapter 6: Conclusions and Recommendations**

This small chapter contains a summary of the important results in this thesis. It also contains some notes on the further research related to the topics discussed in this thesis.

## **1.4 Overview of contributions**

The research objectives that are met in this thesis are

1. A closed form expression for the frequency truncated norms in terms of matrix logarithm for systems given in the state-space.
2. A system theoretical analysis of the downsampling problem, and design of optimal non-casual downsamplers and hold. The effect of noise on the downsampling problem is also analyzed, and optimal non-casual downsamplers and hold are designed in the presence of noise.
3. A frequency domain abstract and state-space solution to optimal sampler design problem with relaxed causality constraint. A method for calculation of the reconstruction error is also obtained.



# Chapter 2

## Sampled-data system theory

To obtain the analog signal (at least approximately) from its samples is the primary aim in signal reconstruction. Sampled data system theory facilitates us in reconstruction and to measure the error of signal reconstruction (see Chapter 1). This chapter contains a general introduction to sampled-data system theory and serves as background material for the later chapters. Specifically, in this chapter we study the sampled data setup shown in Figure 2.3 (on page 11) for the signal reconstruction. Our main aim in this chapter is to describe all components of the sampled-data setup in a mathematical way. We also discuss the concept of lifting, lifting transforms, and signal and system norm that are useful for later chapters. Most of the theory discussed in this chapter is based on the paper by Meinsma and Mirkin [30]. Further details on the topics discussed in this chapter can be found in [66, 57, 3, 2] and the references therein.

### 2.1 Notation

Due to various systems, spaces and transforms in this chapter, it is useful to summarize all the notations in one place. The meaning of these systems, domains and transforms will be cleared in the later sections.

In this thesis, we represent systems by uppercase letters and signals by lowercase letters.

A system in the time domain is represented by calligraphic letter e.g.  $\mathcal{G}$ . A system in the lifted time domain is represented by calligraphic letter with a breve on top e.g.  $\check{\mathcal{G}}$ . A system in the lifted frequency domain is represented by capital letter with a breve on top e.g.  $\check{G}$ . A linear continuous time invariant (LCTI) system in the classic frequency domain is represented by capital letter e.g.  $G(j\omega)$ . A hold in the lifted time domain is represented by calligraphic letter with a grave on top e.g.  $\grave{\mathcal{H}}$ . A hold in the lifted frequency domain is represented by capital letter with a grave on top e.g.  $\grave{H}$ . A sampler in the lifted time domain is represented by calligraphic letter with an acute on top e.g.  $\acute{\mathcal{S}}$ . A sampler in the lifted frequency domain is

System	time domain	lifted time domain	lifted frequency domain	classic frequency domain
Analog	$\mathcal{G}$	$\check{\mathcal{G}}$	$\check{\check{\mathcal{G}}}$	
LCTI	$\mathcal{G}$	$\check{\mathcal{G}}$	$\check{\check{\mathcal{G}}}$	$G$
Hold	$\mathcal{H}$	$\check{\mathcal{H}}$	$\check{\check{\mathcal{H}}}$	
Sampler	$\mathcal{S}$	$\check{\mathcal{S}}$	$\check{\check{\mathcal{S}}}$	
Discrete	$\bar{\mathcal{W}}$	$\bar{\check{\mathcal{W}}}$	$\bar{\check{\check{\mathcal{W}}}}$	$\bar{W}$

**Table 2.1:** Notations for systems in different domains

represented by capital letter with an acute on top e.g.  $\acute{S}$ . Now, we summarize the notations for different systems in table 2.1. Here an analog system maps an analog signal to an analog signal and a discrete system maps a discrete signal to a discrete signal. The definition of a sampler, hold and LCTI systems is given later in this chapter.

An analog or continuous signal is represented by lowercase letter e.g.  $y$ . A discrete signal is represented by a bar on top e.g.  $\bar{y}$ . Square brackets are used to denote the value of a discrete signal at a given integer e.g.  $\bar{y}[k]$  whereas parenthesis are used to denote the value of an analog signal at a given time e.g.  $y(t)$ . An analog (or discrete) signal in the lifted domain is represented by  $\check{y}$  (or  $\bar{y}$ ). In an apologetic way, the lifted  $z$ -transform of a continuous (or discrete) signal  $y$  (or  $\bar{y}$ ) is represented by  $\check{y}$  (or  $\bar{y}$ ) with a suffix ( $z$ ). Similarly, the lifted Fourier transform of a continuous (or discrete) signal  $y$  (or  $\bar{y}$ ) is represented by  $\check{y}$  (or  $\bar{y}$ ) with a suffix ( $e^{j\theta}$ ). Most of the time it is clear from the context if the signal is in the lifted (time) domain or lifted frequency domain ( $z$ -transform or Fourier transform). In case it is really necessary to make a distinction, we use  $\mathfrak{Z}(\check{y})$  (or  $\mathfrak{Z}(\bar{y})$ ) for the  $z$ -transform of lifted signal  $\check{y}$  (or  $\bar{y}$ ). Similarly, we use  $\mathfrak{F}(\check{y})$  (or  $\mathfrak{F}(\bar{y})$ ) for the Fourier transform of lifted signal  $\check{y}$  (or  $\bar{y}$ ).

With a little bit of overloading the notations, the classic continuous (or discrete) time Fourier transform of an analog signal  $y$  (or a discrete signal  $\bar{y}$ ) is represented with different arguments as  $y(j\omega)$  (or  $\bar{y}(e^{j\theta})$ ). Most of the time the signal domain is clear from the context. Sometimes to make distinction between  $y(j\omega)$  and the time domain  $y(t)$  (or due to historic reasons) we represent the classic continuous (or discrete) time Fourier transform of  $y$  (or  $\bar{y}$ ) in capitals as  $Y(j\omega)$  (or  $\bar{Y}(e^{j\theta})$ ). The notations used for signals is summarized in table 2.2 (page 11).

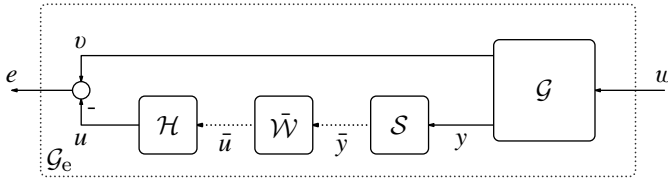
For general discussions (applicable to all type of systems) we put sometimes the tilde on top of the system name to denote all lifted linear  $h$ -time shift invariant system or shift invariant discrete system including samplers and holds also. Time domain, lifted domain and lifted frequency domain systems are differentiated by usual notation given in this section. For example, the time domain systems are represented by  $\mathcal{G}$  (including discrete systems also), lifted time domain systems

Signal	time domain	lifted time domain	lifted Fourier domain	lifted z-domain	classic frequency domain
Continuous	$y$	$\tilde{y}$	$\tilde{y}(e^{j\theta}), \mathfrak{F}(\tilde{y})$	$\tilde{y}(z), \mathfrak{Z}(\tilde{y})$	$y(j\omega), Y(j\omega)$
Discrete	$\bar{y}$	$\bar{y}$	$\bar{y}(e^{j\theta}), \mathfrak{F}(\bar{y})$	$\bar{y}(z), \mathfrak{Z}(\bar{y})$	$\bar{y}(e^{j\theta}), \bar{Y}(e^{j\theta})$

**Table 2.2:** Notations for signals in different domains

are represented by  $\tilde{\mathcal{G}}$  and lifted frequency domain systems are represented by  $\tilde{G}$ . Similarly  $\tilde{\cdot}$  is used to denote all lifted analog or discrete signals. For example, a time domain signal  $y$  or  $\bar{y}$  in the lifted domain is represented by  $\tilde{y}$ .

## 2.2 Sampled-data system



**Figure 2.3:** Sampled-data setup

The sampled-data setup shown in Figure 2.3 is fundamental to about all problems that are considered in this thesis. The setup consists of an analog system  $\mathcal{G}$  known as signal generator or model, a discrete system  $\tilde{W}$ , a sampler  $\mathcal{S}$  and a hold  $\mathcal{H}$ . In this section, we go through these systems one by one.

### 2.2.1 Signal Generator

As discussed in Chapter 1, *signal generators* can be used to model our apriori knowledge about signals  $v$  and  $y$ . This apriori knowledge can be about bandwidth, cross-correlation, spectral density etc. of  $v$  and  $y$ . A detailed discussion about signal generator is given in [30, 7].

In this thesis, we assume that the signal generator  $\mathcal{G}$  which maps an analog signal  $w : \mathbb{R} \rightarrow \mathbb{C}^{n_w}$  to  $y : \mathbb{R} \rightarrow \mathbb{C}^{n_y}$ , is linear and  $h$ -time shift invariant. Here  $n_w$  and  $n_y$  are positive integers. Linearity means  $\mathcal{G}$  satisfies the additivity and homogeneity properties [27]. The  $h$ -time shift invariance of a system means that if we delay the system input by  $kh$  ( $k \in \mathbb{Z}$ ) then the system output is also delayed by  $kh$ . As given in [27, 30], linearity implies that the output  $y$  of the system  $\mathcal{G}$  driven

by input  $w$  is of the form

$$y(t) = \int_{-\infty}^{\infty} g(t, s)w(s)ds \quad (2.1)$$

for some kernel  $g(t, s)$  and  $h$ -time shift invariance of the system  $\mathcal{G}$  implies that the kernel  $g(t, s)$  satisfies

$$g(t, s) = g(t + kh, s + kh)$$

for all  $k \in \mathbb{Z}$ . In this thesis  $h$  is fixed, therefore the linear  $h$ -time shift invariant systems are sometimes called *linear discrete time invariant (LDTI)* systems.

An  $h$ -time shift invariant system is not necessarily  $h'$ -time shift invariant if  $h' \neq kh$  where  $k$  is a positive integer. However, if the system is  $h'$ -time shift invariant for every  $h' \in \mathbb{R}$ , then it is called *linear continuous time invariant (LCTI)*. In other words,  $\mathcal{G}$  is LCTI iff it is of the form (2.1) with kernel  $g(t, s) = g(t - s, 0)$ . In this case,  $g(t - s) := g(t - s, 0)$  and  $g(r)$ ,  $r \in \mathbb{R}$  is known as as the *impulse response* of the LCTI system [27].

## 2.2.2 Sampler

A sampler  $\mathcal{S}$  is a system that maps an analog signal  $y : \mathbb{R} \rightarrow \mathbb{C}^{n_y}$  to a discrete signal  $\bar{y} : \mathbb{Z} \rightarrow \mathbb{C}^{n_{\bar{y}}}$ . Here  $n_y$  and  $n_{\bar{y}}$  are positive integers. We assume that the sampler is linear and  $h$ -time shift invariant. Here  $h$ -time shift invariance means that if we delay the sampler's analog input by  $h$  then its discrete output is delayed by one. Every such sampler is of the form

$$\bar{y} = \mathcal{S}y : \quad \bar{y}[n] = \int_{-\infty}^{\infty} \psi(nh - s)y(s) ds \quad (2.2)$$

for some function  $\psi(t)$ . The function  $\psi(t)$  known as the *sampling function* of sampler  $\mathcal{S}$  and  $h$  is known as the *sampling period*. Although the proof of (2.2) is standard, for completeness it is given Appendix 2.A (page 34).

**Example 2.2.1.** *The ideal sampler  $\mathcal{S}_{\text{idl}}$  is an example of a sampler and it is given by*

$$\bar{y} = \mathcal{S}_{\text{idl}}y : \quad \bar{y}[n] = y(nh). \quad (2.3)$$

*In this case, the ideal sampler can be written in the form (2.2) with the sampling function  $\psi(t) = \delta(t)$ .*

Note that the sampler  $\mathcal{S}$  in (2.2) can be written as a cascade of an LCTI system with impulse response  $\psi(t)$  and the above ideal sampler  $\mathcal{S}_{\text{idl}}$  as

$$\bar{y} = \mathcal{S}_{\text{idl}}(\psi * y)$$

where convolution is defined as  $(\psi * y)(t) := \int_{-\infty}^{\infty} \psi(t - s)y(s)ds$  for all  $t \in \mathbb{R}$ .

### 2.2.3 Hold

A hold is a system which converts a discrete signal  $\bar{u} : \mathbb{Z} \rightarrow \mathbb{C}^{n_{\bar{u}}}$  back to an analog signal  $u : \mathbb{R} \rightarrow \mathbb{C}^{n_u}$ . Here  $n_{\bar{u}}$  and  $n_u$  are positive integers. Note that the input dimension  $n_{\bar{u}}$  can be different from the output dimension  $n_u$  of the hold. We assume that  $\mathcal{H}$  is linear and  $h$ -time shift invariant. Here  $h$ -time shift invariance means that if we delay the hold's discrete input by one then its analog output is delayed by  $h$ . Every such hold is of the form

$$u = \mathcal{H}\bar{u} : \quad u(t) = \sum_{n \in \mathbb{Z}} \phi(t - nh)\bar{u}[n], \quad t \in \mathbb{R} \quad (2.4)$$

for some function  $\phi(t)$  known as the *hold function* or *interpolating kernel*. The derivation of (2.4) is standard and it is given Appendix 2.A (page 35) for reference purpose.

**Example 2.2.2.** A generalized zero order hold  $\mathcal{H}_z$  is an example of a hold and it is given by

$$u = \mathcal{H}_z\bar{u} : \quad u(t) = \phi_z \left( t - \left\lfloor \frac{t}{h} \right\rfloor h \right) \bar{u} \left[ \left\lfloor \frac{t}{h} \right\rfloor \right], \quad t \in \mathbb{R}. \quad (2.5)$$

where  $\lfloor t \rfloor$  denotes the largest integer less than or equal to  $t$  and  $\phi_z : [0, h) \rightarrow \mathbb{C}^{n_u}$ . In this case, the hold function is

$$\phi(t) := \begin{cases} \phi_z(t) & t \in [0, h) \\ 0 & t \notin [0, h) \end{cases}.$$

The ideal zero order hold  $\mathcal{H}_{iz}$  is a special case of the zero order hold  $\mathcal{H}_z$  where  $\phi(t) = \mathbb{1}_{[0, h)}(t)$  and it is given by

$$u = \mathcal{H}_{iz}\bar{u} : \quad u(t) = \bar{u} \left[ \left\lfloor \frac{t}{h} \right\rfloor \right], \quad t \in \mathbb{R}. \quad (2.6)$$

Note that the hold  $\mathcal{H}$  in (2.4) can be written as an LCTI system with impulse response  $\phi(t)$ , preceded by a modulated impulse train as

$$\begin{aligned} u(t) &= \int_{-\infty}^{\infty} \phi(t - s) \sum_{n \in \mathbb{Z}} \delta(s - nh)\bar{u}[n] ds, \quad t \in \mathbb{R} \\ &= \sum_{n \in \mathbb{Z}} \phi(t - nh)\bar{u}[n]. \end{aligned}$$

### 2.2.4 Discrete system

A discrete system  $\bar{\mathcal{W}}$  maps a discrete signal  $\bar{y} : \mathbb{Z} \rightarrow \mathbb{C}^{n_{\bar{y}}}$  to a discrete signal  $\bar{u} : \mathbb{Z} \rightarrow \mathbb{C}^{n_{\bar{u}}}$ . We consider discrete systems which are linear and shift invariant. Shift invariance means that if we delay the discrete system's input by one then



its discrete output is delayed by one. Every such discrete system is given by the following (discrete) convolution,

$$\bar{u} = \bar{\mathcal{W}}\bar{y} \quad : \quad \bar{u}[n] = \sum_{k \in \mathbb{Z}} \bar{w}[n - k]\bar{y}[k], \quad (2.7)$$

where  $\bar{w}[k]$  is known as *impulse response* of  $\bar{\mathcal{W}}$ . If  $\bar{w}[k] = 0 \forall k \neq 0$  then such a discrete system is called *static discrete system*.

In the sampled-data system theory, the time between the samples plays an important role. Therefore a shift invariant discrete system which maps a discrete signal with period  $h$  to a discrete signal with period  $h$  is also called  $h$ -time shift invariant.

The discrete system can be absorbed in the sampler or hold by redefining the sampling or hold function in the sampled-data setup as shown in following corollary.

**Corollary 2.2.3.** *Let  $\mathcal{S}$ ,  $\mathcal{H}$  and  $\bar{\mathcal{W}}$  be as in (2.2), (2.4), and (2.7) respectively. Then, the series interconnection  $\mathcal{H}\bar{\mathcal{W}}$  is a hold with hold function  $\sum_{i \in \mathbb{Z}} \phi(t - ih)\bar{w}[i]$  and the series interconnection  $\bar{\mathcal{W}}\mathcal{S}$  is a sampler with sampling function  $\sum_{i \in \mathbb{Z}} \bar{w}[i]\psi(t - ih)$ .*

*Proof.* See Appendix 2.A (page 35). □

By Corollary 2.2.3, any discrete system following a sampler or preceding a hold can be merged in the sampler or the hold. Therefore, most of the time we consider  $\bar{\mathcal{W}} = I$  (i.e. with impulse response  $\bar{w}[k] = \bar{\delta}[k]$ ) without loss of generality in the sampled data setup.

## 2.3 Lifting

Lifting is now standard in sampled-data literature (see [23, 30, 57] and the references there in for more details). In this section, we give a brief overview of lifting techniques.

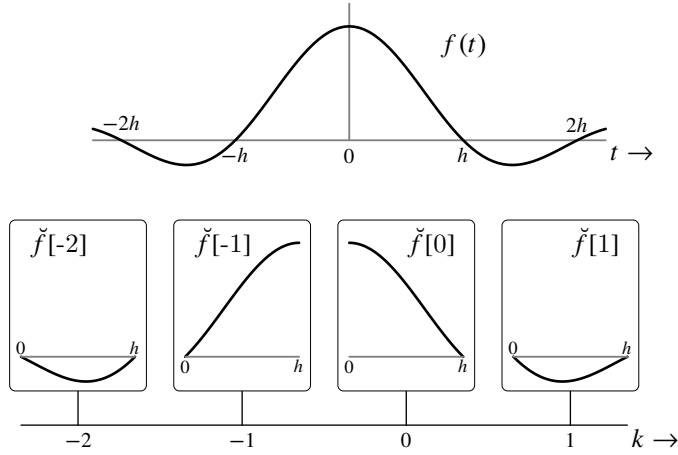
### 2.3.1 Lifting in time domain

Consider a linear  $h$ -time shift invariant system  $\mathcal{G}$  given in (2.1). In order to define the transfer function of such a system, we lift input  $w$  and output  $y$  of the system  $\mathcal{G}$ . The lifting of an analog/continuous time signal is defined as follows.

**Definition 2.3.1.** *For a continuous time signal  $f : \mathbb{R} \rightarrow \mathbb{C}^n$ , the lifted signal  $\check{f} : \mathbb{Z} \rightarrow \{[0, h) \rightarrow \mathbb{C}^n\}$  is the sequence of functions  $\{\check{f}[k]\}$  defined as*

$$\check{f}[k](\tau) := f(kh + \tau), \quad k \in \mathbb{Z}, \tau \in [0, h).$$

*In this context,  $h$  is known as the lifting period.*



**Figure 2.4:** Time domain lifting of  $f(t) = \text{sinc}(t/h) := \frac{\sin(\pi t/h)}{\pi t/h}$ .

**Remark 2.3.2.** The natural domain for  $\tau$  is  $[0, h)$  because then there exists a bijection between the signal  $f$  and its lifted signal  $\check{f}$ . However, sometimes it is beneficial to define  $\check{f}[k](\tau) := f(kh + \tau)$  for arbitrary  $\tau \in \mathbb{R}$ .

Figure 2.4 explains the idea of lifting. It is clear from Definition 2.3.1 that lifting is an invertible process. Sometimes the lifting in Definition 2.3.1 is called *continuous lifting* or *analog lifting*.

**Remark 2.3.3.** In this thesis, unless mentioned differently, lifting always means lifting with lifting period  $h$  which is also the sampling period of the sampler in (2.2). However, theoretically lifting can be done for intervals different from the sampling period of the sampler.

**Remark 2.3.4.** A discrete signal (generated by sampling of an analog signal with sampling period  $h$ ) can be thought of as a sequence whose elements are separated by  $h$  time. Therefore, lifting (with lifting period  $h$ ) of such a discrete signal is defined as the discrete signal itself. However, the sampling period plays a crucial role here. For example if our discrete signal  $\bar{y}$  is generated by sampling an analog signal  $y$  with sampling period  $h/4$ . Then we have four samples in the interval  $h$ . Therefore lifting (with lifting period  $h$ ) is a sequence whose elements contain four samples (see Figure 4.5). Discrete lifting is discussed in more detail in Chapter 4. Since in this chapter we always perform lifting with respect to the sampling period i.e. the lifting period and the sampling period are the same, discrete signal (say  $\bar{y}$ ) and the lifted discrete signal are the same. Hence, discrete signal and the lifted discrete signal are represented by the same symbol  $\bar{y}$ .

Lifting the input and the output of a system will naturally define the lifting of the system. In the rest of this section, we define the lifted analog system, hold and sampler.

If  $y = \mathcal{G}w$ , then we denote the mapping from the lifted  $w$  to the lifted  $y$  by  $\check{\mathcal{G}}$  i.e.  $\check{y} = \check{\mathcal{G}}\check{w}$ . It is shown in [30, §3] that if  $\mathcal{G}$  given in (2.1) is linear  $h$ -time shift invariant then

$$\check{y} = \check{\mathcal{G}}\check{w} : \check{y}[k] = \sum_{i \in \mathbb{Z}} \check{\mathcal{G}}[k-i]\check{w}[i], \quad k \in \mathbb{Z} \quad (2.8)$$

where  $\check{\mathcal{G}}[k] : \{[0, h) \rightarrow \mathbb{C}^{n_w}\} \rightarrow \{[0, h) \rightarrow \mathbb{C}^{n_y}\}$  is the *lifted impulse response system* of the system  $\check{\mathcal{G}}$  given by

$$(\check{\mathcal{G}}[k]x)(\tau) = \int_0^h g(kh + \tau, \sigma)x(\sigma) d\sigma \quad \tau \in [0, h). \quad (2.9)$$

Linear continuous time invariant (LCTI) systems are special cases of linear  $h$ -time shift invariant systems. If  $\mathcal{G}$  is LCTI, then the *lifted impulse response system*  $\check{\mathcal{G}}[k]$  of lifted system  $\check{\mathcal{G}}$  given in (2.8) is given by

$$(\check{\mathcal{G}}[k]\check{w})(\tau) = \int_0^h g(kh + \tau - \sigma)\check{w}(\sigma) d\sigma \quad \tau \in [0, h), \quad (2.10)$$

where  $g(t)$  is the impulse response of the LCTI system  $\mathcal{G}$ . The word *system* is used in *lifted impulse response system* to emphasize the fact that  $\check{\mathcal{G}}[k]$  is an operator.

**Example 2.3.5.** Let  $\mathcal{G}$  be the LCTI system with impulse response

$$g(t) := \mathbb{1}_{[0, h)}(t).$$

Then the kernel  $g(kh + \tau - \sigma)$  of the lifted impulse response system  $\check{\mathcal{G}}[k]$  is given by

$$g(kh + \tau - \sigma) = \begin{cases} \mathbb{1}_{[0, h)}(\tau - \sigma) & k = 0 \\ \mathbb{1}_{[-h, 0)}(\tau - \sigma) & k = 1 \\ 0 & \text{elsewhere} \end{cases}$$

Note that  $\tau, \sigma \in [0, h)$ .

Lifting the input  $y$  and output  $\bar{y}$  of a sampler  $\mathcal{S}$  in (2.2) gives a *lifted sampler*

$$\bar{y} = \acute{\mathcal{S}}\check{y} : \bar{y}[k] = \sum_{i \in \mathbb{Z}} \acute{\mathcal{S}}[k-i]\check{y}[i] \quad (2.11)$$

where  $\acute{\mathcal{S}}[k] : \{[0, h) \rightarrow \mathbb{C}^{n_y}\} \rightarrow \mathbb{C}^{n_{\bar{y}}}$  is the *lifted impulse response system* of the lifted system  $\acute{\mathcal{S}}$  given by

$$\acute{\mathcal{S}}[k]x = \int_0^h \psi(kh - \sigma)x(\sigma) d\sigma.$$

Similarly, lifting the input  $\bar{u}$  and the output  $u$  of a hold  $\mathcal{H}$  in (2.4) gives a *lifted hold*

$$\check{u} = \check{\mathcal{H}}\bar{u} : \quad \check{u}[k] = \sum_{i \in \mathbb{Z}} \check{\mathcal{H}}[k-i]\bar{u}[i] \quad (2.12)$$

where  $\check{\mathcal{H}}[k] : \mathbb{C}^{n_{\bar{u}}} \rightarrow \{[0, h) \rightarrow \mathbb{C}^{n_u}\}$  is the *lifted impulse response system* of the lifted system  $\check{\mathcal{H}}$  given by

$$(\check{\mathcal{H}}[k]\bar{x})(\tau) := \check{\phi}(\tau)\bar{x}, \quad \tau \in [0, h).$$

Shift invariance of a lifted system means that if its lifted input is delayed by 1, then its lifted output is delayed by 1 as well. Shift invariance of the lifted systems  $\check{\mathcal{G}}$ ,  $\check{\mathcal{S}}$  and  $\check{\mathcal{H}}$  is the consequence of the fact that the corresponding time domain systems  $\mathcal{G}$ ,  $\mathcal{S}$ , and  $\mathcal{H}$  are  $h$ -time shift invariant [30]. However, the advantage with the lifted system is that they behave like shift-invariant discrete systems (see (2.20),(2.22), (2.23)). Therefore, we can expect that most of the theory for a discrete system may have something analogous (e.g. convolution,  $z$  and Fourier transforms) for the lifted system as well. This is indeed the case as we will see in later sections.

Now we consider the cascade of two  $h$ -time shift invariant systems  $\mathcal{G}_1$  and  $\mathcal{G}_2$  with kernel  $g_1$  and  $g_2$  respectively. Using (2.1), we have

$$y = \mathcal{G}_1\mathcal{G}_2w : \quad y(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(t, s)g_2(s, r)w(r) dr ds \quad (2.13)$$

where  $t \in \mathbb{R}$ . Using (2.8), it can be proved that the seemingly difficult integration in (2.13) is transformed to the following discrete convolution in the lifted domain

$$\check{y} = \check{\mathcal{G}}_1\check{\mathcal{G}}_2\check{u} : \quad \check{y}[k] = \sum_{n \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \check{\mathcal{G}}_1[k-i]\check{\mathcal{G}}_2[i-n]\check{u}[n] \quad (2.14)$$

where  $\check{\mathcal{G}}_i[k]$ ,  $k \in \mathbb{Z}$  is the lifted impulse response system of the lifted system  $\check{\mathcal{G}}_i$ . Let  $\check{\mathcal{G}}_3[k] := \sum_{i \in \mathbb{Z}} \check{\mathcal{G}}_1[k-i]\check{\mathcal{G}}_2[i]$  then (2.14) is equivalent to

$$\check{y} = \check{\mathcal{G}}_1\check{\mathcal{G}}_2\check{u} : \quad \check{y}[k] = \sum_{n \in \mathbb{Z}} \check{\mathcal{G}}_3[k-n]\check{u}[n].$$

Therefore, lifting translates the series interconnection into a familiar convolution. Thus lifting puts the inter-sample behavior of the system in the background in such a way that we can treat the lifted system as a linear shift invariant discrete system. However, the inter-sample behavior of the system is not lost after lifting. The advantage of lifting is that we can use convolution (as shown in this section). However, this advantage come at the price of difficult impulse responses.

### 2.3.2 Lifting in frequency domain

After lifting, a linear  $h$ -time shift invariant system can be treated like linear shift-invariant discrete systems. Therefore we can apply frequency domain methods to these systems. In this section we define the  $z$ -transform and the Fourier transform of signals and systems.

### Lifted $z$ and Fourier transform

As lifted signals are sequences, we define the (lifted)  $z$ -transform of lifted signals as

**Definition 2.3.6.** The  $z$ -transform  $\check{\mathfrak{Z}}(\check{f})$  of a lifted signal  $\check{f}$  is defined as

$$\check{\mathfrak{Z}}(\check{f}) = \check{f}(z; \tau) := \sum_{k \in \mathbb{Z}} \check{f}[k](\tau) z^{-k} = \sum_{k \in \mathbb{Z}} f(kh + \tau) z^{-k} \quad (2.15)$$

where  $\tau \in [0, h)$ .  $\check{\mathfrak{Z}}(\check{f})$  is also called lifted  $z$ -transform of the signal  $f$ .

**Remark 2.3.7.** As discussed in Remark 2.3.4, in this chapter the lifting period and the sampling period are the same. Therefore, the discrete signal (say  $\bar{y}$ ) and the lifted discrete signal are the same. For this reason, the lifted  $z$ -transform of such a signal is represented by  $\bar{y}(z)$ .

Similarly the (lifted) Fourier transform is defined as:

**Definition 2.3.8.** The Fourier transform  $\check{\mathfrak{F}}(\check{f})$  of a lifted signal  $\check{f}$  is defined as

$$\check{\mathfrak{F}}(\check{f}) = \check{f}(e^{j\theta}; \tau) := \sum_{k \in \mathbb{Z}} \check{f}[k](\tau) e^{-j\theta k} = \sum_{k \in \mathbb{Z}} f(kh + \tau) e^{-j\theta k} \quad (2.16)$$

where  $\tau \in [0, h)$ .  $\check{\mathfrak{F}}(\check{f})$  is also called lifted Fourier transform of the signal  $f$ .

Clearly, for a given  $\tau$ ,  $\check{f}(e^{j\theta}; \tau)$  is the discrete time Fourier transform [27] of the sequence  $\check{f}[k](\tau)$  and  $\check{f}(e^{j\theta}; \tau)$  is periodic in  $\theta$  with period  $2\pi$ .

In most of the cases we deal with real signals. However, most of the results in later chapters are formulated in the lifted frequency domain. To check whether a given lifted Fourier transform corresponds to a real signal or not, the following straight-forward result is useful.

**Corollary 2.3.9.** If the lifted Fourier transform  $\check{f}(e^{j\theta}; \tau)$  of a signal  $f$  exists then  $f$  is real if and only if  $\check{f}(e^{j\theta}; \tau) = \overline{\check{f}(e^{-j\theta}; \tau)}$  for all  $\theta \in [-\pi, \pi]$  and  $\tau \in [0, h)$ .

*Proof.* See Appendix 2.A (page 35).  $\square$

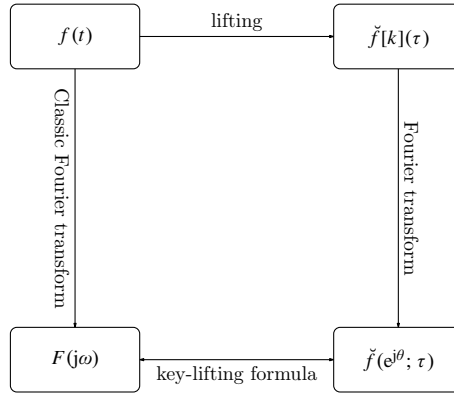
So far we were bit sloppy about the existence of Fourier transforms and  $z$ -transform. However, for the following important result which is similar to Poisson summation formula, the existence of various transforms do matter [30, 6].

**Theorem 2.3.10** (Key lifting formula [30]). Let  $f$  be an analog signal such that  $f(t)e^{-s_0 t}$  belongs to  $L^2(\mathbb{R})$  for some  $s_0 \in \mathbb{C}$ . Then the two-sided Laplace transform  $F(s)$  of  $f(t)$  exists and we have the following properties

$$\check{f}(e^{s_0 h}; \tau) = \frac{1}{h} \sum_{k \in \mathbb{Z}} F(s_k) e^{s_k \tau} \quad (2.17a)$$

$$F(s_k) = \int_0^h \check{f}(e^{s_0 h}; \tau) e^{-s_k \tau} d\tau \quad (2.17b)$$

for all  $\tau \in [0, h)$ , where  $s_k := s_0 + j\frac{2\pi k}{h}$ .



**Figure 2.5:** Relation between various transforms and Key lifting formula.

*Proof.* See [30]. □

Equation (2.17) relates the (lifted)  $z$ -transform (for  $z = e^{s_0}$ ) with the Laplace transform and can be interpreted as a bijection between  $\{F(s_k)\}_{k \in \mathbb{Z}}$  and  $\check{f}(e^{s_0}; \tau)$ . As a special case, when  $s_0 = \frac{j\theta}{h}$ , we have the bijection between the (lifted) Fourier transform and the classical Fourier transform:

$$\check{f}(e^{j\theta}; \tau) = \frac{1}{h} \sum_{k \in \mathbb{Z}} F(j\omega_k) e^{j\omega_k \tau} \quad (2.18a)$$

$$F(j\omega_k) = \int_0^h \check{f}(e^{j\theta}; \tau) e^{-j\omega_k \tau} d\tau \quad (2.18b)$$

where  $\omega_k := \frac{\theta + 2\pi k}{h}$ . For further detail and applications of the key lifting formula see [30, §IV.A]. Figure 2.5 explains the relation between the various transforms and the Key lifting formula.

**Remark 2.3.11.** *The equalities in (2.17a) and (2.18a) are in the  $L^2$  sense.*

### Transfer function

The transfer function of a linear  $h$ -time shift invariant system  $\mathcal{G}$  is defined as the  $z$ -transform of its lifted impulse response system

$$\check{G}(z) := \sum_{k \in \mathbb{Z}} \check{G}[k] z^{-k}. \quad (2.19)$$

Sometimes we call  $\check{G}(z)$  as *lifted transfer function* of  $\mathcal{G}$ . In the rest of this section we define transfer functions for arbitrary linear  $h$ -time shift invariant systems, including samplers and holds. See [30], for more detailed discussion on (lifted) transfer function.

Taking the  $z$ -transform of the output  $\check{y}$  and the input  $\check{w}$  of a lifted system  $\check{\mathcal{G}}$  in (2.8) results in *lifted frequency domain system*

$$\mathfrak{Z}(\check{y}) = \check{G}\mathfrak{Z}(\check{w}) : \check{y}(z) = \check{G}(z)\check{w}(z)$$

where  $\check{G}(z) : \{[0, h) \rightarrow \mathbb{C}^{n_w}\} \rightarrow \{[0, h) \rightarrow \mathbb{C}^{n_y}\}$  (it is the  $z$ -transform of the lifted impulse response system of the lifted system  $\check{\mathcal{G}}$ ) is given by (see [30, §4])

$$\check{y}(z) = \check{G}(z)\check{w}(z) : \check{y}(z; \tau) = \int_0^h \check{g}(z; \tau, \sigma)\check{w}(z; \sigma)d\sigma. \quad (2.20)$$

Here  $\tau \in [0, h)$  and  $\check{g}(z; \tau, \sigma)$  is the lifted  $z$ -transform of the kernel  $g(t, s)$  of  $\mathcal{G}$  with respect to its first variable  $t$  i.e.

$$\check{g}(z; \tau, \sigma) := \sum_{k \in \mathbb{Z}} g(\tau + kh, \sigma)z^{-k}, \quad \tau, \sigma \in [0, h). \quad (2.21)$$

$\check{G}(z)$  is called the *transfer function* of the lifted system  $\check{\mathcal{G}}$  (or *lifted transfer function* of  $\mathcal{G}$ ). By the above equation, the transfer function  $\check{G}(z)$  is an operator whose kernel is given by  $\check{g}(z; \tau, \sigma)$ . As a special case, if the system  $\mathcal{G}$  is LCTI (see (2.10)), then the transfer function  $\check{G}(z)$  is an operator whose kernel is given by  $\check{g}(z; \tau - \sigma, 0)$ .

**Remark 2.3.12.** *If  $\mathcal{G}$  is LCTI then  $\check{g}(z; \tau, \sigma)$  is a function of  $\tau - \sigma$  for a given  $z$  where  $\tau, \sigma \in [0, h)$ . However the converse is not true always. Consider the following system  $\check{u} = \check{G}\check{w}$  defined as*

$$\check{u}(z; \tau) = \int_0^h e^{\tau - \sigma} \check{w}(z; \sigma)d\sigma, \quad \tau \in [0, h)$$

Here,  $\check{g}(z; \tau - \sigma, 0) = e^{\tau - \sigma}$  is a function of  $\tau - \sigma$ . In the time domain we have

$$u(kh + \tau) = \int_0^h e^{\tau - \sigma} w(kh + \sigma)d\sigma = \int_{kh}^{kh+h} e^{kh + \tau - s} w(s)ds.$$

Hence,

$$u(t) = \int_{-\infty}^{\infty} e^{t-s} \mathbb{1}_{[0, h)} \left( \left\lceil \frac{t}{h} \right\rceil h - s \right) w(s)ds$$

where  $\lceil t \rceil$  is the smallest integer greater than or equal to  $t$ . Clearly, the system is not LCTI.

**Example 2.3.13.** *Let  $\mathcal{G}$  be as in Example 2.3.5 i.e.  $\mathcal{G}$  is an LCTI system with impulse response*

$$g(t) := \mathbb{1}_{[0, h)}(t).$$

Then, by using (2.21), the lifted transfer function  $\check{G}(z)$  of  $\mathcal{G}$  is an operator whose kernel is given by

$$\check{g}(z; \tau - \sigma) = \mathbb{1}_{[0, h)}(\tau - \sigma) + z^{-1} \mathbb{1}_{[-h, 0)}(\tau - \sigma).$$

Similarly, taking the  $z$ -transform of the output  $\bar{y}$  and the input  $\check{y}$  of the lifted sampler  $\check{S}$  in (2.11) results in *lifted frequency domain sampler*

$$\check{3}(\bar{y}) = \check{S}\check{3}(\check{y}) : \quad \bar{y}(z) = \check{S}(z)\check{y}(z)$$

where  $\check{S}(z) : \{[0, h) \rightarrow \mathbb{C}^{n_{\bar{y}}}\} \rightarrow \mathbb{C}^{n_{\check{y}}}$  (it is the  $z$ -transform of the lifted impulse response system of the lifted sampler  $\check{S}$ ) is given by (see [30, §4])

$$\bar{y}(z) = \check{S}(z)\check{y}(z) : \quad \bar{y}(z) = \int_0^h \check{\psi}(z; -\sigma)\check{y}(z; \sigma)d\sigma. \quad (2.22)$$

Here  $\check{\psi}(z)$  is the lifted  $z$ -transform of the sampling function  $\psi(t)$ .  $\check{S}(z)$  is called the *transfer function* of the lifted sampler  $\check{S}$  (or *lifted transfer function* of  $\mathcal{S}$ ). By the above equation, the transfer function  $\check{S}(z)$  is an operator whose kernel is given by  $\check{\psi}(z; -\sigma)$ .

Similarly, taking the  $z$ -transform of the output  $\check{u}$  and the input  $\bar{u}$  of the lifted hold  $\check{\mathcal{H}}$  in (2.12) results in *lifted frequency domain hold*

$$\check{3}(\check{u}) = \check{H}\check{3}(\bar{u}) : \quad \check{u}(z) = \check{H}(z)\bar{u}(z)$$

where  $\check{H}(z) : \mathbb{C}^{n_{\bar{u}}} \rightarrow \{[0, h) \rightarrow \mathbb{C}^{n_{\check{u}}}\}$  (it is the  $z$ -transform of the lifted impulse response system of the lifted hold  $\check{\mathcal{H}}$ ) is given by (see [30, §4])

$$\check{u}(z) = \check{H}(z)\bar{u}(z) : \quad \check{u}(z; \tau) = \check{\phi}(z; \tau)\bar{u}(z), \quad \tau \in [0, h). \quad (2.23)$$

Here  $\check{\phi}(z)$  is the lifted  $z$ -transform of the hold function  $\phi(t)$ .  $\check{H}(z)$  is called the *transfer function* of the lifted hold  $\check{\mathcal{H}}$  (or *lifted transfer function* of  $\mathcal{H}$ ).

The advantage of the lifted  $z$ -transform is more visible in the series interconnection of two linear  $h$ -time shift invariant systems  $\mathcal{G}_1$  and  $\mathcal{G}_2$  with transfer function  $\check{G}_1(z)$  and  $\check{G}_2(z)$  respectively. Using (2.14), it can be proved that the transfer function of cascade  $\mathcal{G}_1\mathcal{G}_2$  is given by the composition  $\check{G}_1(z)\check{G}_2(z)$ .

## 2.4 Lifted domain spaces and norms

So far with the exception of Key lifting formula, spaces of signals are not discussed. In this section, we review signal and system norms in various spaces. For detailed discussion on this topic see [31, 57, 3, 2] and the references therein.

### 2.4.1 Lifted signal space and norm

$\ell^2(\mathbb{B}, \mathcal{H})$  is the Hilbert space of sequences mapping  $\mathbb{B} \subseteq \mathbb{Z}$  to a Hilbert space  $\mathcal{H}$  [8, Proposition I.6.2] i.e.

$$\ell^2(\mathbb{B}, \mathcal{H}) := \{f : \mathbb{B} \rightarrow \mathcal{H} \mid \sum_{i \in \mathbb{B}} \|f_i\|_{\mathcal{H}}^2 < \infty\}.$$



Here  $f := \{f_i\}_{i \in \mathbb{B}}$ . The inner product in this space is

$$\langle x, y \rangle = \sum_{i \in \mathbb{Z}} \langle x_i, y_i \rangle_{\mathcal{H}}.$$

The norm in this space is denoted by  $\|\cdot\|_2$ .

**Remark 2.4.1.** We use the shorthand  $\ell^2(\mathbb{Z})$  to denote  $\ell^2(\mathbb{Z}, \mathcal{H})$  if the Hilbert space  $\mathcal{H}$  is understood from the context. Whenever  $\mathbb{B} = \mathbb{Z}$ , we just write  $\ell^2$ .

$L^2(\mathbb{B}, \mathbb{C}^n)$  is the Hilbert space of square integrable functions  $f : \mathbb{B} \rightarrow \mathbb{C}^n$  where  $\mathbb{B} \subseteq \mathbb{R}$  and  $n$  is a positive integer. The norm in this space is denoted by  $\|\cdot\|_2$ . This space has inner product

$$\langle x, y \rangle = \int_{\mathbb{B}} \langle x(\tau), y(\tau) \rangle_{\mathbb{C}^n} d\tau$$

**Remark 2.4.2.** If  $n$  is unambiguous/unspecified then we just write  $L^2(\mathbb{B})$ . Whenever  $\mathbb{B} = \mathbb{R}$ , we just write  $L^2$ .

Now, consider the norm of an analog signal  $f \in L^2$ ,

$$\begin{aligned} \|f\|_2^2 &= \int_{-\infty}^{\infty} \|f(t)\|_2^2 dt = \sum_{k \in \mathbb{Z}} \int_0^h \|f(kh + \tau)\|_2^2 d\tau \\ &= \sum_{k \in \mathbb{Z}} \int_0^h \|\check{f}[k](\tau)\|_2^2 d\tau \\ &= \sum_{k \in \mathbb{Z}} \|\check{f}[k]\|_{L^2[0, h]}^2 =: \|\check{f}\|_2^2. \end{aligned} \quad (2.24)$$

This defines the norm of a lifted signal  $\check{f}$ . This also shows that lifting an analog time signal  $f \in L^2$ , by definition results in a lifted signal  $\check{f}$  with the same norm in the Hilbert space  $\ell^2(\mathbb{Z}, L^2[0, h])$  ( $L^2[0, h] := L^2([0, h], \mathbb{C}^n)$ ) (see [57, chapter 10], [30], [8, Proposition I.6.2]). The space  $\ell^2(\mathbb{Z}, L^2[0, h])$  has inner product

$$\langle \check{x}, \check{y} \rangle = \sum_{k \in \mathbb{Z}} \int_0^h \langle \check{x}[k](\tau), \check{y}[k](\tau) \rangle_{\mathbb{C}^n} d\tau.$$

Equation (2.24) also implies that  $L^2$  is isomorphic to  $\ell^2(\mathbb{Z}, L^2[0, h])$  [67, §7.4]. Therefore, the norms in both of the spaces can be denoted by  $\|\cdot\|_2$ .

For a discrete signal  $\bar{f} \in \ell^2(\mathbb{Z}, \mathbb{C}^n)$ , the lifted signal is the same, therefore they both have the same norm.

We can decompose  $\ell^2(\mathbb{Z})$  as the orthogonal sum of the spaces  $\ell^2(\mathbb{Z}_l^+)$  and  $\ell^2(\mathbb{Z}_l^-)$  consisting of signals that are zero in  $\mathbb{Z}_l^-$  and  $\mathbb{Z}_l^+$  for a given  $l \in \mathbb{Z}$ . Here  $\mathbb{Z}_l^+$  is the set of all integers greater than or equal to  $l$  and  $\mathbb{Z}_l^-$  is the set of all integers smaller than  $l$ . In short,  $\ell^2(\mathbb{Z}) = \ell^2(\mathbb{Z}_l^+) \oplus \ell^2(\mathbb{Z}_l^-)$ . These spaces are important in relation to causality discussed later in Section 2.5.

The space  $L^2(\mathbb{T}, \mathcal{H})$  represents the Hilbert space consisting of functions  $p(z)$  mapping from unit circle  $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$  to a separable Hilbert space  $\mathcal{H}$ , with norm

$$\|p\|_2 := \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \|p(e^{i\theta})\|_{\mathcal{H}}^2 d\theta} < \infty.$$

If an analog lifted signal  $\check{f}$  is in  $\ell^2(\mathbb{Z}, L^2[0, h])$  then its Fourier transform  $\mathfrak{F}(\check{f})$  belongs to  $L^2(\mathbb{T}, L^2[0, h])$  because

$$\begin{aligned} \|\mathfrak{F}(\check{f})\|_2^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\check{f}(e^{i\theta})\|_{L^2[0, h]}^2 d\theta \\ &= \sum_{k \in \mathbb{Z}} \|\check{f}[k]\|_{L^2[0, h]}^2 \\ &= \|\check{f}\|_{\ell^2(\mathbb{Z}, L^2[0, h])}^2 = \|f\|_2^2 < \infty \end{aligned}$$

The above can be proved by using the fact that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\theta(m-k)} d\theta = \delta[m - k]. \quad (2.25)$$

The equivalence of  $\|\mathfrak{F}(\check{f})\|_2 = \|\check{f}\|_{\ell^2(\mathbb{Z}, L^2[0, h])}$  is known as *Parseval identity* which says that  $\ell^2(\mathbb{Z}, L^2[0, h])$  is isomorphic to  $L^2(\mathbb{T}, L^2[0, h])$ .

Similar to the analog case, the Parseval identity between discrete signal  $\bar{f} \in \ell^2(\mathbb{Z}, \mathbb{C}^n)$  and its Fourier transform  $\mathfrak{F}(\bar{f}) \in L^2(\mathbb{T}, \mathbb{C}^n)$  can be stated as

$$\|\mathfrak{F}(\bar{f})\|_2 = \|\bar{f}\|_{\ell^2(\mathbb{Z}, \mathbb{C}^n)}$$

This means that  $\ell^2(\mathbb{Z}, \mathbb{C}^n)$  is isomorphic to  $L^2(\mathbb{T}, \mathbb{C}^n)$ .

**Remark 2.4.3.** We use the shorthand  $L^2(\mathbb{T})$  to denote  $L^2(\mathbb{T}, \mathcal{H})$  if the Hilbert space  $\mathcal{H}$  is understood from the context.

The *Hardy space*  $H^2$  is the Hilbert space of analytic functions  $\tilde{f}(z) : \mathbb{C} \setminus \overline{\mathbb{D}} \rightarrow \mathcal{H}$  ( $\mathcal{H}$  is a separable Hilbert-space) such that

$$\|\tilde{f}\|_2 := \sqrt{\sup_{r>1} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\tilde{f}(re^{i\theta})\|_{\mathcal{H}}^2 d\theta \right)} < \infty$$

where  $\overline{\mathbb{D}}$  is the closed unit disk in the complex plane  $\mathbb{C}$ . The space  $H^2$  can be identified as a closed subspace of  $L^2(\mathbb{T}, \mathcal{H})$  [44, chapter 5]. The orthogonal complement of  $H^2$  in  $L^2(\mathbb{T})$  exists and it is denoted by  $(H^2)^\perp$ . This  $(H^2)^\perp$  is the Hilbert space of analytic functions  $\tilde{f}(z) : \mathbb{D} \rightarrow \mathcal{H}$  such that

$$\|\tilde{f}\|_2 := \sqrt{\sup_{r<1} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\tilde{f}(re^{i\theta})\|_{\mathcal{H}}^2 d\theta \right)} < \infty.$$

Here  $\mathbb{D}$  is the open unit disk in the complex plane  $\mathbb{C}$ .

Finally,  $z^l H^2$  denotes a Hilbert space of analytic functions  $\tilde{f}(z) : \mathbb{C} \setminus \overline{\mathbb{D}} \rightarrow \mathcal{H}$  such that  $z^{-l} \tilde{f}(z) \in H^2$ . If  $\mathcal{H}$  is either  $L^2[0, h)$  or  $\mathbb{C}^n$ , it can be shown that  $\ell^2(\mathbb{Z}_-^+, \mathcal{H})$  is isomorphic to  $z^l H^2$  via  $z$ -transform [12, §2.5]. These spaces are important in relation to causality discussed later in Section 2.5.

## 2.4.2 Adjoint systems and conjugate transfer function

In this section we discuss adjoints of lifted systems. We do this for a lifted  $h$ -time shift invariant system  $\check{G}$ , lifted hold  $\check{H}$  and lifted sampler  $\check{S}$ .

Lifting is an isometric isomorphism between  $L^2$  and  $\ell^2(\mathbb{Z}, L^2[0, h))$ , therefore it preserves inner products [67, §7.4]. Similarly, the Fourier transform is an isometric isomorphism between  $\ell^2(\mathbb{Z}, L^2[0, h))$  and  $L^2(\mathbb{T}, L^2[0, h))$  (also between  $\ell^2(\mathbb{Z}, \mathbb{C}^n)$  and  $L^2(\mathbb{T}, \mathbb{C}^n)$ ), therefore it also preserves inner products. The adjoint and Fourier transform operations commute as the inner product is preserved by the Fourier transform. Similarly, the adjoint and lifting operations commute as the inner product is preserved by the lifting.

It is shown in [30] that the kernel of adjoint  $\mathcal{G}^*$  of the system  $\mathcal{G}$  given in (2.1) is

$$g^\sim(s, t) := (g(t, s))^*$$

where  $*$  denote complex conjugate transpose. The lifted  $z$ -transform of the above given kernels are related as

$$g^\sim(z; \sigma, \tau) := g(1/\bar{z}; \tau, \sigma)^*.$$

The system which has the kernel  $g^\sim(z; \sigma, \tau)$  is denoted by  $\check{G}^\sim(z)$  and it is known as the *conjugate* of transfer function  $\check{G}(z)$ . It is shown in [30] that for  $z = e^{j\theta}$ , the conjugate  $\check{G}^\sim(e^{j\theta})$  is the adjoint of  $\check{G}(e^{j\theta})$  with respect to  $L^2[0, h)$ .

Again by [30], the kernel  $\phi(t)$  of the adjoint  $\mathcal{S}^*$  of the sampler  $\mathcal{S}$  given in (2.2) is

$$\phi(t) := \psi(-t)^*$$

and the kernel  $\phi(z; \tau)$  of the conjugate  $\check{S}^\sim(z)$  of the transfer function  $\check{S}(z)$  given in (2.22) is

$$\phi(z; \tau) := \psi(1/\bar{z}; -\tau)^*.$$

Note that the adjoint  $\mathcal{S}^*$  of a sampler  $\mathcal{S}$  is a hold [30].

Similarly, the kernel  $\psi(t)$  of the adjoint  $\mathcal{H}^*$  of the hold  $\mathcal{H}$  given in (2.4) is

$$\psi(t) := \phi(-t)^*$$

and the kernel  $\psi(z; \tau)$  of the conjugate  $\check{H}^\sim(z)$  of the transfer function  $\check{H}(z)$  given in (2.23) is

$$\psi(z; \tau) := \phi(1/\bar{z}; \tau)^*.$$

Note that the adjoint  $\mathcal{H}^*$  of hold  $\mathcal{H}$  is a sampler [30].

### 2.4.3 System norms

In this section, we review the definition of some standard system norms used in sampled-data system theory for shift invariant discrete systems and linear  $h$ -time shift invariant systems including samplers and holds. Norms defined in this section are very standard and discussed in great detail in [30, 3, 2, 12]. In this section  $\tilde{\mathbb{S}}$  means either  $L^2[0, h)$  or  $\mathbb{C}^n$  norm. Also,  $\tilde{\mathbb{R}}$  denotes the space (with induced 2-norm) of all bounded operators and  $\tilde{\mathbb{R}}_{HS}$  denotes the space of all Hilbert-Schmidt operators, mapping  $\tilde{\mathbb{S}}_i$  to  $\tilde{\mathbb{S}}_o$ . Here  $\tilde{\mathbb{S}}_i$  and  $\tilde{\mathbb{S}}_o$  can be  $L^2[0, h)$  or  $\mathbb{C}^n$ . Also, in this section  $\mathcal{G}$  can be a linear shift invariant discrete systems or a linear  $h$ -time shift invariant systems including samplers and holds.

#### $L^\infty$ system norm

The space  $L^\infty$  can be defined as the space of linear shift invariant discrete systems and linear  $h$ -time shift invariant systems with norm defined as [2, 30]:

$$\|\mathcal{G}\|_{L^\infty} := \operatorname{ess\,sup}_{\theta \in [-\pi, \pi]} \|\tilde{G}(e^{j\theta})\| < \infty \quad (2.26)$$

where  $\tilde{G}(e^{j\theta})$  is the (lifted) transfer function of  $\mathcal{G}$  and  $\|\tilde{G}(e^{j\theta})\|$  is given by

$$\|\tilde{G}(e^{j\theta})\| = \sup_{\|x\|_2=1} \|\tilde{G}(e^{j\theta})x\|_2, \quad x \in \tilde{\mathbb{S}}.$$

The above definition of the  $L^\infty$  norm is equivalent to the induced norm interpretation given as

$$\|\mathcal{G}\|_{L^\infty} := \sup_{\|x\|_{\mathbb{S}_i}=1} \|\mathcal{G}x\|_{\mathbb{S}_o}$$

where  $\mathbb{S}_i$  and  $\mathbb{S}_o$  can be  $L^2(\mathbb{R})$  or  $\ell^2(\mathbb{Z})$  depending upon whether the signal in concern is analog or discrete respectively [30, §V.C]. For example, an analog system in  $L^\infty$  is a bounded operator from  $L^2(\mathbb{R})$  to  $L^2(\mathbb{R})$ , a sampler in  $L^\infty$  is a bounded operator from  $L^2(\mathbb{R})$  to  $\ell^2(\mathbb{Z})$ , a hold in  $L^\infty$  is a bounded operator from  $\ell^2(\mathbb{Z})$  to  $L^2(\mathbb{R})$ , and a discrete system in  $L^\infty$  is bounded operator from  $\ell^2(\mathbb{Z})$  to  $\ell^2(\mathbb{Z})$ . An interesting fact is that the ideal sampler does not belong to  $L^\infty$  [30].

The following result is important for later chapters.

**Lemma 2.4.4** (Essentially from [31]). *If  $\mathcal{G} \in L^\infty$ , then its transfer function  $\tilde{G}(e^{j\theta})$  is a bounded operator at almost all  $\theta \in [-\pi, \pi]$ .*

*Proof.* By (2.26),  $\|\mathcal{G}\|_{L^\infty}$  is finite iff  $\|\tilde{G}(e^{j\theta})\|$  is finite at almost all  $\theta$ .  $\square$

#### $z^l H^\infty$ system norm

The *Hardy space*  $H^\infty$  is the set of analytic transfer functions  $\tilde{G}(z) : \mathbb{C} \setminus \bar{\mathbb{D}} \rightarrow \tilde{\mathbb{R}}$  with finite norm given by

$$\|\tilde{G}\|_{H^\infty} := \operatorname{ess\,sup}_{z \in \mathbb{C} \setminus \bar{\mathbb{D}}} \|\tilde{G}(z)\|$$

where  $\overline{\mathbb{D}}$  is the closed unit disk in the complex plane  $\mathbb{C}$  and  $\|\tilde{G}(z)\|$  is given by

$$\|\tilde{G}(z)\| = \sup_{\|x\|_2=1} \|\tilde{G}(z)x\|_2 \quad x \in \tilde{\mathbb{S}}$$

Similar to the  $H^2$  signal norm case, the space  $H^\infty$  can be considered as a closed subspace of  $L^\infty$ .

For a given  $l$ ,  $z^l H^\infty$  is the subspace of  $L^\infty$  that contains all transfer functions  $\tilde{G}(z)$  such that  $z^{-l} \tilde{G}(z) \in H^\infty$  [30].

### $L^2$ system norm

The space  $L^2$  can be defined as the space of linear shift invariant discrete system and linear  $h$ -time shift invariant systems with norm defined as [2, 30]:

$$\|\mathcal{G}\|_{L^2} := \sqrt{\frac{1}{2\pi h} \int_{-\pi}^{\pi} \|\tilde{G}(e^{j\theta})\|_{HS}^2 d\theta} \quad (2.27)$$

$$= \sqrt{\frac{1}{h} \sum_{k \in \mathbb{Z}} \|\tilde{\mathcal{G}}[k]\|_{HS}^2} < \infty \quad (2.28)$$

where  $\tilde{G}(e^{j\theta})$  is the (lifted) transfer function of  $\mathcal{G}$ ,  $\tilde{\mathcal{G}}[k]$  is the lifted impulse response of  $\mathcal{G}$ , and  $\|\cdot\|_{HS}$  stand for Hilbert-Schmidt norm of a linear operator [69, §8.1].

The following result is important for later chapters.

**Lemma 2.4.5** (Essentially from [31]). *If  $\mathcal{G} \in L^2$ , then its transfer function  $\tilde{G}(e^{j\theta})$  is a Hilbert-Schmidt operator at almost all  $\theta \in [-\pi, \pi]$*

*Proof.* By (2.27),  $\|\mathcal{G}\|_{L^2}$  is finite iff  $\|\tilde{G}(e^{j\theta})\|_{HS}$  is finite at almost all  $\theta$ . □

Neither  $L^2$  nor  $L^\infty$  is a subset of the other. However, in the case of transfer functions that have uniformly bounded rank on the unit circle, we can state the following Lemma [30].

**Lemma 2.4.6.** *If  $\mathcal{G}$  is in  $L^\infty$  and  $\text{rank } \tilde{G}(e^{j\theta}) \leq r$  for almost all  $\theta \in [-\pi, \pi]$ , then  $\mathcal{G} \in L^2$ . Here  $r$  is a non-negative integer.*

*Proof.* See [30, proposition 5.3] □

Lemma 2.4.6 says that if  $\text{rank } \tilde{G}(e^{j\theta})$  is uniformly bounded for almost all  $\theta \in [-\pi, \pi]$ , then  $\tilde{G} \in L^\infty$  implies  $\tilde{G} \in L^2$ . The lifted output (or the lifted input) of a sampler (or a hold) given by (2.22) ( or (2.23)) belongs to  $\mathbb{C}^{n_{\tilde{y}}}$  ( or  $\mathbb{C}^{n_{\tilde{u}}}$  ). Therefore, if samplers or holds are in  $L^\infty$  then they are in  $L^2$  by Lemma 2.4.6. Moreover, if the sampling function of a sampler and the hold function of a hold are in  $L^2$  then the following result holds.

**Lemma 2.4.7.** *Consider a sampler and a hold given by (2.2) and (2.4) respectively. If the sampling function  $\psi$  of the sampler and the hold function  $\phi$  of the hold belong to  $L^2$ , then the sampler and hold belong to  $L^\infty \cap L^2$ .*

*Proof.* See Appendix 2.A (page 36). □

For a generalization of Lemma 2.4.7, see [30, §VI(A)] and the references therein.

### $z^l \mathbb{H}^2$ system norm

The *Hardy space*  $\mathbb{H}^2$  is the set of analytic transfer functions  $\tilde{G}(z) : \mathbb{C} \setminus \overline{\mathbb{D}} \rightarrow \tilde{\mathbb{R}}_{HS}$  with finite norm given by

$$\|\mathcal{G}\|_{L^2} := \sqrt{\frac{1}{2\pi h} \int_{-\pi}^{\pi} \|\tilde{G}(e^{j\theta})\|_{HS}^2 d\theta} = \sqrt{\frac{1}{h} \sum_{k \in \mathbb{N}} \|\tilde{\mathcal{G}}[k]\|_{HS}^2}$$

Note that the summation in the above is over non-negative integers only.

Similar to the  $H^2$  signal norm case, the space  $\mathbb{H}^2$  can be considered as a closed subspace of  $L^2$  (see also [57, Theorem 12.2.1]).

For a given  $l$ ,  $z^l \mathbb{H}^2$  is the subspace of  $L^2$  that contains all transfer functions  $\tilde{G}(z)$  such that  $z^{-l} \tilde{G}(z) \in \mathbb{H}^2$ .

**Remark 2.4.8.** *The standard  $\mathbb{H}^2$  norm of a discrete system given in [71] does not contain the factor  $\sqrt{\frac{1}{h}}$ . However, this scaling is constant, therefore it does not affect the optimization results in this thesis.*

## 2.5 Causality

Causality loosely speaking says that the effect of an event must happen after the event has occurred. This section describes the meaning of causality for different systems in the sampled-data setup. Causality for  $h$ -time shift invariant system is not trivial as we will see in this section.

An analog system  $\mathcal{G}$  which maps an analog signal to an analog signal, is defined (*classic*) *causal* if

$$\Pi_T \mathcal{G}(I - \Pi_T) = 0, \quad \forall T \in \mathbb{R}, \quad (2.29)$$

where the *truncation operator*  $\Pi_T$  is defined as

$$(\Pi_T u)(t) := \begin{cases} u(t) & t < T \\ 0 & t \geq T \end{cases}.$$

If the analog system is continuous time invariant then it is sufficient to check the above condition at only one time instant (say  $T = 0$ ) to establish (classic) causality.

A standard result is that a linear analog system  $\mathcal{G}$  is (classic) causal iff it is given by

$$y = \mathcal{G}u : \quad y(t) = \int_{-\infty}^{\infty} g(t, s)w(s)ds \quad (2.30)$$

where  $g(t, s) = 0 \forall s > t$ .

Moreover, if the analog system is linear, (classic) causal and  $h$ -time shift invariant then we have the following standard result.

**Lemma 2.5.1.** *Let  $\mathcal{G}$  be a linear  $h$ -time shift invariant system given by (2.1). Then  $\mathcal{G}$  is (classic) causal iff its impulse response system  $\check{\mathcal{G}}[k]$  defined in (2.9) has the following form*

$$x_o = \check{\mathcal{G}}[k]x_i : \quad x_o(\tau) = \begin{cases} 0 & k < 0 \\ \int_0^\tau g(\tau, \sigma)x_i(\sigma) d\sigma & k = 0 \\ \int_0^h g(kh + \tau, \sigma)x_i(\sigma) d\sigma & k > 0 \end{cases}$$

where  $\tau, \sigma \in [0, h)$ .

*Proof.* See Appendix 2.A (page 36). □

**Remark 2.5.2.** *Note that in Lemma 2.5.1,  $\check{\mathcal{G}}[0]$  has kernel  $g(\tau, \sigma)\mathbb{1}_{[0, \infty)}(\tau - \sigma)$  for all  $\tau, \sigma \in [0, h)$ . This implies not only the index  $k$  (i.e.  $\check{\mathcal{G}}[k] = 0$  for  $k < 0$ ) but also the structure of  $\check{\mathcal{G}}[0]$  in the impulse response system  $\check{\mathcal{G}}[k]$  plays an important role in identifying the (classic) causality of a linear, (classic) causal and  $h$ -time shift invariant system  $\mathcal{G}$ .*

We call an analog signal  $y$  *causal* if  $y(t) = 0, \forall t < 0$ . Now, it follows from the above definition that an LCTI system  $\mathcal{G}$  is (classic) causal if the system output is causal for all causal inputs i.e. if its impulse response  $g(t) = 0 \forall t < 0$ .

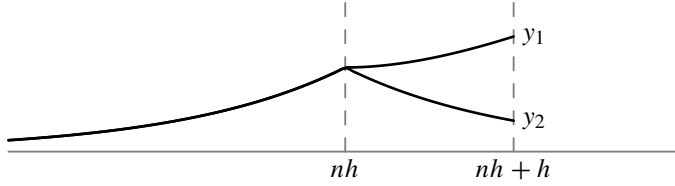
Similarly a discrete system  $\check{\mathcal{G}}$  which maps a discrete signal to a discrete signal, is defined (classic) *causal* if

$$\bar{\Pi}_k \check{\mathcal{G}}(I - \bar{\Pi}_k) = 0 \quad \forall k \in \mathbb{Z}, \quad (2.31)$$

where the *truncation operator*  $\bar{\Pi}_k$  is defined as

$$(\bar{\Pi}_k \bar{u})[n] := \begin{cases} \bar{u}[n] & n < k \\ 0 & n \geq k \end{cases}$$

If the discrete system is shift invariant then it is sufficient to check the above condition at only one instant (say  $k = 0$ ) to establish (classic) causality. We call a discrete signal  $\bar{y}$  *causal* if  $\bar{y}[n] = 0 \forall n < 0$ . Similarly, for a given integer  $l$ , we call a discrete signal  $\bar{y}$   *$l$ -causal* if  $\bar{y}[n + l]$  is causal. Therefore, it follows from the above definition that the linear shift invariant discrete system given in (2.7) is causal if the system output is casual for all causal inputs i.e. if its impulse response  $\bar{w}(k) = 0$  for all integers  $k < 0$ .



**Figure 2.6:** Two analog input signals  $y_1$  and  $y_2$  used in Example 2.5.4 which are same up to time  $nh$  but different afterwards.

Similarly, we can define (classic) causality of samplers and holds. A linear  $h$ -time shift invariant sampler  $\mathcal{S}$  is defined (classic) causal if

$$\bar{\Pi}_k \mathcal{S}(I - \Pi_{kh}) = 0 \quad \forall k \in \mathbb{Z}, \quad (2.32)$$

where as a linear  $h$ -time shift invariant hold  $\mathcal{H}$  is defined (classic) causal if

$$\Pi_{kh} \mathcal{H}(I - \bar{\Pi}_k) = 0 \quad \forall k \in \mathbb{Z}. \quad (2.33)$$

The following lemma explains how the (classic) causality of a linear  $h$ -time shift invariant sampler is related to its sampling function.

**Lemma 2.5.3.** *Let a linear  $h$ -time shift invariant sampler  $\mathcal{S}$  be given by (2.2). Then  $\mathcal{S}$  is (classic) causal iff its sampling function  $\psi(t) = 0 \forall t \leq -h$ .*

Although the above lemma is standard (see [30, §VI.2]), the proof is given in Appendix 2.A (page 37) for reference purpose.

Lemma 2.5.3 says that a (classic) causal linear  $h$ -time shift invariant sampler  $\mathcal{S}$  equivalently is given by

$$\bar{y}[n] = \int_{-\infty}^{nh+h} \psi(nh-t)y(t)dt. \quad (2.34)$$

Therefore, the output  $\bar{y}[n]$  of the sampler depends upon the input within time interval  $(-\infty, nh+h)$ . To understand this better, let us consider the following example.

**Example 2.5.4.** *Assume that  $\mathcal{S}$  is a (classic) causal sampler. Also assume that  $y_1$  and  $y_2$  are two analog signals which are the same upto time  $t = nh$  but different afterwards (Figure 2.6). Also denote  $\bar{y}_1 = \mathcal{S}y_1$  and  $\bar{y}_2 = \mathcal{S}y_2$ . Since  $y_1(t) \neq y_2(t) \forall t \in (nh, nh+h]$ , using (2.34) we have typically that  $y_1[n] \neq y_2[n]$ . Thus the present output of the sampler depends upon the future inputs.*

*Now, consider a (classic) causal system  $\mathcal{G}$ . Denote  $u_1 = \mathcal{G}y_1$  and  $u_2 = \mathcal{G}y_2$ . Even though,  $y_1(t) \neq y_2(t) \forall t \in (nh, nh+h]$ , we have that  $u_1(nh) = u_2(nh)$ . Thus the present output of the system does not depends upon the future inputs.*

Example 2.5.4 shows that the (classic) causal sampler output depends not only on the present and the past inputs but also on the future inputs in time interval



$[nh, (n + 1)h)$ . If the present output of a system depends upon the present and past inputs only then that system is called *input/output causal*. However this definition of input/output causality is not equivalent to the classic causality of sampler because the present output at  $n$  depends upon the future inputs in time interval  $[nh, (n + 1)h)$ . If we need that the output  $\bar{y}[n]$  of the sampler depend upon the present and the past inputs only, then we need *strict (classic) causality* of the sampler which by definition means

$$\bar{\Pi}_{k-1}\mathcal{S}(I - \Pi_{kh}) = 0 \quad k \in \mathbb{Z}.$$

The advantage of having (2.32) as the definition of sampler's causality is that it is aligned with the (classic) causality definition of the hold, the LCTI systems and the discrete systems. The difference between classic causality and input/output causality is explained the following example.

**Example 2.5.5.** *The non-equivalence of input/output causality with the classic causality can be shown with an example of ideal zero order hold  $\mathcal{H}_{iz}$  given in (2.6) with hold function  $\mathbb{1}_{[0,h)}(t)$ . Here the step function  $\mathbb{1}_{[0,h)}(t)$  is 1 if  $t \in [0, h)$ , otherwise it is zero. This hold is classic causal as well as input/output causal. The adjoint of  $\mathcal{H}_{iz}$  is a sampler with sampling function  $\mathbb{1}_{[0,h)}(-t)$  and it is anti-causal in input/output sense. However by definition of causality in (2.32) both of the systems are (classic) causal.*

*Similarly, a zero order hold  $\mathcal{H}_z$  given in (2.5) is classic and input/output causal. However, its adjoint which is a sampler, is classic causal but not input/output causal.*

The following lemma explains how the (classic) causality of a linear  $h$ -time shift invariant hold is related to its hold function.

**Lemma 2.5.6.** *Let a linear  $h$ -time shift invariant hold  $\mathcal{H}$  be given by (2.4). Then  $\mathcal{H}$  is (classic) causal iff its hold function  $\phi(t) = 0 \forall t \leq 0$ .*

Although the above lemma is standard (see [30, §VI.2]), the proof is given in Appendix 2.A (page 37) for reference purpose.

Lemma 2.5.6 says that a (classic) causal linear  $h$ -time shift invariant hold  $\mathcal{H}$  is given by

$$u(t) = \sum_{n \leq \lfloor \frac{t}{h} \rfloor, n \in \mathbb{Z}} \phi(t - nh)\bar{u}[n] \quad (2.35)$$

where  $\lfloor t \rfloor$  means the greatest integer less than or equal to  $t$ .

Equation (2.35) implies that the hold is both classic causal and input/output causal. The criterion given in Lemma 2.5.6 for the causality of holds looks like the criterion of LCTI systems and shift invariant discrete systems. In fact the notion of input/output causality and the classic causality are the same for holds, LCTI systems and shift invariant discrete systems.

Now, we define causality of lifted systems. A *shift invariant (lifted) system*  $\tilde{\mathcal{G}}$  (can be lifted analog system  $\tilde{\mathcal{G}}$  or discrete system  $\tilde{\mathcal{G}}$  or lifted sampler  $\tilde{\mathcal{G}}$  or lifted hold  $\tilde{\mathcal{G}}$ ) is defined (*lifted*) *causal* if

$$\tilde{\Pi}_k \tilde{\mathcal{G}}(I - \tilde{\Pi}_k) = 0, \quad \forall k \in \mathbb{Z}. \quad (2.36)$$

where the *truncation operator*  $\tilde{\Pi}_k$  is defined as

$$(\tilde{\Pi}_k \tilde{u})[n] := \begin{cases} \tilde{u}[n] & n < k \\ 0 & n \geq k \end{cases}.$$

where the lifted  $\tilde{u}$  is  $\check{u}$  if the signal is analog or  $\bar{u}$  if the signal is discrete.

**Remark 2.5.7.** *Lifting the definition of (classic) causality of a discrete system, sampler and hold given in (2.31), (2.32) and (2.33) respectively, leads to the (lifted) causality definition given in (2.36). However, for an analog system  $\mathcal{G}$  we have*

$$\tilde{\Pi}_k \check{u}[n](\tau) := \begin{cases} \check{u}[n](\tau) & n < k \\ 0 & n \geq k \end{cases}.$$

Therefore,  $\tilde{\Pi}_k \check{\mathcal{G}}(I - \tilde{\Pi}_k) = 0$  in (2.36) means  $\Pi_{kh} \mathcal{G}(I - \Pi_{kh}) = 0$ . This is not equivalent to the classic causality definition given in (2.29) because there it must be true for every time instant  $T \in \mathbb{R}$  not just at  $kh$ . In other words, classic causality implies lifted causality but not otherwise for an analog system in general. However, if  $\mathcal{G}$  is continuous time invariant then the classic causality and the lifted causality are equivalent. Most of the times in this thesis the signal models  $\mathcal{G}$  are LCTI therefore the classic causality and the lifted causality are equivalent. However, in general, care must be taken in the study of causality [30, remark 6.1].

The (lifted) causality leads the following simple condition on the lifted impulse response system.

**Lemma 2.5.8.** *Given a lifted system  $\tilde{\mathcal{G}}$  (can be lifted analog system  $\tilde{\mathcal{G}}$  or discrete system  $\tilde{\mathcal{G}}$  or lifted sampler  $\tilde{\mathcal{G}}$  or lifted hold  $\tilde{\mathcal{G}}$ ). Now,  $\tilde{\mathcal{G}}$  is (lifted) causal iff its lifted impulse response system  $\check{\mathcal{G}}[k] = 0 \forall k < 0$ .*

*Proof.* We have

$$\check{y}[k] = \sum_{i \in \mathbb{Z}} \check{\mathcal{G}}[k - i] \check{w}[i], \quad k \in \mathbb{Z}$$

For (lifted) causality we must have zero output for input  $\check{w}[i] = 0 \forall i < 0$  i.e.

$$\check{y}[k] = \sum_{i=0}^{\infty} \check{\mathcal{G}}[k - i] \check{w}[i] = 0, \quad \forall k < 0$$

This implies  $\check{\mathcal{G}}[k] = 0 \forall k < 0$ . □

As a special case if  $\mathcal{G}$  is an analog (lifted) causal system then the impulse response system  $\check{\mathcal{G}}[k]$  defined in (2.9) has the following form

$$x_o = \check{\mathcal{G}}[k]x_i : \quad x_o(\tau) = \begin{cases} 0 & k < 0 \\ \int_0^h g(\tau, \sigma)x_i(\sigma) d\sigma & k = 0 \\ \int_0^h g(kh + \tau, \sigma)x_i(\sigma) d\sigma & k > 0 \end{cases}$$

where  $\tau, \sigma \in [0, h)$ . Here  $\check{\mathcal{G}}[0]$  does not have the structure needed in Lemma 2.5.1 for classic causality. This again leads to the fact that the lifted causality is not same as the classic causality for analog systems in general.

Now, we show that if  $\mathcal{G}$  is LCTI then  $\check{\mathcal{G}}[0]$  have the structure required in Lemma 2.5.1 for classic causality.

**Lemma 2.5.9.** *Let  $\mathcal{G}$  be a linear analog system. If  $\mathcal{G}$  is LCTI and lifted causal then the impulse response system  $\check{\mathcal{G}}[k]$  defined in (2.9) has the following form*

$$x_o = \check{\mathcal{G}}[k]x_i : \quad x_o(\tau) = \begin{cases} 0 & k < 0 \\ \int_0^\tau g(\tau, \sigma)x_i(\sigma) d\sigma & k = 0 \\ \int_0^h g(kh + \tau, \sigma)x_i(\sigma) d\sigma & k > 0 \end{cases}$$

where  $\tau, \sigma \in [0, h)$ .

*Proof.* It follows from Lemma 2.5.8 that  $g(kh + \tau - \sigma) = 0 \forall k < 0$ . This means  $g(t) = 0$  for  $t < 0$ . This further implies

$$g(\tau - \sigma) = 0, \text{ if } \sigma > \tau$$

□

Lemma 2.5.9 shows the lifted causality is the same as classic causality for LCTI analog systems.

In summary, we have seen that the lifted, classic and input/output are not always equivalent (see Example 2.5.5 and Remark 2.5.7).

**Remark 2.5.10.** *In this thesis, we consider the lifted causality of the systems most of the time. Therefore, from now on, whenever we refer to causality we mean (lifted) causality in this thesis.*

Similar to the discrete signals, we call a lifted signal  $\tilde{y}$  causal if  $\tilde{y}[n] = 0, \forall n < 0$ . Similarly, for a given integer  $l$ , we call a lifted signal  $\tilde{y}$   $l$ -causal if  $\tilde{y}[n+l]$  is causal.

For a given integer  $l$ , a shift invariant lifted system  $\tilde{\mathcal{G}}$  is defined  $l$ -causal or relaxed causal if

$$\tilde{\Pi}_{k-l}\tilde{\mathcal{G}}(I - \tilde{\Pi}_k) = 0, \quad k \in \mathbb{Z}$$

This means that for a causal lifted input, the lifted output of an  $l$ -causal system is  $l$ -causal. In other words, the present lifted output at  $k$  depends upon all lifted

inputs up to  $k + l$ . Here, a causal system means  $l = 0$  and a strictly causal system means  $l = -1$ .

Anti-causality is just opposite of the causality. A system is defined *anti-causal* if its adjoint (see Section 2.4.2) is causal. Note that the relaxed causal systems with  $l > 0$  are neither causal nor anti-causal.

## 2.6 Stability

Stability of a system loosely speaking means that the output of the system is bounded for a bounded input. At any point of time, we never want that due to some bounded noise or external disturbances the output of our system grow unboundedly. Therefore, stability is desired for each component of the sampled-data setup given in Figure 2.3. In sampled-data system theory stability is defined as follows [30].

**Definition 2.6.1.** *A linear  $h$ -time shift invariant system  $\mathcal{G}$  is defined stable if its lifted system  $\tilde{\mathcal{G}}$  is a bounded operator from  $\ell^2(\mathbb{Z})$  to  $\ell^2(\mathbb{Z})$ .*

It is clear from Definition 2.6.1 that an analog system  $\mathcal{G}$  is stable if it is a bounded operator from  $L^2(\mathbb{R})$  to  $L^2(\mathbb{R})$ , a discrete system  $\bar{W}$  is stable if it is a bounded operator from  $\ell^2(\mathbb{Z})$  to  $\ell^2(\mathbb{Z})$ , a sampler  $\mathcal{S}$  is stable if it is a bounded operator from  $L^2(\mathbb{R})$  to  $\ell^2(\mathbb{Z})$ , and a hold  $\mathcal{H}$  is stable if it is a bounded operator from  $\ell^2(\mathbb{Z})$  to  $L^2(\mathbb{R})$ .

**Proposition 2.6.2.** *A linear  $h$ -time shift invariant system  $\mathcal{G}$  is stable iff its induced  $L^\infty$  norm is finite i.e.  $\|\mathcal{G}\|_{L^\infty} < \infty$ .*

*Proof.* See [30, theorem 6.1]. □

Note that an ideal sampler does not belong to  $L^\infty$  therefore it is not stable [30].

**Example 2.6.3.** *Let an LCTI analog system  $\mathcal{G} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  be given by  $G(s) = C(sI - A)^{-1}B + D$  in the Laplace domain, where  $A, B, C, D$  are constant matrices. Then  $\mathcal{G} \in L^\infty$  (hence stable) if  $A$  does not have eigenvalues on the imaginary axis [11, §2.2]. In this case, the impulse response  $g(t)$  of the system  $\mathcal{G}$  has a causal part determined by the poles of the system in left half plane and anti-causal part determined by the poles of the system in right half plane. For example, let  $A = \begin{bmatrix} A_c & 0 \\ 0 & A_a \end{bmatrix}$ ,  $B = \begin{bmatrix} B_c \\ B_a \end{bmatrix}$ , and  $C = [C_c \quad C_a]$  where  $A_c$  has eigenvalues strictly in the left half plane and  $A_a$  has eigenvalues strictly in the right half plane. Then the impulse response of the system is given by  $g(t) = g_c(t) + g_a(t)$  where*

$$g_c(t) = C_c e^{A_c t} B_c \mathbb{1}_{[0, \infty)}(t) + D \delta(t)$$

$$g_a(t) = C_a e^{A_a t} B_a \mathbb{1}_{(-\infty, 0)}(t).$$

Here  $g_c(t)$  is the causal part and  $g_a(t)$  is the anti-causal part of the impulse response  $g(t)$ . For more details see [64, §3.4.2]. Since  $\int_{-\infty}^{\infty} |g(t)| dt$  is defined, the system is stable [15, theorem 0.11].

Let a linear shift-invariant discrete system  $\bar{\mathcal{G}} : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  be given by  $\bar{G}(z) = C(zI - A)^{-1}B + D$  in the  $z$ -domain where  $A, B, C, D$  are constant matrices. Similar to the analog system case, it can be proved that  $\bar{\mathcal{G}} \in \mathcal{L}^\infty$  (hence stable) if  $A$  does not have eigenvalues on the unit circle. Here, the impulse response has a causal part determined by the poles (strictly) in the unit circle and a strictly anti-causal part determined by the poles (strictly) outside the unit circle.

### 2.6.1 Stability with causality

It is shown in [30] that if a (lifted) system  $\tilde{G}$  (analog, discrete, sampler or hold) is  $l$ -causal and stable then  $\tilde{G}$  belongs to  $z^l \mathcal{H}^\infty$ . The proof of the following proposition can be found in [30].

**Proposition 2.6.4.** *A bounded linear  $h$ -time shift invariant system  $\tilde{G} : H^2 \rightarrow z^l H^2$  has induced norm  $\|\tilde{G}\|_{\mathcal{L}^\infty}$  and  $\tilde{G} \in z^l \mathcal{H}^\infty$ .*

**Example 2.6.5.** *Let an LCTI analog system  $\mathcal{G} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  be given by  $G(s) = C(sI - A)^{-1}B + D$  in the Laplace domain, where  $A, B, C, D$  are constant matrices. Then  $\mathcal{G} \in \mathcal{H}^\infty$  if  $A$  is Hurwitz (i.e. all eigenvalues of  $A$  have strictly negative real part) [16, §2.3].*

Let a linear shift-invariant discrete system  $\bar{\mathcal{G}} : \ell^2 \rightarrow \ell^2$  be given by  $\bar{G}(z) = C(zI - A)^{-1}B + D$  in the  $z$ -domain, where  $A, B, C, D$  are constant matrices. Similar to the analog system case, it can be proved that  $\bar{\mathcal{G}} \in \mathcal{H}^\infty$  if  $A$  is Schur (i.e. all eigenvalues of  $A$  are strictly in the unit circle) [14, §4.3].

## 2.7 Concluding remarks

In this chapter we reviewed some fundamentals of the sampled-data system. We also reviewed the concept of lifting, lifting transforms, lifted transfer functions, lifted signal and system norms.

## 2.A Proofs

*Proof of (2.2).* A linear sampler is given by

$$\bar{y}[k] := \int_{-\infty}^{\infty} \psi_1(kh, s)y(s) ds$$

where  $\psi_1$  is the kernel of the sampler. Since the sampler is  $h$ -time shift invariant we have that  $\psi_1(kh, s) = \psi_1(kh + mh, s + mh)$  for any integer  $m$ . In particular for  $m = -k$ , we have

$$\bar{y}[k] := \int_{-\infty}^{\infty} \psi_1(0, s - kh)y(s) ds.$$

Defining  $\psi(kh - s) := \psi_1(0, s - kh)$ , we have the desired form.  $\square$

*Proof of (2.4).* The proof is essentially from [27]. Let  $\phi$  be the output of a hold  $\mathcal{H}$  with input  $\bar{\delta}$  i.e.

$$\phi := \mathcal{H}\bar{\delta}$$

Then by  $h$ -shift invariance of the hold, if the input is delayed by an integer  $k$  i.e.  $\bar{\delta}[\cdot - k]$ , then the output is  $\phi(\cdot - kh)$ . Since  $\bar{u}[n] = \sum_{k \in \mathbb{Z}} \bar{\delta}[n - k]\bar{u}[k]$ , by linearity we have

$$\mathcal{H}\bar{u} := \mathcal{H} \sum_{k \in \mathbb{Z}} \bar{\delta}[\cdot - k]\bar{u}[k] = \sum_{k \in \mathbb{Z}} \phi(t - kh)\bar{u}[k]$$

□

*Proof of Corollary 2.2.3.* Using (2.4) and (2.7), we have

$$\begin{aligned} u = \mathcal{H}\bar{W}\bar{y} \quad : \quad u(t) &= \sum_{n \in \mathbb{Z}} \phi(t - nh) \left( \sum_{k \in \mathbb{Z}} \bar{w}[n - k]\bar{y}[k] \right) \\ &= \sum_{k \in \mathbb{Z}} \left( \sum_{n \in \mathbb{Z}} \phi(t - nh)\bar{w}[n - k] \right) \bar{y}[k] \\ &= \sum_{k \in \mathbb{Z}} \left( \sum_{i \in \mathbb{Z}} \phi(t - (i + k)h)\bar{w}[i] \right) \bar{y}[k] \end{aligned}$$

Therefore,  $\mathcal{H}\bar{W}$  has hold function  $\sum_{i \in \mathbb{Z}} \phi(t - ih)\bar{w}[i]$ .

Similarly, using (2.2) and (2.7), we have

$$\begin{aligned} \bar{u} = \bar{W}S y \quad : \quad \bar{u}[n] &= \sum_{k \in \mathbb{Z}} \bar{w}[n - k] \int_{-\infty}^{\infty} \psi(kh - s)y(s) ds \\ &= \int_{-\infty}^{\infty} \left( \sum_{k \in \mathbb{Z}} \bar{w}[n - k]\psi(kh - s) \right) y(s) ds \\ &= \int_{-\infty}^{\infty} \left( \sum_{i \in \mathbb{Z}} \bar{w}[i]\psi(nh - ih - s) \right) y(s) ds \end{aligned}$$

Therefore,  $\bar{W}S$  has sampling function  $\sum_{i \in \mathbb{Z}} \bar{w}[i]\psi(t - ih)$ . □

*Proof of Corollary 2.3.9.* Given that  $\check{f}(e^{j\theta}; \tau)$  of the lifted signal  $\check{f}$  exists. If  $f$  is real, then  $\check{f}$  is real. In that case,

$$\overline{\check{f}(e^{-j\theta}, \tau)} = \overline{\sum_k f(kh + \tau)e^{j\theta k}} = \sum_k f(kh + \tau)e^{-j\theta k} = \check{f}(e^{j\theta}, \tau).$$

On the other hand if  $\check{f}(e^{j\theta}; \tau) = \overline{\check{f}(e^{-j\theta}, \tau)}$  is true then,

$$\begin{aligned}
& \check{f}(e^{j\theta}; \tau) = \overline{\check{f}(e^{-j\theta}, \tau)} \\
& \Rightarrow \int_{-\pi}^{\pi} \check{f}(e^{j\theta}; \tau) e^{j\theta k} d\theta = \int_{-\pi}^{\pi} \overline{\check{f}(e^{-j\theta}, \tau)} e^{j\theta k} d\theta \\
& \Rightarrow \int_{-\pi}^{\pi} \check{f}(e^{j\theta}; \tau) e^{j\theta k} d\theta = \int_{-\pi}^{\pi} \overline{\check{f}(e^{-j\theta}, \tau) e^{-j\theta k}} d\theta \\
& \Rightarrow \int_{-\pi}^{\pi} \check{f}(e^{j\theta}; \tau) e^{j\theta k} d\theta = \int_{-\pi}^{\pi} \overline{\check{f}(e^{j\theta}, \tau) e^{j\theta k}} d\theta, \text{ changed } \theta = -\theta \\
& \Rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \check{f}(e^{j\theta}; \tau) e^{j\theta k} d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\check{f}(e^{j\theta}, \tau) e^{j\theta k}} d\theta \\
& \Rightarrow \check{f}[k](\tau) = \overline{\check{f}[k](\tau)}
\end{aligned}$$

Therefore  $\check{f}$  is real, which implies  $f(t)$  is real.  $\square$

*Proof of Lemma 2.4.7.* Using [30, Proposition 5.1], we have

$$\acute{S}(e^{j\theta}) \acute{S}^{\sim}(e^{j\theta}) = \frac{1}{h} \sum_{k \in \mathbb{Z}} \psi(j\omega_k) \psi^{\sim}(j\omega_k)$$

Since  $\psi(t) \in L^2$ , therefore  $\psi(j\omega) \in L^2(j\mathbb{R})$ . Let for a given  $\theta \in [-\pi, \pi]$ ,  $\psi_{\theta}[k] = \psi(j\omega_k)$ . Since  $\psi(j\omega) \in L^2(j\mathbb{R})$ , we have  $\psi_{\theta} \in \ell^2$ . Therefore,

$$\sum_{k \in \mathbb{Z}} \psi(j\omega_k) \psi^{\sim}(j\omega_k) = \|\psi_{\theta}\|_2^2$$

converges for almost all  $\theta$ . Hence, the singular values of  $\acute{S}(e^{j\theta}) \acute{S}^{\sim}(e^{j\theta})$  are finite for almost all  $\theta$ . Therefore  $\|S\|_{L^{\infty}} = \text{ess sup}_{\theta \in [-\pi, \pi]} \|\acute{S}(e^{j\theta})\|_{\infty}$  is finite. The same is true for hold by duality.

Note that the transfer function of samplers and holds have uniformly bounded rank at almost all  $\theta \in [-\pi, \pi]$ . Therefore, if samplers and holds are in  $L^{\infty}$  then they are in  $L^2$  by Lemma 2.4.6.  $\square$

*Proof of Lemma 2.5.1.* Since  $\mathcal{G}$  is linear, we have

$$y(t) = \int_{-\infty}^{\infty} g(t, s) w(s) ds$$

Now  $\mathcal{G}$  is (classic) causal iff  $g(t, s) = 0 \forall s > t$ . This implies

$$y(t) = \int_{-\infty}^t g(t, s) w(s) ds$$

Lifting  $y(t)$ , we have

$$y(kh + \tau) = \int_{-\infty}^{kh + \tau} g(kh + \tau, s) w(s) ds$$

where  $\tau = [0, h)$ . The above is equivalent to

$$y(kh + \tau) = \int_{kh}^{kh+\tau} g(kh + \tau, s)w(s)ds + \sum_{i=-\infty}^{k-1} \int_{ih}^{(i+1)h} g(kh + \tau, s)w(s)ds$$

Using  $h$ -time shift invariance and change of variables ( $s = ih + \sigma$ ), we have

$$\begin{aligned} \check{y}[k](\tau) &= \int_0^\tau g(\tau, \sigma)\check{w}[k](\sigma)d\sigma \\ &+ \sum_{i=-\infty}^{k-1} \int_0^h g((k-i)h + \tau, \sigma)\check{w}[i](\sigma)d\sigma \end{aligned}$$

□

*Proof of Lemma 2.5.3.* Since the sampler is (classic) causal, we have that

$$\bar{y}[n] = \int_{-\infty}^{\infty} \psi(nh - t)y(t)dt = 0 \quad \forall n < 0$$

for all causal input signals  $y$ . This implies that (mind the interval of integration)

$$\bar{y}[n] = \int_0^{\infty} \psi(nh - t)y(t)dt = 0 \quad \forall n < 0 \quad (2.37)$$

This implies that  $\psi(nh - t) = 0 \forall n < 0, t \geq 0$  which means  $\psi(t - h) = 0, \forall t \leq 0$ . This further implies that (classic) causal sampler is given by

$$\begin{aligned} \bar{y}[n] &= \int_{-h}^{\infty} \psi(t)y(nh - t)dt \\ &= \int_{-\infty}^{nh+h} \psi(nh - t)y(t)dt \end{aligned}$$

□

*Proof of Lemma 2.5.6.* Since the hold is (classic) causal, we have that

$$u(t) = \sum_{n \in \mathbb{Z}} \phi(t - nh)\bar{u}[n] = 0 \quad \forall t < 0$$

for all causal input signals  $\bar{u}$ . This implies that

$$u(t) = \sum_{n \in \mathbb{N}} \phi(t - nh)\bar{u}[n] = 0 \quad \forall t < 0$$

This implies that  $\phi(t - nh) = 0 \forall t < 0, n \geq 0$  which means  $\phi(t) = 0 \forall t \leq 0$ . Therefore, a (classic) causal hold is given by

$$u(t) = \sum_{n \leq \lfloor \frac{t}{h} \rfloor, n \in \mathbb{Z}} \phi(t - nh)\bar{u}[n].$$

□





# Chapter 3

## Frequency truncated norms

### 3.1 Introduction

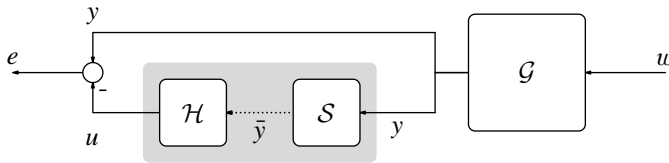


Figure 3.1: Sampled signal reconstruction setup

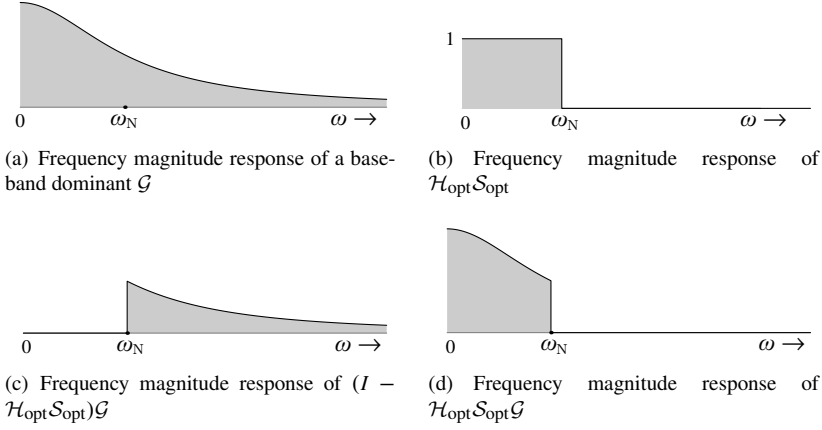
A fairly general sampled-data setup is shown in Figure 3.1. Here an analog signal  $y$  is sampled by a sampler  $\mathcal{S}$  (with sampling period  $h$ ) to produce a discrete signal  $\bar{y}$ . Then the hold  $\mathcal{H}$  converts the discrete signal  $\bar{y}$  back to the analog domain. This reconstructed signal  $u$  must resemble our original analog signal  $y$ . To check the quality of the reconstruction process,  $u$  is compared with  $y$ . To measure the quality of reconstruction process, it is common in sampled-data system theory to use the  $L^2$  or  $L^\infty$  norm [3, 2, 30] of the mapping from  $w$  to  $e = y - u$ . This mapping is given by

$$\mathcal{G}_e := (I - \mathcal{H}\mathcal{S})\mathcal{G}.$$

The joint design of an  $L^2$  or  $L^\infty$  optimal sampler-and-hold given a real linear continuous time invariant system (LCTI)  $\mathcal{G}$  is well known (see [31] and the references there in). As an example, consider an LCTI system  $\mathcal{G} \in L^\infty \cap L^2$  with frequency response  $G(j\omega)$  that is *baseband dominant* i.e

$$|G(j\omega)| \geq |G(j(\omega + 2n\omega_N))| \quad \forall n \in \mathbb{Z}, \omega \in [0, \omega_N] \quad (3.1)$$

where  $\omega_N := \frac{\pi}{h}$  is the Nyquist frequency corresponding to sampling period  $h$  (see Figure 3.2(a)). Now, it follows from [31, example 6.1] that both the  $L^2$  and  $L^\infty$



**Figure 3.2:** Frequency magnitude responses at various points in sample-data reconstruction setup if the frequency response is base-band dominant and the sampler and hold are  $L^2$  or  $L^\infty$  optimal.

optimal sampler is a cascade of the ideal low-pass filter which passes the frequency band  $[0, \omega_N]$ , followed by the ideal sampler i.e.

$$\bar{y} = \mathcal{S}_{\text{opt}}y : \quad \bar{y}[k] = \frac{1}{h} \int_{-\infty}^{\infty} \text{sinc}_h(kh - \tau)y(\tau) d\tau$$

and the optimal hold is the sinc interpolator

$$u = \mathcal{H}_{\text{opt}}\bar{y} : \quad u(t) = \sum_{k \in \mathbb{Z}} \bar{y}[k] \text{sinc}_h(t - kh).$$

Here  $\text{sinc}_h(t) = \sin(\omega_N t)/(\omega_N t)$ . Therefore, by the Shannon sampling theorem, the cascade  $\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}$  is the ideal low pass filter with cut-off frequency  $\omega_N$  and it is LCTI (see Figure 3.2(b)). This cascade  $\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}$  results in an LCTI  $\mathcal{G}_e = (I - \mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}})\mathcal{G}$  whose frequency response is the same as that of  $\mathcal{G}$  but without the band passed by  $\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}$  (see Figure 3.2(c)). Since  $\mathcal{G}_e$  is LCTI, its squared  $L^2$  norm can be calculated as:

$$\|\mathcal{G}_e\|_{L^2}^2 = \frac{1}{\pi} \int_{\omega_N}^{\infty} \|G(j\omega)\|^2 d\omega = \frac{1}{\pi} \text{tr} \int_{\omega_N}^{\infty} G^\sim(j\omega)G(j\omega) d\omega$$

where  $G(j\omega)$  is the frequency response of the LCTI system  $\mathcal{G}$ . Similarly, since  $\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}\mathcal{G}$  is an LCTI system with frequency response shown in Figure 3.2(d), its squared  $L^2$  system norm can be obtained as

$$\|\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}\mathcal{G}\|_{L^2}^2 = \frac{1}{\pi} \int_0^{\omega_N} \|G(j\omega)\|^2 d\omega = \frac{1}{\pi} \text{tr} \int_0^{\omega_N} G^\sim(j\omega)G(j\omega) d\omega.$$

Therefore, the calculation of a truncated finite and a semi-infinite integral is required for obtaining the system norms  $\|\mathcal{G}_e\|_{L^2}$  and  $\|\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}\mathcal{G}\|_{L^2}^2$  (if finite). Generalizing the situation, the basic integrals that need to be calculated can be written as

$$\|\mathcal{G}\|_{\omega_l}^2 := \frac{1}{\pi} \text{tr} \int_{\omega_l}^{\infty} G^{\sim}(j\omega)G(j\omega) d\omega \quad (3.2)$$

and

$$\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2 := \frac{1}{\pi} \text{tr} \int_{\omega_l}^{\omega_u} G^{\sim}(j\omega)G(j\omega) d\omega \quad (3.3)$$

where  $\mathcal{G}$  is LCTI and  $\omega_u > \omega_l \geq 0$ . For a given  $\omega_l$  and  $\omega_u$ ,  $\|\mathcal{G}\|_{\omega_l}$  and  $\|\mathcal{G}\|_{[\omega_l, \omega_u]}$  are known as *frequency truncated norms*. In the same way, if  $G(j\omega)$  is not baseband-dominant then the optimal  $\mathcal{HS}$  cancels the frequency band(s) in which the frequency response of  $\mathcal{G}$  is dominant [31]. In this case, the optimal sampler comprises of the ideal sampler and an ideal passband filter whose frequency pattern might be rather complicated. But in any case the optimal  $\mathcal{HS}$  cancels some finite frequency bands from  $|G(j\omega)|$ . Again to calculate the  $\|\mathcal{G}_e\|_{L^2}$  (or  $\|\mathcal{H}_{\text{opt}}\mathcal{S}_{\text{opt}}\mathcal{G}\|_{L^2}$ ) in this case, the fundamental requirement is to calculate integrals of type (3.2) and (3.3). Another interesting case arises when  $G(j\omega)$  has an imaginary axis pole  $j\omega$  for some  $\omega \geq 0$ . In this case,  $\mathcal{HS}$  cancels a frequency band of  $|G(j\omega)|$  that contains this imaginary axis pole in order to have a finite  $\|\mathcal{G}_e\|_{L^2}$  (see [31] for details). Again the fundamental fact is that the calculation of the  $\|\mathcal{G}_e\|_{L^2}$  (if finite) in this situation requires the calculation of integrals of type (3.2) and (3.3).

In summary, the calculation of integrals (3.2) and (3.3) are required in the sampled-data system theory. These integrals are also studied in the context of model reduction by [17] for a proper rational  $G(s)$  with real coefficients. However, there are some subtleties, in particular related to finiteness of the semi-infinite integral (3.2) and to the problem posed by imaginary poles of  $G(s)$ . Similar to [17], only proper rational systems are considered in this chapter. An advantage of using rational  $G(s)$  is that it provides a good approximation of a wide range of LCTI transfer functions and that it is computationally efficient due to its finite dimension. An advantage of a proper rational  $G(s)$  with real coefficients is that it can be realized in state space as  $G(s) := D + C(sI - A)^{-1}B$  where  $A$ ,  $B$ ,  $C$  and  $D$  are real matrices. This realization will be used throughout this chapter. Clearly the semi-infinite integral (3.2) is infinite if  $D \neq 0$ . Therefore,  $D$  is assumed zero whenever the semi-infinite integral is discussed. For the finite integral there is no such condition. All the subtleties related to imaginary axis poles of a given proper rational system  $G(s)$ , along with expressions for the integrals (3.2) and (3.3) are the main content of this chapter. This chapter is based on the papers [33, 34].

### 3.2 A generic version

If a proper rational system  $G(s)$  is realized as  $D + C(sI - A)^{-1}B$  then

$$G^\sim(s)G(s) = \hat{D} + \hat{C}(sI - A)^{-1}\hat{B}$$

where (see [70, §3.4])

$$\left[ \begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array} \right] := \left[ \begin{array}{cc|c} A & 0 & B \\ -C^T C & -A^T & -C^T D \\ \hline D^T C & B^T & D^T D \end{array} \right]. \quad (3.4)$$

Now, we have

$$\text{tr} \int G^\sim(j\omega)G(j\omega) d\omega = \text{tr} \hat{D}\omega + \text{tr} \int \hat{C}(j\omega I - A)^{-1}\hat{B} d\omega. \quad (3.5)$$

To evaluate the integral in the right hand side of the above equation, we consider a generic problem of integrating a transfer function  $K(s) := \tilde{C}(sI - \tilde{A})^{-1}\tilde{B}$  over a part of imaginary axis i.e.

$$\int_{\omega_l}^{\infty} K(j\omega)d\omega = \int_{\omega_l}^{\infty} \tilde{C}(j\omega I - \tilde{A})^{-1}\tilde{B} d\omega \quad (3.6)$$

where  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$  are possibly complex matrices and  $\omega_l \geq 0$ .

To understand the subtleties in the integration of  $K(j\omega)$ , let us consider an example.

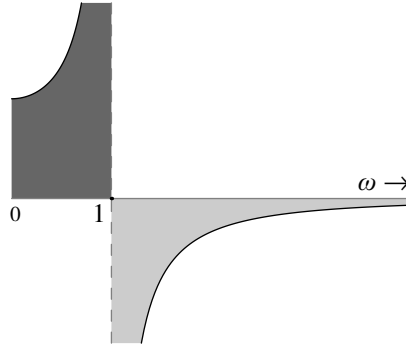
**Example 3.2.1.** Consider  $K(s) = 1/(s^2 + 1)$ . It has poles at  $\pm j$  on the imaginary axis. Consider the following semi-infinite integral with positive real  $\omega_l$  :

$$\begin{aligned} \int_{\omega_l}^{\infty} K(j\omega)d\omega &= \int_{\omega_l}^{\infty} 1/(-\omega^2 + 1) d\omega \\ &= \begin{cases} \frac{1}{2} \log \left( \frac{\omega_l - 1}{\omega_l + 1} \right) & \omega_l > 1 \\ \text{not defined} & \omega_l < 1 \end{cases}. \end{aligned}$$

Note that the above integral exists only for  $\omega_l > 1$ . Now consider the following finite integral with real  $\omega_l$  and  $\omega_u$  such that  $\omega_u \geq \omega_l$ :

$$\begin{aligned} \int_{\omega_l}^{\omega_u} K(j\omega)d\omega &= \int_{\omega_l}^{\omega_u} 1/(-\omega^2 + 1) d\omega \\ &= \begin{cases} \frac{1}{2} \log \left( \frac{(\omega_u + 1)(\omega_l - 1)}{(\omega_u - 1)(\omega_l + 1)} \right) & 1 \notin [\omega_l, \omega_u] \\ \text{not defined} & 1 \in [\omega_l, \omega_u] \end{cases} \end{aligned}$$

Here, again integration over the imaginary axis pole 1 is not defined. But otherwise it is defined and finite even if the limits  $\omega_u$  and  $\omega_l$  are less than 1. This result is also visible from Figure 3.3. If both the limits  $\omega_u$  and  $\omega_l$  lie in the same shaded region of the Figure 3.3 then  $\int_{\omega_l}^{\omega_u} K(j\omega) d\omega$  is defined.



**Figure 3.3:**  $K(j\omega)$  with respect to frequency  $\omega$ . The shaded region shows the integrable portion of  $K(j\omega)$ .

Example 3.2.1 shows the role of imaginary axis poles of  $K(s)$  in the existence of truncated integrals. Another important factor to note in the example is that  $\log$  is a multi-valued function over  $\mathbb{C}$ . To make it single-valued and analytic we take from now on, the *principal logarithm*. If a complex number  $z \in \mathbb{C}$  does not lie on the non-positive real axis  $\bar{\mathbb{R}}^-$ , then its principal logarithm  $\log(z)$  is defined as [48]

$$\log(z) := \log(|z|) + j \operatorname{Arg}(z) \quad (3.7)$$

where  $\operatorname{Arg}(z) \in (-\pi, \pi]$  is the principle argument function [48, chapter 3]. Similar to Example 3.2.1, it can be expected that integrating the transfer matrix  $K(s)$  results in a logarithm of some matrix. A short overview of matrix functions and the logarithm in particular is given in Section 3.2.1. After equipped with matrix logarithm, an anti-derivative of  $K(j\omega)$  can be obtained, hence the integral (3.6). This is discussed in Section 3.2.2.

### 3.2.1 Matrix functions and principal logarithm

The purpose of this section is to provide a quick overview of the theory of matrix functions and principal logarithm as described in Chapter 1 of [19] along with some theorem useful in evaluating integral (3.6).

The following is a standard result in Linear Algebra.

**Lemma 3.2.2** (Jordan canonical form). *Every matrix  $A \in \mathbb{C}^{n \times n}$  can be written in a Jordan canonical form*

$$A = Z \operatorname{diag}(J_1, J_2, \dots, J_p) Z^{-1}, \quad (3.8)$$

where the Jordan block  $J_k$  is a matrix of the form

$$J_k = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k \times m_k},$$

$Z$  is an invertible matrix, and  $\lambda_1, \lambda_2, \dots, \lambda_p$  are the eigenvalues of  $A$ .

*Proof.* See [54, chapter 5]. □

**Remark 3.2.3.** An eigenvalue of  $A$  can be associated with more than one Jordan block. Therefore in general the number of distinct eigenvalues can be less than or equal to  $p$ . In this section, we denote all distinct eigenvalues of  $A$  by  $\lambda_1, \dots, \lambda_s$  and the order of the largest Jordan block associated with  $\lambda_i$  with  $n_i$ .

The following terminology is needed in the definition of *Matrix functions*.

**Definition 3.2.4.** A function  $f : \mathbb{C} \rightarrow \mathbb{C}$  is said to be defined on the spectrum of the matrix  $A$  if

$$f^{(k)}(\lambda_i), \quad k = 0, \dots, n_i - 1, \quad i = 1, \dots, s$$

exist. Here  $f^{(k)}$  denotes the  $k$ -th derivative of  $f$ , and  $\lambda_i$  and  $n_i$  are as in Lemma 3.2.2 and Remark 3.2.3.

Now, the matrix function can be defined as follows:

**Definition 3.2.5.** Let (3.8) be a Jordan canonical form of the matrix  $A \in \mathbb{C}^{n \times n}$ . If a function  $f : \mathbb{C} \rightarrow \mathbb{C}$  is defined on the spectrum of  $A$ , then

$$f(A) := Z \operatorname{diag}(f(J_1), f(J_2), \dots, f(J_p))Z^{-1}$$

where

$$f(J_k) := \begin{bmatrix} f(\lambda_k) & f^{(1)}(\lambda_k) & \dots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & \ddots & \vdots \\ & & \ddots & f^{(1)}(\lambda_k) \\ & & & f(\lambda_k) \end{bmatrix}.$$

Clearly, for a complex matrix  $A$ ,  $\log(A)$  is defined only when the scalar function  $\log(z)$ ,  $z \in \mathbb{C}$  is analytic on the spectrum of  $A$  (see Definition 3.2.4). A standard convention is to use the principal logarithm and this is analytic on the set  $\mathbb{C} \setminus \bar{\mathbb{R}}^-$ . This implies that, to have a proper definition of  $\log(A)$  its eigenvalues must not lie on the non-positive real axis  $\bar{\mathbb{R}}^-$ . Similar to the scalar principal logarithm, the principal logarithm of a matrix  $A$  is thus defined as follows,

**Definition 3.2.6** (Principal logarithm). *If a complex matrix  $A$  does not have eigenvalues on the non-positive real axis  $\bar{\mathbb{R}}^-$  then the principal logarithm of  $A$  is defined as*

$$\log(A) := Z \operatorname{diag}(f(J_1), f(J_2), \dots, f(J_p)) Z^{-1} \quad (3.9)$$

where

$$\log(J_k) := \begin{bmatrix} \log(\lambda_k) & \frac{1}{\lambda_k} & \cdots & \frac{(-1)^{m_k-2}}{(m_k-1)\lambda_k^{m_k-1}} \\ & \log(\lambda_k) & \ddots & \vdots \\ & & \ddots & \frac{1}{\lambda_k} \\ & & & \log(\lambda_k) \end{bmatrix}$$

where the scalar principal logarithm is defined in (3.7).

Existence and uniqueness of the principal logarithm of a matrix  $A$  are proved in [19, Theorem 1.28 and 1.31].

The following properties are directly visible from the definition of matrix function and are proved in [19].

**Lemma 3.2.7.** *If a complex matrix  $A$  does not have eigenvalues on the non-positive real axis  $\bar{\mathbb{R}}^-$ , then its principal logarithm  $\log(A)$  exists and satisfies the following properties:*

1.  $\log(\lambda_k)$  is an eigenvalue of  $\log(A)$  if  $\lambda_k$  is an eigenvalue of  $A$ .
2.  $\log(\operatorname{diag}\{A_1, A_2, \dots, A_n\}) = \operatorname{diag}\{\log(A_1), \log(A_2), \dots, \log(A_n)\}$ .
3.  $\log(TAT^{-1}) = T \log(A) T^{-1}$ .
4.  $\log(A)^* = \log(A^*)$ .
5. If  $A$  is real, then  $\log(A)$  is real.

*Proof.* All the results follow from [19, Theorem 1.13] and [19, Theorem 1.18].  $\square$

### Matrix integral

In this section, we will use the definition of matrix function and logarithm to derive some important results which eventually helps in evaluating the integral (3.6). We start with the following small but useful lemma.

**Lemma 3.2.8.** *If  $\omega_u$  and  $\omega_l$  are real numbers then the complex function*

$$f_{\log}(z) := \log\left(\frac{\omega_u + jz}{\omega_l + jz}\right), \quad z \in \mathbb{C}$$

where  $\log$  means the scalar principal logarithm (3.7), is well-defined and analytic iff  $z \notin [j\omega_l, j\omega_u]$ . Here  $[j\omega_l, j\omega_u]$  denotes the set of convex combinations of  $j\omega_l$  and  $j\omega_u$ .



*Proof.* See Appendix 3.A (page 53). □

The above lemma is used to prove the following important theorem that has application in evaluating the integration of a rational transfer function.

**Theorem 3.2.9.** *Given a matrix  $A \in \mathbb{C}^{n \times n}$  with Jordan canonical form given by (3.8). Define  $\omega_{\max} := \max \omega_k$  where the maximum<sup>1</sup> is taken over all imaginary eigenvalues  $j\omega_k$  of  $A$ . Then,*

1. *on  $\omega \in (\omega_{\max}, \infty)$ , the function  $\log(\omega I + jA)$  is an analytic anti-derivative of  $(\omega I + jA)^{-1}$ .*
2. *if  $A$  has no imaginary eigenvalue  $j\omega$  with  $\omega \in [\omega_l, \omega_u]$  where  $\omega_l$  and  $\omega_u$  are real numbers, then*

$$\int_{\omega_l}^{\omega_u} (\omega I + jA)^{-1} d\omega = \log \left( (\omega_u I + jA)(\omega_l I + jA)^{-1} \right)$$

where  $\log$  is the principal logarithm (see Definition 3.2.6).

*Proof.* See Appendix 3.A (page 53). □

It is interesting to note that if  $\omega < \omega_{\max}$ , then anti-derivative  $\int (\omega I + jA)^{-1} d\omega$  cannot be written as  $\log(\omega I + jA)$  simply because it has an eigenvalue on the non-positive real axis  $\bar{\mathbb{R}}^-$ . But a finite integral exists if limits  $\omega_l$  and  $\omega_u$  satisfy the condition given in part (2) of Theorem 3.2.9.

Following are a few extensions of Theorem 3.2.9:

**Corollary 3.2.10.** *Given a matrix  $A \in \mathbb{C}^{n \times n}$ , then*

1.  $\lim_{\omega \rightarrow \infty} \log(\omega I + jA) - \log(\omega)I = 0$ .
2.  $\lim_{\omega \rightarrow \infty} \text{imag}(\log(\omega I + jA)) = 0$ .

where  $\log$  is the principal logarithm (see Definition 3.2.6).

*Proof.* See Appendix 3.A (page 55). □

The following result is also useful:

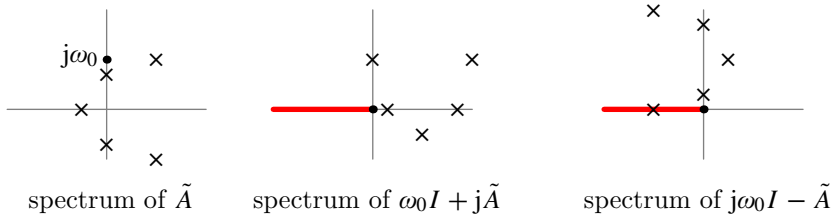
**Proposition 3.2.11.**

1. *If  $A$  is Hurwitz then  $\text{imag}(\log(jA)) = \text{imag}(\log(-A)) - \frac{\pi}{2}I$*
2. *If  $A$  is Hurwitz and real, then  $\text{imag}(\log(jA)) = -\frac{\pi}{2}I$*

*Proof.* See Appendix 3.A (page 55). □

---

<sup>1</sup> $\omega_{\max} = -\infty$  if  $A$  has no eigenvalues on imaginary axis.



**Figure 3.4:** Given a spectrum of  $\tilde{A}$  and given an  $\omega_0 > \omega_{\max}$  (left), the other two plots show the spectrum of  $\omega_0 I + j\tilde{A}$  and  $j\omega_0 I - \tilde{A}$ .

### 3.2.2 Integration of a transfer function

We return to the problem of integration of a strictly proper rational transfer function  $K(s) := \tilde{C}(sI - \tilde{A})^{-1}\tilde{B}$ . Equipped with the matrix logarithm, an anti-derivative of  $K(j\omega)$  can be obtained as:

**Lemma 3.2.12.** Define  $\omega_{\max} := \max \omega_k$  where the maximum is taken over all imaginary eigenvalues  $j\omega_k$  of  $\tilde{A}$ . An anti-derivative of  $K(j\omega) := \tilde{C}(j\omega I - \tilde{A})^{-1}\tilde{B}$  with  $\tilde{A}, \tilde{B}, \tilde{C}$  complex matrices, on  $(\omega_{\max}, \infty)$  is

$$\int K(j\omega)d\omega = -j\tilde{C} \log(\omega I + j\tilde{A})\tilde{B}. \quad (3.10)$$

*Proof.* Follows from Theorem 3.2.9(1). □

Analogous to the scalar case  $\int 1/(j(\omega - \omega_N)) d\omega = -j \log(\omega - \omega_N)$  for all  $\omega > \omega_N$ , it seems intuitive that (3.10) is an anti-derivative of

$$K(j\omega) = -j\tilde{C}(\omega I + j\tilde{A})^{-1}\tilde{B}. \quad (3.11)$$

However, this intuition may lead us to a wrong conclusion that the anti derivative of  $K(j\omega) = \tilde{C}(j\omega I - \tilde{A})^{-1}\tilde{B}$  on  $(\omega_{\max}, \infty)$  is

$$-j\tilde{C} \log(j\omega I - \tilde{A})\tilde{B} \quad (3.12)$$

as we see next. It is generally wrong because as  $\omega$  varies in  $(\omega_{\max}, \infty)$  some eigenvalues of  $j\omega I - \tilde{A}$  may cross the branch cut (the non positive real axis) of the principal logarithm, and this makes the candidate anti-derivative (3.12) discontinuous (and wrong), see Figure 3.4(right). Extracting  $j$  from the realization of  $K(j\omega)$  as done in (3.10) avoids this problem because now the matrix whose logarithm we take,  $\omega I + j\tilde{A}$ , by construction has no eigenvalues on the branch cut when  $\omega \in (\omega_{\max}, \infty)$ , see Figure 3.4(middle). All this is happening because the non-positive real axis  $\mathbb{R}^-$  is used as a branch cut for the principal logarithm, which is standard across literature. If the non-positive imaginary axis would have been chosen as branch cut for logarithm, then extraction of  $j$  from the realization of  $K(j\omega)$  would not have been needed at all.

Now it is a matter of putting the limits in the anti-derivative of  $K(j\omega)$  to evaluate the semi-infinite integral (3.6):

**Theorem 3.2.13.** *Suppose that  $K(s) = \tilde{C}(sI - \tilde{A})^{-1}\tilde{B}$  with  $\tilde{A}, \tilde{B}, \tilde{C}$  complex matrices. If  $\tilde{C}\tilde{B} = 0$ , then*

$$\int_{\omega_l}^{\infty} K(j\omega) d\omega = j\tilde{C} \log(\omega_l I + j\tilde{A})\tilde{B} \quad (3.13)$$

*provided that  $\omega_l > \omega_{\max} := \max \omega_k$  where the maximum is taken over all imaginary eigenvalues  $j\omega_k$  of  $\tilde{A}$ . If  $\tilde{C}\tilde{B} \neq 0$ , then the semi-infinite integral in the left hand side of (3.13) is not defined.*

*Proof.* Using Lemma 3.2.12, Corollary 3.2.10(1) and  $\tilde{C}\tilde{B} = 0$ , we obtain

$$\begin{aligned} \int_{\omega_l}^{\infty} K(j\omega) d\omega &= -j \int_{\omega_l}^{\infty} \tilde{C}(\omega I + j\tilde{A})^{-1}\tilde{B} d\omega \\ &= j\tilde{C} \log(\omega_l I + j\tilde{A})\tilde{B} - j \lim_{\omega \rightarrow \infty} \log(\omega)\tilde{C}\tilde{B} \\ &= j\tilde{C} \log(\omega_l I + j\tilde{A})\tilde{B}. \end{aligned}$$

Clearly if  $\tilde{C}\tilde{B} \neq 0$  then we do not have finite norm. □

The condition  $\tilde{C}\tilde{B} = 0$  is required to control the growth of the logarithm function as frequency  $\omega$  increases. The condition  $\tilde{C}\tilde{B} = 0$  is equivalent to  $K(s)$  having relative degree 2 or more. This can be seen as follows:

$$\begin{aligned} K(s) &= s^{-1}\tilde{C}(I - \tilde{A}/s)^{-1}\tilde{B} \\ &= s^{-1}\tilde{C}(I + \tilde{A}/s + (\tilde{A}/s)^2 + \dots)\tilde{B}, & \text{for } |s| > \|\tilde{A}\| \\ &= \tilde{C}\tilde{B}s^{-1} + \tilde{C}\tilde{A}\tilde{B}s^{-2} + \tilde{C}\tilde{A}^2\tilde{B}s^{-3} + \dots \end{aligned}$$

From this it is clear that the relative degree of  $K(s)$  is 2 iff  $\tilde{C}\tilde{B} = 0$ .

Sometimes, we just need the trace of the integral in (3.6). In this case, the condition  $\tilde{C}\tilde{B} = 0$  will change to condition  $\text{tr } \tilde{C}\tilde{B} = 0$  as shown in the following corollary.

**Corollary 3.2.14.** *Suppose that  $K(s) = \tilde{C}(sI - \tilde{A})^{-1}\tilde{B}$  with  $\tilde{A}, \tilde{B}, \tilde{C}$  complex matrices. If  $\text{tr } \tilde{C}\tilde{B} = 0$ , then*

$$\int_{\omega_l}^{\infty} \text{tr } K(j\omega) d\omega = j \text{tr } \tilde{C} \log(\omega_l I + j\tilde{A})\tilde{B} \quad (3.14)$$

*provided that  $\omega_l > \omega_{\max} := \max \omega_k$  where the maximum is taken over all imaginary eigenvalues  $j\omega_k$  of  $\tilde{A}$ . If  $\text{tr } \tilde{C}\tilde{B} \neq 0$ , then the semi-infinite integral in the left hand side of (3.14) is not defined.*

*Proof.* Similar to the proof of Theorem 3.2.13. □

### 3.2.3 Finite integral for proper $K(s)$

If a proper rational transfer function  $K(s) = \tilde{C}(sI - \tilde{A})^{-1}\tilde{B}$  has relative degree 1, then the semi-infinite integral (3.13) does not exist because  $\lim_{\omega \rightarrow \infty} K(j\omega) = \tilde{C}\tilde{B} \neq 0$  is nonzero. In Theorem 3.2.16, the condition  $\tilde{C}\tilde{B} = 0$  is required only to make the integral finite at  $\omega = \infty$ , but if the upper limit is finite then there is no need of this condition. Similarly, if  $K(s)$  has a state space realization  $\tilde{D} + \tilde{C}(sI - \tilde{A})^{-1}\tilde{B}$  with  $\tilde{D} \neq 0$ , then also the semi-infinite integral (3.13) does not exist, because  $\lim_{\omega \rightarrow \infty} K(j\omega) = \tilde{D}$  nonzero. However, a finite integral may still exist in the above cases:

**Proposition 3.2.15.** *Let  $K(s) = \tilde{C}(sI - \tilde{A})^{-1}\tilde{B} + \tilde{D}$  be a realization with  $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}$  real matrices. Then*

$$\int_{\omega_l}^{\omega_u} K(j\omega) d\omega = -j\tilde{C}[\log(\omega_u I + j\tilde{A}) - \log(\omega_l I + j\tilde{A})]\tilde{B} + \tilde{D}(\omega_u - \omega_l) \quad (3.15)$$

provided that  $\omega_l$  and  $\omega_u$  are finite, and greater than  $\omega_{\max} := \max \omega_k$  where the maximum is taken over all imaginary eigenvalues  $j\omega_k$  of  $\tilde{A}$ .

*Proof.* Follows from Lemma 3.2.12. □

Proposition 3.2.15 requires that both the limits should be greater than  $\omega_{\max}$ . However, in the scalar case (see Example 3.2.1 and Figure 3.3) and in Theorem 3.2.9(2), the only requirement is that the imaginary axis poles should not lie in between the limits of integration. This suggests that the finite integral in the Proposition 3.2.15 can be further generalized. Indeed.

**Theorem 3.2.16.** *Given finite  $\omega_u, \omega_l \in \mathbb{R}$ . Let  $K(s) = \tilde{C}(sI - \tilde{A})^{-1}\tilde{B} + \tilde{D}$  be a realization with real  $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}$  matrices and suppose that  $\tilde{A}$  has no imaginary eigenvalue  $j\omega$  with  $\omega \in [\omega_l, \omega_u]$ . Then,*

$$\int_{\omega_l}^{\omega_u} K(j\omega) d\omega = -j\tilde{C} \log(\Omega)\tilde{B} + \tilde{D}(\omega_u - \omega_l).$$

where

$$\tilde{\Omega} := (\omega_u I + j\tilde{A})(\omega_l I + j\tilde{A})^{-1} = (j\omega_u I - \tilde{A})(j\omega_l I - \tilde{A})^{-1}. \quad (3.16)$$

*Proof.* Follows from Theorem 3.2.9(2). □

Notice that the extraction of  $j$  from  $(j\omega I - \tilde{A})$  — which was needed earlier to avoid eigenvalues on the branch cut  $\bar{\mathbb{R}}^-$  — cancels in the formula for  $\tilde{\Omega}$ . In fact, any non-zero complex number can be extracted. In any case  $\tilde{\Omega}$  has no eigenvalues on the non-positive real axis  $\bar{\mathbb{R}}^-$  if  $\tilde{A}$  has no imaginary eigenvalue  $j\omega$  with  $\omega \in [\omega_l, \omega_u]$  (see Lemma 3.2.8 also).

The remaining subsections contain applications and special cases of Theorem 3.2.13, Corollary 3.2.14 and Theorem 3.2.16.

### 3.3 Computation of frequency truncated $L^2$ -norm

We return to the truncated integrals (3.2) and (3.3). If  $G(s) := C(sI - A)^{-1}B$ , then  $G \sim G$  can be realized as (3.4) with  $D = 0$ . In this case  $\text{tr} \hat{C} \hat{B} = 0$ , so the semi-infinite integral (3.2) can be obtained via Corollary 3.2.14.

**Proposition 3.3.1.** *For a real system  $\mathcal{G}$  with given transfer matrix  $G(s) := C(sI - A)^{-1}B$ , the truncated norm (3.2) satisfies*

$$\|\mathcal{G}\|_{\omega_l}^2 = \frac{1}{\pi} \text{tr} \left( j \hat{C} \log(\omega_l I + j \hat{A}) \hat{B} \right) \quad (3.17)$$

provided that  $\omega_l > \omega_{\max} := \max |\omega_k|$  where the maximum is taken over all imaginary eigenvalues  $j\omega_k$  of  $A$ , and  $\hat{A}$ ,  $\hat{B}$ , and  $\hat{C}$  are as in (3.4) with  $D = 0$ .

*Proof.* Since  $\hat{A}$  is real,  $\max |\omega_k|$  is same as  $\max \omega_k$ . Also, as  $\text{tr} \hat{C} \hat{B} = 0$ , the result follows from Corollary 3.2.14.  $\square$

If the  $A$  matrix of  $G(s)$  does not have a pole on the imaginary axis, then  $\omega_{\max}$  can be taken as  $-\infty$ . This means  $\omega_l$  is allowed to be anything, in particular zero. This case arises if  $G(s)$  is stable which is equivalent to say that  $A$  does not have eigenvalues on the imaginary axis (see Example 2.6.3). Considering only the stable and causal  $G(s)$  case (see Example 2.6.5), it will be shown in next section that this leads to a more computationally efficient form of (3.17). A discussion similar to the stable and causal systems also follows for stable and anti-causal systems.

#### 3.3.1 Stable and causal $G(s)$

It is a classic result that the squared  $L^2$ -norm

$$\|\mathcal{G}\|_{L^2}^2 := \frac{1}{\pi} \text{tr} \int_0^\infty G \sim(j\omega) G(j\omega) d\omega \quad (3.18)$$

of a stable and causal finite dimensional system  $\mathcal{G}$  with given transfer matrix  $G(s) = C(sI - A)^{-1}B$  can be computed via the solution of a linear equation. Specifically, if  $A$  is Hurwitz then

$$\|\mathcal{G}\|_{L^2}^2 = \text{tr}(B^T P B) \quad (3.19)$$

where  $P$  is the unique solution of the Lyapunov equation

$$A^T P + P A = -C^T C, \quad (3.20)$$

see e.g. [70, Lemma 2.1]. If  $A$  is a square matrix of dimension  $n$ , then  $\|\mathcal{G}\|_{L^2}^2$  obtained via (3.19) is given in terms of an  $n \times n$  matrix whereas  $\|\mathcal{G}\|_{L^2}^2$  obtained via (3.17) is given in terms of an  $2n \times 2n$  matrix. This suggests that there may exist a simplification of (3.17) that requires only a  $n \times n$  matrix, in case of a Hurwitz  $A$  matrix.

**Theorem 3.3.2.** *Let  $G(s) = C(sI - A)^{-1}B$  be a realization with  $A, B, C$  real matrices and  $A$  Hurwitz. Then*

$$\|\mathcal{G}\|_{\omega_l}^2 = -\frac{2}{\pi} \operatorname{imag} \operatorname{tr}(B^T P \log(\omega_l I + jA)B) \quad (3.21)$$

$$= \|\mathcal{G}\|_{L^2}^2 - \frac{2}{\pi} \operatorname{imag} \operatorname{tr}(B^T P \log(j\omega_l I - A)B) \quad (3.22)$$

where  $P$  is the unique solution of the Lyapunov equation (3.20).

*Proof.* See Appendix 3.A (page 55). □

In Theorem 3.3.2, Hurwitzness of  $A$  helps in two different ways. First, a solution of the Lyapunov equation (3.20) is guaranteed to exist, and second, the eigenvalues of the matrices  $\omega_l I + jA$  and  $j\omega_l I - A$  do not lie on the non-positive real axis  $\mathbb{R}^-$ . This guarantees the existence of  $\log(\omega_l I + jA)$  and  $\log(j\omega_l I - A)$  for all  $\omega_l \in \mathbb{R}$ .

For  $\omega_l = 0$  we recover (3.19). Indeed for  $\omega_l = 0$ , (3.21) reduces to

$$\begin{aligned} \|\mathcal{G}\|_{\omega_l=0}^2 &= \frac{-2}{\pi} \operatorname{tr} \left( B^T P [\operatorname{imag} \log(jA)] B \right) \\ &= \frac{-2}{\pi} \operatorname{tr} \left( B^T P \left[ -\frac{\pi}{2} I \right] B \right) = \operatorname{tr} (B^T P B). \end{aligned}$$

Here Proposition 3.2.11 is used, which states that  $\operatorname{imag}(\log(jA)) = -\frac{\pi}{2}I$  for every Hurwitz and real matrix  $A$ .

The following example shows an application of Theorem 3.3.2:

**Example 3.3.3.** *Given a stable and causal system  $\mathcal{G}$  with transfer function*

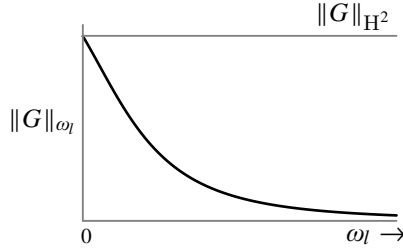
$$G(s) = \frac{1}{(s+1)^2(s+2)}.$$

The state-space realization of  $G(s)$  is  $C(sI - A)^{-1}B + D$  where

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] = \left[ \begin{array}{ccc|c} -4 & -5 & -2 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \end{array} \right].$$

Figure 3.5 (page 52) shows the truncated norm  $\|\mathcal{G}\|_{\omega_l}$  with respect to truncation frequency  $\omega_l$ , as calculated by (3.21). As expected, the truncated norm converges to the  $H_2$  norm of the system when  $\omega_l$  tends to zero and converges to 0 when  $\omega_l$  tends to infinity.

The next section explains the computation of the finite integral (3.3).



**Figure 3.5:** Truncated norm for Example 3.3.3 at different truncation frequencies.

### 3.3.2 Finite integral for $G(s)$

Similar to Section 3.2.3, assume that  $G(s)$  is proper with a state space realization  $G(s) := D + C(sI - A)^{-1}B$  with a non-zero  $D$ . In this case, the semi-infinite integral (3.2) does not exist because  $\lim_{\omega \rightarrow \infty} G^{\sim}(j\omega)G(j\omega) = D^*D$  is a nonzero quantity. Nevertheless, a finite integral may still exist.

**Theorem 3.3.4.** Take  $\omega_u, \omega_l \in \mathbb{R}$ . Let  $G(s) = C(sI - A)^{-1}B + D$  be a realization with real  $A, B, C, D$  matrices and suppose that  $A$  has no imaginary eigenvalue  $j\omega$  with  $\omega \in [\omega_l, \omega_u]$ . Then,

$$\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2 = \frac{1}{\pi} \left( -j \operatorname{tr} \hat{C} \log(\hat{\Omega}) \hat{B} + \operatorname{tr} \hat{D}(\omega_u - \omega_l) \right) \quad (3.23)$$

where  $\hat{\Omega} := (\omega_u I + j\hat{A})(\omega_l I + j\hat{A})^{-1}$ , and  $\hat{A}, \hat{B}, \hat{C}$  and  $\hat{D}$  are as defined in (3.4).

*Proof.* Follows from Theorem 3.2.16.  $\square$

#### Finite integral for stable and causal $G(s)$

Similar to Section 3.3.1, if  $A$  is Hurwitz then the following simplified form is possible.

**Theorem 3.3.5.** Let  $G(s) = D + C(sI - A)^{-1}B$  be a realization with  $A, B, C, D$  real matrices and  $A$  Hurwitz. Then

$$\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2 = \frac{2}{\pi} \operatorname{imag} \operatorname{tr} \left( (B^T P + D^T C) \log(\Omega) B \right) + \frac{1}{\pi} \operatorname{tr} D^T D (\omega_u - \omega_l) \quad (3.24)$$

where  $\Omega := (\omega_u I + jA)(\omega_l I + jA)^{-1}$  and  $P$  is the unique solution of the Lyapunov equation (3.20).

*Proof.* See Appendix 3.A (page 56).  $\square$

## 3.4 Conclusions

This chapter gives expressions in terms of principal logarithms for frequency truncated norms of stable and unstable systems. There exist computationally efficient algorithms, not relying on the Jordan form, to compute the principal logarithm. One such algorithm is implemented in MATLAB [10]. A computationally efficient form for frequency truncated norm is also obtained for stable systems in this chapter.

### 3.A Proofs

*Proof of Lemma 3.2.8.* Since the principal logarithm by definition is not defined and analytic on the non-positive real axis  $\bar{\mathbb{R}}^-$ ,  $f_{\log}(z)$  is not defined and analytic iff  $(\omega_u + jz)(\omega_l + jz)^{-1}$  lies on  $\bar{\mathbb{R}}^-$ . This means that there exists a real  $r \geq 0$  such that

$$(\omega_u + jz)(\omega_l + jz)^{-1} = -r.$$

This gives

$$z = j \frac{\omega_u + r\omega_l}{r + 1}. \quad (3.25)$$

This  $z$  is an imaginary number because  $r$ ,  $\omega_l$  and  $\omega_u$  are real. Since  $r \geq 0$ , equation (3.25) shows that  $z$  is a convex combination of  $j\omega_u$  and  $j\omega_l$ . But this contradicts the assumption that  $z \notin [j\omega_l, j\omega_u]$ .  $\square$

*Proof of Theorem 3.2.9.* Let  $\lambda_k$  be the eigenvalue of Jordan block  $J_k \in \mathbb{C}^{m_k \times m_k}$  (see (3.8)). For a given  $\omega \in \mathbb{R}$ , the function  $1/(\omega + jz)$  is analytic with respect to  $z$  as long as  $z \neq j\omega$ . Therefore, using Definition 3.2.4 and (3.8), if  $j\omega$  is not an eigenvalue of  $A$ , we have

$$(\omega I + jA)^{-1} = Z (\text{diag}(F_1(\omega), \dots, F_p(\omega))) Z^{-1}$$

where  $F_k(\omega) \in \mathbb{C}^{m_k \times m_k}$  is defined as

$$F_k(\omega) := \begin{bmatrix} (\omega + j\lambda_k)^{-1} & -j(\omega + j\lambda_k)^{-2} & \cdots & (-j)^{m_k-1}(\omega + j\lambda_k)^{-m_k} \\ & (\omega + j\lambda_k)^{-1} & \ddots & \vdots \\ & & \ddots & -j(\omega + j\lambda_k)^{-2} \\ & & & (\omega + j\lambda_k)^{-1} \end{bmatrix}. \quad (3.26)$$

Let  $j\omega_1, \dots, j\omega_q$  be all distinct imaginary eigenvalues of  $A$  with  $\omega_1 \leq \omega_2 \leq \dots \leq \omega_q$ . Define  $\omega_0 := -\infty$  and  $\omega_{q+1} := \infty$ , then for a given  $n \in \{0, \dots, q\}$

$$\int (\omega I + jA)^{-1} d\omega = Z (\text{diag}(L_1(\omega), \dots, L_p(\omega))) Z^{-1}, \quad (3.27)$$



is an anti-derivative of  $(\omega I + jA)^{-1}$  at all  $\omega \in (\omega_n, \omega_{n+1})$ , where  $L_k(\omega) \in \mathbb{C}^{m_k \times m_k}$  is defined as

$$L_k(\omega) := \begin{bmatrix} f_{\lambda_k}(\omega) & j(\omega + j\lambda_k)^{-1} & \cdots & -(-j)^{m_k-1} \frac{(\omega + j\lambda_k)^{-m_k+1}}{m_k-1} \\ & f_{\lambda_k}(\omega) & \ddots & \vdots \\ & & \ddots & j(\omega + j\lambda_k)^{-1} \\ & & & f_{\lambda_k}(\omega) \end{bmatrix} \quad (3.28)$$

and

$$f_{\lambda_k}(\omega) := \begin{cases} \log(\omega + j\lambda_k), & \text{if } \omega + j\lambda_k \notin \bar{\mathbb{R}}^- \\ \log(-\omega - j\lambda_k), & \text{if } \omega + j\lambda_k \in \mathbb{R}^- \end{cases} \quad (3.29)$$

Since the principal log is not defined and analytic on  $\bar{\mathbb{R}}^-$ ,  $f_{\lambda_k}(\omega)$  is defined in this way. Note that  $f_{\lambda_k}(\omega)$  in the above equation is defined only for  $\omega \in (\omega_n, \omega_{n+1})$ , therefore we have either  $f_{\lambda_k}(\omega) = \log(\omega + j\lambda_k)$  or  $f_{\lambda_k}(\omega) = \log(-\omega - j\lambda_k)$ .

*Part (1):* If real  $\omega > \omega_{\max}$ , then by Definition 3.2.6:

$$\log(\omega I + jA) = Z \left( \text{diag}(L_1(\omega), \dots, L_p(\omega)) \right) Z^{-1} \quad (3.30)$$

because  $\log(\omega + jz)$  is well-defined and analytic on the spectrum of  $A$  and

$$\frac{d^n}{dz^n} \log(\omega + jz) = -(-j)^n (\omega + jz)^{-n} (n-1)!$$

*part (2):* If no eigenvalue of  $A$  is in  $[j\omega_l, j\omega_u]$  then the anti-derivative in (3.27) is defined for all  $\omega \in [\omega_l, \omega_u]$ . Therefore

$$\int_{\omega_l}^{\omega_u} (\omega I + jA)^{-1} d\omega = Z \left( \text{diag}(L_1(\omega), \dots, L_p(\omega)) \Big|_{\omega_l}^{\omega_u} \right) Z^{-1},$$

where  $L_k(\omega) \Big|_{\omega_l}^{\omega_u} := L_k(\omega_u) - L_k(\omega_l)$  is given by,

$$\begin{bmatrix} \log \left( \frac{\omega_u + j\lambda_k}{\omega_l + j\lambda_k} \right) & \cdots & -(-j)^{m_k-1} \left( \frac{(\omega_u + j\lambda_k)^{-m_k+1}}{m_k-1} - \frac{(\omega_l + j\lambda_k)^{-m_k+1}}{m_k-1} \right) \\ & \ddots & \vdots \\ & & \log \left( \frac{\omega_u + j\lambda_k}{\omega_l + j\lambda_k} \right) \end{bmatrix}.$$

for all eigenvalues  $\lambda_k$  of  $A$ .

Note that  $\log((\omega_u + jz)(\omega_l + jz)^{-1})$  is a well-defined and analytic function for all  $z \notin [j\omega_l, j\omega_u]$  (see Lemma 3.2.8), in particular at all eigenvalues of  $A$  as  $A$  has no eigenvalues in  $[j\omega_l, j\omega_u]$ . Therefore by Definition 3.2.6

$$\log \left( (\omega_u I + jA)(\omega_l I + jA)^{-1} \right) = Z \left( \text{diag}(L_1(\omega), \dots, L_p(\omega)) \Big|_{\omega_l}^{\omega_u} \right) Z^{-1}.$$

as  $\frac{d^n}{dz^n} \log \left( \frac{\omega_u + jz}{\omega_l + jz} \right) = -(-j)^n ((\omega_u + jz)^{-n} - (\omega_l + jz)^{-n}) (n-1)!$   $\square$

*Proof of Corollary 3.2.10.* For  $\omega > \omega_{\max}$ , (3.30) can be used.

1. Since  $\lim_{\omega \rightarrow \infty} \log(\omega + j\lambda_k) - \log(\omega) = 0$  and  $\lim_{\omega \rightarrow \infty} (\omega + j\lambda_k)^{-j} = 0$  for an integer  $j > 0$ , we have that  $\lim_{\omega \rightarrow \infty} L_k(\omega) - \log(\omega)I_{m_k} = 0$  (see (3.28)).
2. Since  $\log(\omega) \in \mathbb{R}$ , we have that

$$\lim_{\omega \rightarrow \infty} \text{imag}(L_k(\omega) - \log(\omega)I_{m_k}) = 0 = \lim_{\omega \rightarrow \infty} \text{imag}(L_k(\omega))$$

□

*Proof of Proposition 3.2.11.* Let  $z \in \mathbb{C}$  and  $\text{real}(z) < 0$ , then using the branch  $\mathbb{C} \setminus \mathbb{R}^-$  for scalar logarithm, we have

$$\log(jz) = \log(-z) - j\frac{\pi}{2}.$$

Since  $A$  is a Hurwitz matrix, the functions  $\log(jz)$  and  $\log(-z)$  are defined on the spectrum of  $A$ . Hence,  $\log(-A)$  and  $\log(jA)$  exist by Definition 3.2.6. Therefore, using [19, Theorem 1.15a] (or Definition 3.2.6),

$$\log(jA) = \log(-A) - j\frac{\pi}{2}I$$

and therefore,

$$\begin{aligned} \text{imag}(\log(jA)) &= \text{imag}(\log(-A) - j\frac{\pi}{2}I) \\ &= \text{imag}(\log(-A)) - \frac{\pi}{2}I. \end{aligned}$$

If  $A \in \mathbb{R}^{n \times n}$ , then  $\log(-A)$  is real by Lemma 3.2.7(5). Hence,  $\text{imag}(\log(-A)) = 0$ . □

*Proof of Theorem 3.3.2.* Using (3.17), the truncated norm  $\|\mathcal{G}\|_{\omega_l}^2$  is given by

$$\|\mathcal{G}\|_{\omega_l}^2 = \frac{1}{\pi} \text{tr} \left( j\hat{C} \log(\omega_l I + j\hat{A})\hat{B} \right)$$

where  $\hat{C} = [0 \ B^T]$  and  $\hat{B} = [B \ 0]^T$  as  $D = 0$ .

Define  $T := \begin{bmatrix} I & 0 \\ -P & I \end{bmatrix}$ , then using (3.20) and Lemma 3.2.7(3, 4), we have

$$\log(\omega_l I + j\hat{A}) = T^{-1} \log(\omega_l I + jT\hat{A}T^{-1})T \quad (3.31)$$

$$= T^{-1} \text{diag}\{\log(\omega_l I + jA), \log(\omega_l I - jA^T)\}T. \quad (3.32)$$

Therefore, using above  $\|\mathcal{G}\|_{\omega_l}^2$  is given as

$$\|\mathcal{G}\|_{\omega_l}^2 = \frac{1}{\pi} \text{tr} \left( j\hat{C} \log(\omega_l I + j\hat{A})\hat{B} \right)$$

$$\begin{aligned}
&= \frac{1}{\pi} \operatorname{tr} \left( \mathbf{j} \hat{C} T^{-1} \operatorname{diag} \{ \log(\omega_l I + \mathbf{j}A), \log(\omega_l I - \mathbf{j}A^T) \} T \hat{B} \right) \\
&= \frac{1}{\pi} \mathbf{j} \operatorname{tr} \left( B^T P \log(\omega_l I + \mathbf{j}A) B - B^T \log(\omega_l I - \mathbf{j}A^T) P B \right) \\
&= -\frac{2}{\pi} \operatorname{tr} \operatorname{imag} (B^T P \log(\omega_l I + \mathbf{j}A) B)
\end{aligned}$$

where the last step follows from Lemma 3.2.7(4) which says  $\log(A)^* = \log(A^*)$ . Since  $-(\mathbf{j}\omega_l I - A)$  is Hurwitz for every  $\omega \in \mathbb{R}$ , therefore using Theorem 3.2.11(1),

$$\begin{aligned}
\operatorname{imag} \log(\omega_l I + \mathbf{j}A) &= \operatorname{imag} \log(-\mathbf{j}(\mathbf{j}\omega_l I - A)) \\
&= \operatorname{imag} \log(\mathbf{j}\omega_l I - A) - \frac{\pi}{2} I.
\end{aligned}$$

Hence,  $\|\mathcal{G}\|_{\omega_l}^2$  equals

$$\|\mathcal{G}\|_{\omega_l}^2 = \operatorname{tr} B^T P B - \frac{2}{\pi} \operatorname{tr} \operatorname{imag} (B^T P \log(\mathbf{j}\omega_l I - A) B).$$

□

*Proof of Theorem 3.3.5.* Using (3.23), the truncated norm  $\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2$  is given by

$$\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2 = \frac{1}{\pi} \operatorname{tr} \left( -\mathbf{j} \hat{C} \log(\hat{\Omega}) \hat{B} + \operatorname{tr} \hat{D}(\omega_u - \omega_l) \right)$$

where  $\hat{\Omega} := (\omega_u I + \mathbf{j}\hat{A})(\omega_l I + \mathbf{j}\hat{A})^{-1}$ . Define  $T := \begin{bmatrix} I & 0 \\ -P & I \end{bmatrix}$ , then using (3.20) and Lemma 3.2.7(3,2), we have

$$\begin{aligned}
\log(\hat{\Omega}) &= T^{-1} \log \left( (\omega_u I + \mathbf{j}T \hat{A} T^{-1})(\omega_l I + \mathbf{j}T \hat{A} T^{-1})^{-1} \right) T \\
&= T^{-1} \operatorname{diag} \{ \log(\Omega), \log(\Omega^*) \} T.
\end{aligned}$$

where  $\Omega := (\omega_u I + \mathbf{j}A)(\omega_l I + \mathbf{j}A)^{-1}$ . Therefore, using above  $\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2$  is given as

$$\begin{aligned}
\|\mathcal{G}\|_{[\omega_l, \omega_u]}^2 &= -\frac{1}{\pi} \operatorname{tr} \left( \mathbf{j} \hat{C} \log(\hat{\Omega}) \hat{B} \right) + \frac{1}{\pi} \operatorname{tr} \hat{D}(\omega_u - \omega_l) \\
&= -\frac{1}{\pi} \operatorname{tr} \left( \mathbf{j} \hat{C} T^{-1} \operatorname{diag} \{ \log(\Omega), \log(\Omega^*) \} T \hat{B} \right) + \frac{1}{\pi} \operatorname{tr} \hat{D}(\omega_u - \omega_l) \\
&= -\frac{1}{\pi} \mathbf{j} \operatorname{tr} \left( (B^T P + D^T C) \log(\Omega) B - B^T \log(\Omega^*) (P B + C^T D) \right) \\
&\quad + \frac{1}{\pi} \operatorname{tr} \hat{D}(\omega_u - \omega_l) \\
&= \frac{2}{\pi} \operatorname{tr} \operatorname{imag} \left( (B^T P + D^T C) \log(\Omega) B \right) + \frac{1}{\pi} \operatorname{tr} \hat{D}(\omega_u - \omega_l)
\end{aligned}$$

where the last step follows from Lemma 3.2.7(4) which says  $\log(\Omega)^* = \log(\Omega^*)$ . □

# Chapter 4

## Non-causal downsampling

### 4.1 Introduction

In most of the signal processing applications, information is taken from analog sources and utilized in analog format. Due to the high quality of digital transmission, information is often converted in a discrete form at the front end from an analog source and converted back to analog at the rear end. Sometimes in applications like image, audio, video etc, it is necessary to change the sampling rate of the discrete signal. There are several reasons for this sampling rate conversion such as different clocks in two digital hardware, etc. The process of increasing the sampling rate is known as *upsampling* or *interpolation* whereas decreasing the sampling rate is known as *downsampling* or *decimation* [9,46]. The reduction factor of the sampling rate can be an integer or a rational number greater than one. Instead of imposing any structure, we just assume that the downsampling process is linear and  $h$ -time shift invariant. One approach to the problem of reconstruction of an analog signal from a discrete signal uses the theory of sampled-data system (see Chapter 2). The two distinctive features of this approach are the use of signal generation models and the use of system norms for the measurement of reconstruction performance. The sampled-data system theory can also be applied to downsampling problem, see [20]. The sampled-data set-up for the downsampling problem is shown in Figure 4.1. Here the continuous time signal  $y$  is sampled with

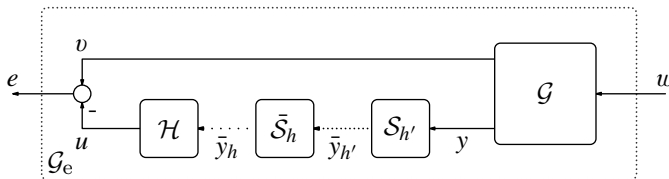


Figure 4.1: Downsampling in sampled-data setting

sampling period  $h'$  by a given sampler  $\mathcal{S}_{h'}$ , then the sampling rate of the resulting discrete signal  $\bar{y}_{h'}$  is changed by a downsampler  $\bar{\mathcal{S}}_h$  to a slower rate of  $\frac{1}{h} := \frac{1}{Mh'}$ , where  $M$  is a positive integer known as the *downsampling factor*. The output  $\bar{y}_h$  of the downsampler is converted back to the analog domain by a Hold  $\mathcal{H}$ . Any discrete filter in between  $\bar{\mathcal{S}}_h$  and  $\mathcal{H}$  can be combined in either  $\bar{\mathcal{S}}_h$  or  $\mathcal{H}$  without loss of generality. The continuous time signals  $y$  and  $v$  are modeled as the output of a known linear continuous time invariant (LCTI) system  $\mathcal{G}$  driven by a known set of signals  $w$  (see Figure 4.1).  $\mathcal{G}$  can be thought as a model of frequency characteristics of the signals  $y$  and  $v$ . The choice of  $\mathcal{G}$  depends upon the application at hand. The main aim in the downsampling problem is to construct  $u$  as close as possible to  $v$  with design parameter  $\bar{\mathcal{S}}_h$  and  $\mathcal{H}$ , given  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ . To take care of factors like noise, the reconstructed signal  $u$  is compared with  $v$  instead of  $y$ . Also normalization of the signal  $w$  allows us to write the reconstruction performance in terms of the norm of the system  $\mathcal{G}_e$  that maps  $w$  to  $e$  in Figure 4.1. We use the  $L^2$  and  $L^\infty$  system norms (defined in Chapter 2) to measure the reconstruction performance in this chapter. The advantage of using sampled-data system theory for downsampling is that it utilizes the inter-sample information available.

A general approach for downsampling in the signal processing literature is to somehow band-limit the input signal  $y$  to the Nyquist frequency  $\omega_N := \pi/h$  (or any multiple of the Nyquist frequency if one allows filter banks) to prevent aliasing errors. This approach leads to error-free reconstruction. However in practice, perfect band-limitedness is not possible, therefore we can think of using some error criteria (such as  $L^2$  or  $L^\infty$ ). The downsampling problem from the sampled-data viewpoint is studied first by Ishii *et. al* [20], where they use the  $H^\infty$  error criteria to design causal optimal downsamplers  $\bar{\mathcal{S}}_h$  (given  $\mathcal{H}$ ,  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ ). Nagahara and Yamamoto [43] designed a computationally efficient  $H^\infty$  optimal downsamplers (given  $\mathcal{H}$ ,  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ ). The problem of designing non-causal downsamplers (given  $\mathcal{H}$ ,  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ ) and the problem of designing non-causal holds (given  $\bar{\mathcal{S}}_h$ ,  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ ) is essentially solved in [31, Proposition 2.1 and 3.1]. The problem of designing the optimal non-causal downsamplers and holds simultaneously (given  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ ) is first studied in [31], but for a very limited class of the signal generators  $\mathcal{G}$ . [31] uses  $L^2$  and  $L^\infty$  system norm for measuring the reconstruction performance.

In this chapter, we show how to extend the downsampling result of [31] for all LCTI signal generators  $\mathcal{G}$ . Here, the problem of designing joint non-causal downsampler-and-hold, given  $\mathcal{S}_{h'}$  and  $\mathcal{G}$ , is considered. This leads to a generic treatment of the downsampling problem. In this chapter, we will also see how to quantify the signal reconstruction error and its variation in the presence of noise. This chapter is based on the papers [50, 52].

The outline of this chapter is as follows. We describe mathematical representations of the sampled-data setup for the downsampling problem in Section 4.3. Then we define discrete lifting, discrete lifted-transforms, and lifted transfer functions in Sections 4.4 and 4.5. An important result known as *Rank theorem* is discussed in the Section 4.7. Mathematical formulations and simplifications of the downsampling problem is done in Section 4.8. Section 4.9 and Section 4.10 contain the main results of this chapter i.e. solution of the downsampling problem

with different error norms along with some examples and an expression for the error norm. The effect of noise on the downsampling problem is discussed in the Section 4.11.

## 4.2 Notations

Throughout this chapter  $h$  denotes the sampling period,  $\omega_N := \pi/h$  is the associated Nyquist frequency. Also  $h' = h/M$ , where  $M$  is a positive integer known as the *downsampling factor*. Linear discrete time invariant (LDTI) system means a linear  $h$ -time shift invariant system.

A discrete signal  $\bar{y}$  with interval  $h'$  means that the samples in the discrete signal are separated by fixed  $h'$  time. Sometimes we write  $\ell^2(\mathbb{Z}, \mathbb{C}^n)$  as  $\ell_{h'}^2$  in order to emphasize the fact that the discrete signals in  $\ell_{h'}^2$  are generated from sampling a continuous time signal with sampling period  $h'$ . In all other cases, discrete signals are assumed to be generated from sampling a continuous time signal with sampling period  $h$  unless otherwise stated.

We use the same notation of Chapter 2 given in Section 2.1 for signals and systems. Similar to Section 2.1, the meaning of some of the systems, domains and transforms will be cleared in the later sections. In addition to Section 2.1, a downsampler  $\bar{S}_h$  in the lifted time domain is represented by calligraphic letter with a acute on top e.g.  $\bar{S}$ . A downsampler in the lifted frequency domain is represented by capital letter with a acute on top e.g.  $\bar{S}_h$ . If  $\bar{y}$  is a discrete signal with interval  $h'$  then in lifted domain it is represented as  $\bar{y}$ . In an apologetic way, the lifted  $z$ -transform of the signal  $\bar{y}$  is represented by  $\bar{y}$  with a suffix ( $z$ ). Similarly, the lifted Fourier transform is represented by  $\bar{y}$  with a suffix ( $e^{j\theta}$ ). Most of the time it is clear from the context if the signal is in lifted (time) domain or lifted frequency domain (lifted  $z$ -transform or Fourier transform). In case it is really necessary to make a distinction, we use  $\mathfrak{Z}(\bar{y})$  for lifted  $z$ -transform and  $\mathfrak{F}(\bar{y})$  for lifted Fourier transform of the signal  $\bar{y}$ . With a little bit of overloading the notations, the classic discrete time Fourier transform of a discrete signal  $\bar{y}$  with interval  $h'$  is represented with different arguments as  $\bar{y}(e^{j\nu})$ . Most of the time the signal domain is clear from the context. Sometimes to make distinction with the time domain (or due to historic reasons) we represent the classic discrete time Fourier transform of  $\bar{y}$  in capitals as  $\bar{Y}(e^{j\nu})$ . In addition to tables 2.1 and 2.2, the Table 4.2 (page 60) summarizes notation for downsamplers and discrete signals with interval  $h'$  in different domains.

SVD stands for singular value decomposition. Also the left singular vectors of an operator  $A$  are the eigenvectors of  $AA^*$  and the right singular vectors of an operator  $A$  are the eigenvectors of  $A^*A$ .

Throughout

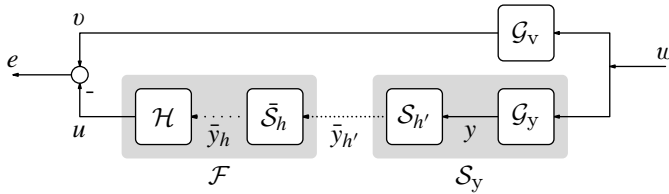
$$\omega_k := \frac{\theta + 2\pi k}{h}, \quad k \in \mathbb{Z}$$

for a given  $\theta \in [-\pi, \pi]$ .

	time domain	lifted time domain	lifted frequency domain		classic frequency domain
Downsampler	$\bar{\mathcal{S}}_h$	$\acute{\mathcal{S}}_h$	$\acute{\mathcal{S}}_h$		
Discrete signal with interval $h'$	$\bar{y}$	$\vec{y}$	$\frac{z}{\mathfrak{Z}(\vec{y})}$	$\frac{\text{Fourier transform}}{\mathfrak{F}(\vec{y})}$	$\frac{\bar{y}(e^{j\nu})}{\bar{Y}(e^{j\nu})}$

**Table 4.2:** Notations for downsampler and discrete signal (with interval  $h'$ ) in different domains

### 4.3 Sampled data setup for downsampling



**Figure 4.3:** Downsampling in sampled-data setting

In this section we review standard mathematical descriptions of the components of the sampled-data set up given in Figure 4.1 (for details see Chapter 2).

The sampled data setup for the downsampling problem is shown in more detail in Figure 4.3. Here

- the *signal generator*  $\mathcal{G} = \begin{bmatrix} \mathcal{G}_v \\ \mathcal{G}_y \end{bmatrix}$  is an LCTI system with finite  $L^2$  and  $L^\infty$  system norm [70]. The systems  $\mathcal{G}_v$  and  $\mathcal{G}_y$  are the partition of  $\mathcal{G}$  with respect to the signals  $v$  and  $y$  respectively.
- the *sampler*  $\mathcal{S}_{h'}$  is a linear  $h'$ -time shift invariant device which samples an analog signal  $y : \mathbb{R} \rightarrow \mathbb{C}$  at every  $h'$  time instant and gives a discrete signal  $\bar{y}_{h'} : \mathbb{Z} \rightarrow \mathbb{C}$ . More specifically

$$\bar{y}_{h'} = \mathcal{S}_{h'} y : \quad \bar{y}_{h'}[n] = \int_{-\infty}^{\infty} \psi(nh' - s)y(s) ds. \quad (4.1)$$

where  $\psi(t)$  is such that its classic frequency response  $\psi(j\omega)$  is uniformly bounded. The  $\psi(t)$  is known as the *sampling function*. The only difference

between the sampler  $\mathcal{S}_{h'}$  defined here and the sampler defined in (2.2) of Chapter 2 is that  $\mathcal{S}_{h'}$  samples faster than the sampler  $\mathcal{S}$  in (2.2).

The condition that  $\psi(j\omega)$  is uniformly bounded is needed to allow the ideal sampler  $\mathcal{S}_{\text{idl}}$  given in (2.3) even though it is not stable [30]. The ideal sampler has sampling function  $\delta(t)$  with uniformly bounded classic fourier transform  $\delta(j\omega) = 1$ .

- The *downsampler*  $\bar{\mathcal{S}}_h$  is a device which operates on a discrete signal  $\bar{y}_{h'} : \mathbb{Z} \rightarrow \mathbb{C}$  with interval  $h'$  and produces a discrete signal  $\bar{y}_h : \mathbb{Z} \rightarrow \mathbb{C}^r$  with interval  $h = Mh'$  at the output. Here  $M$  is called *downsampling factor* and  $r$  is a positive integer. We assume the downsampler to be linear and  $M$  shift invariant, i.e.

$$\bar{y}_h = \bar{\mathcal{S}}_h \bar{y}_{h'} : \quad \bar{y}_h[n] = \sum_{k \in \mathbb{Z}} \bar{\chi}[Mn - k] \bar{y}_{h'}[k], \quad n \in \mathbb{Z} \quad (4.2)$$

where  $\bar{\chi} = \begin{bmatrix} \bar{\chi}_0 \\ \vdots \\ \bar{\chi}_{r-1} \end{bmatrix}$  and  $\bar{\chi}_i$  is a sequence mapping  $\mathbb{Z}$  to  $\mathbb{C}$  for all  $i \in$

$[0, \dots, r-1]$ . The  $\bar{\chi}$  is known as *discrete sampling function*. When  $\bar{\chi} = \bar{\delta}$  where  $\bar{\delta}$  is unit impulse sequence, then the downsampler is called as *ideal downsampler* and will be denoted by  $\bar{\mathcal{S}}_{\text{idl}}$ . Note that any downsampler can be seen as a series combination of  $h'$ -shift invariant discrete system with impulse response  $\bar{\chi}[n]$  followed by the ideal downsampler  $\bar{\mathcal{S}}_{\text{idl}}$ .

Since in the sampled-data system theory time between the samples also play an important role, therefore the  $M$ -shift invariance of the downsamplers is equivalent to  $h = Mh'$  time shift invariance.

- The *hold*  $\mathcal{H}$  is a device which converts a discrete signal  $\bar{y}_h : \mathbb{Z} \rightarrow \mathbb{C}^r$  with interval  $h$  back to an analog signal  $u : \mathbb{R} \rightarrow \mathbb{C}$ . Here  $r$  is a positive integer. We assume  $\mathcal{H}$  to be linear and  $h$ -time shift invariant, i.e.

$$u = \mathcal{H} \bar{y}_h : \quad u(t) = \sum_{n \in \mathbb{Z}} \phi(t - nh) \bar{y}_h[n], \quad t \in \mathbb{R}$$

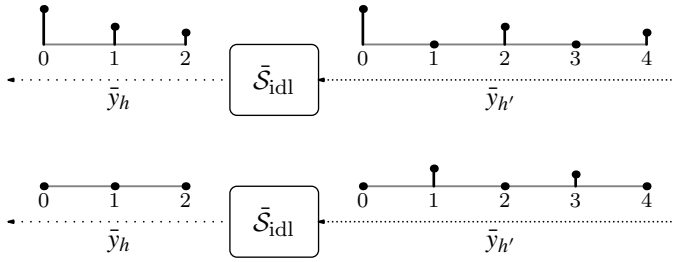
where  $\phi(t) := [\phi_0(t) \quad \phi_1(t) \quad \dots \quad \phi_{r-1}(t)]$  and  $\phi_i : \mathbb{R} \rightarrow \mathbb{C}$  for all  $i \in [0, \dots, r-1]$ . The  $\phi(t)$  is known as *hold function* or *interpolating kernel*.

Since the output of the downsampler is the input of the hold, the output dimension of the downsampler and the input dimension of the hold must be same.

Although in earlier chapters the terms interpolator and hold are used to denote the same operator, in this chapter we use interpolator to mean the following.

**Definition 4.3.1** (Interpolator). *An interpolator  $\mathcal{F}$  is a device which converts a discrete signal  $\bar{y}_{h'} : \mathbb{Z} \rightarrow \mathbb{C}$  with interval  $h'$  back to an analog signal  $u : \mathbb{R} \rightarrow \mathbb{C}$ .*





**Figure 4.4:** The output of the ideal downsampler  $\bar{S}_{idl}$  with down-sampling factor 2 (top). The output is not delayed by a sample if the input is delayed by a sample (bottom).

Note that in this chapter, the interpolator's input is a discrete signal with interval  $h'$  whereas the hold's input is a discrete signal with interval  $h$ .

A downsampler followed by a hold (i.e.  $\mathcal{H}\bar{S}_h$ ) is an important component throughout this chapter, so it deserves a special name:

**Definition 4.3.2** (Hybrid interpolator). *Any interpolator  $\mathcal{F}$  is a hybrid interpolator of order  $r$  if it can be represented as a downsampler (with output dimension  $r$ ) followed by a hold (with input dimension  $r$ ) i.e.  $\mathcal{F} := \mathcal{H}\bar{S}_h$ .*

**Remark 4.3.3.** *In this chapter, we assumed the dimensions of signals  $y$ ,  $v$ ,  $\bar{y}_{h'}$  and  $u$  in the sampled-data setup (see Figure 4.3) are one. This is just for simplicity. The result can be easily generalized to the case where the dimensions are greater than one.*

## 4.4 Discrete lifting in time domain

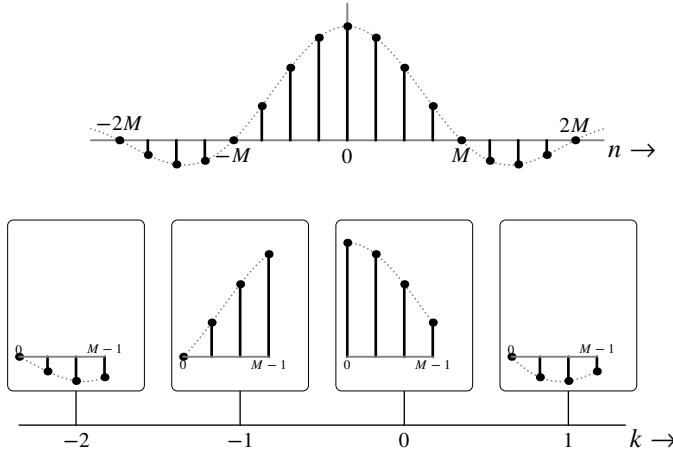
In general a downsampler given in (4.2) is a linear but not shift invariant system (i.e. if we shift the input by one then its output is not delayed by one). This fact can be seen in the following example.

**Example 4.4.1.** *Consider the ideal downsampler  $\bar{S}_{idl}$  that downsamples the input signal by a factor of 2,*

$$\bar{y}_h = \bar{S}_{idl}\bar{y}_{h'} : \quad \bar{y}_h[n] = \bar{y}_{h'}[2n], \quad n \in \mathbb{Z}.$$

*Assume that the input  $\bar{y}_{h'}[n]$  is non-zero only if  $n$  is even. Clearly,  $\bar{y}_h[n]$  is non-zero for this input (see Figure 4.4(top)). However, if we delay our input  $\bar{y}_{h'}$  by one sample then output is not delayed by one sample. In fact, we have zero output in this particular case (see Figure 4.4(bottom)).*

Example 4.4.1 clearly shows that downsamplers are not shift invariant system in general. This implies that techniques like transfer functions, frequency responses etc. available for the linear shift invariant system cannot be written for



**Figure 4.5:** Discrete lifting  $\vec{f}$  of  $\bar{f}[n] = \text{sinc}(n/M)$  with  $M = 4$ .

a downsampler. However, a downsampler with downsampling factor  $M$  is  $M$  shift invariant. This means that if the input of the downsampler is delayed by  $M$  then the output is delayed by 1 as shown below.

$$\begin{aligned} \bar{y}_h[n-1] &= \sum_{k \in \mathbb{Z}} \bar{\chi}[M(n-1)-k] \bar{y}_{h'}[k] = \sum_{k \in \mathbb{Z}} \bar{\chi}[Mn - (k+M)] \bar{y}_{h'}[k] \\ &= \sum_{k \in \mathbb{Z}} \bar{\chi}[Mn - k] \bar{y}_{h'}[k - M]. \end{aligned}$$

To write the transfer function of the downsampler we need to transform the downsampler such that it is shift invariant. It means if we delay the input of the transformed downsampler by 1 then its output must be delayed by 1 only. This goal can be achieved if we *lift* the input of the downsampler [57]. Similar to lifting of an analog signal, we can define lifting of a discrete signal with sampling interval  $h'$  as (see e.g. [65, 41] and the references cited therein).

**Definition 4.4.2.** For any discrete signal  $\bar{f} : \mathbb{Z} \rightarrow \mathbb{C}^{n_f}$  sampled at interval  $h' = h/M$ , the lifting (with respect to  $h$ )  $\vec{f} : \mathbb{Z} \rightarrow \{\mathbb{M} \rightarrow \mathbb{C}^{n_f}\}$  is the sequence of discrete functions  $\{\vec{f}[k]\}$  defined as

$$\vec{f}[k; m] = \bar{f}[Mk + m], \quad k \in \mathbb{Z}, m \in \mathbb{M}$$

Here  $n_f$  is an positive integer and  $\mathbb{M} := \{0, 1, \dots, M-1\}$ .

For a given  $k$ ,  $\vec{f}[k]$  is a finite sequence which maps  $\mathbb{M}$  to  $\mathbb{C}^{n_f}$ . Recall that the lifting of an analog signal  $f : \mathbb{R} \rightarrow \mathbb{C}^{n_f}$  results in lifted signal  $\vec{f}$  where  $\vec{f}[k]$  is a function which maps  $[0, h)$  to  $\mathbb{C}^{n_f}$  for a given  $k$ . Figure 4.5 shows the concept of discrete lifting.

**Remark 4.4.3.** The natural domain for  $m$  in Definition 4.4.2 is  $\mathbb{M}$  because then there exists a bijection between the signal  $f$  and its lifted signal  $\tilde{f}$ . However, sometimes it is beneficial to define  $\tilde{f}[k; m] := f(kh + m)$  for arbitrary  $m \in \mathbb{Z}$ .

The discrete time lifting is similar to the *polyphase decomposition* (see e.g. [61]). It is clear from the above definition that if a discrete signal  $\tilde{f}$  is generated by sampling an analog signal with sampling period  $h$  i.e.  $h' = h$  and  $M = 1$  then its discrete lifting is the same as the discrete signal  $\tilde{f}$ . For this reason the discrete lifting of such a signal is represented by  $\tilde{f}$ .

Now, lifting the input  $\tilde{y}_{h'}$  turns the downsampler  $\tilde{y}_h = \tilde{S}_h \tilde{y}_{h'}$  in (4.2) into *lifted downsampler*  $\hat{S}_h$  which is given by the following convolution

$$\begin{aligned} \tilde{y}_h = \hat{S}_h \tilde{y}_{h'} : \quad \tilde{y}_h[n] &= \sum_{i \in \mathbb{Z}} \sum_{m \in \mathbb{M}} \tilde{\chi}[Mn - Mi - m] \tilde{y}_{h'}[Mi + m], \quad n \in \mathbb{Z} \\ &= \sum_{i \in \mathbb{Z}} \sum_{m \in \mathbb{M}} \tilde{\chi}[n - i; -m] \tilde{y}_{h'}[i; m] \\ &=: \sum_{i \in \mathbb{Z}} \hat{S}_h[n - i] \tilde{y}_{h'}[i] \end{aligned} \quad (4.3)$$

where for each  $k \in \mathbb{Z}$ ,  $\hat{S}_h[k] : \{\mathbb{M} \rightarrow \mathbb{C}\} \rightarrow \mathbb{C}$  is the *lifted impulse response system* of the lifted system  $\hat{S}_h$ . The lifted system  $\hat{S}_h$  is shift invariant as

$$\begin{aligned} \tilde{y}_h[n + 1] &= \sum_{i \in \mathbb{Z}} \hat{S}_h[n + 1 - i] \tilde{y}_{h'}[i] \\ &= \sum_{i \in \mathbb{Z}} \hat{S}_h[n - i] \tilde{y}_{h'}[i + 1] \end{aligned}$$

Since the sampler  $S_{h'}$  samples at interval  $h'$  which is different from  $h$ , we can (discrete) lift the output of the sampler  $S_{h'}$  with respect to  $h$ . Lifting the input and output of the sampler  $\tilde{y}_{h'} = S_{h'} y$  result in the *lifted sampler*

$$\begin{aligned} \tilde{y}_{h'} = \hat{S}_{h'} \check{y} : \quad \tilde{y}_{h'}[k; m] &= \sum_{i \in \mathbb{Z}} \int_0^h (\psi(Mkh' + mh' - (hi + \tau))) \times \\ &\quad y(hi + \tau) d\tau, \quad k \in \mathbb{Z}, m \in \mathbb{M} \\ &= \sum_{i \in \mathbb{Z}} \int_0^h (\psi((k - i)h + (mh' - \tau))) \check{y}[i](\tau) d\tau \\ &= \sum_{i \in \mathbb{Z}} \int_0^h (\check{\psi}[k - i](mh' - \tau)) \check{y}[i](\tau) d\tau \\ &=: \sum_{i \in \mathbb{Z}} \hat{S}_{h'}[k - i] \check{y}[i] \end{aligned} \quad (4.4)$$

where for each  $k \in \mathbb{Z}$ ,  $\hat{S}_{h'}[k] : \{[0, h) \rightarrow \mathbb{C}\} \rightarrow \{\mathbb{M} \rightarrow \mathbb{C}\}$  is the *lifted impulse response system* of the lifted system  $\hat{S}_{h'}$ . Clearly the system  $\hat{S}_{h'}$  is shift invariant.

The advantage of (discrete/analog) lifting is that the lifted sampler  $\tilde{S}_{h'}$  and the lifted downsampler  $\tilde{S}_h$  are shift invariant which will allow us later to write a transfer function and frequency response of these systems in Section 4.5.1. However, this comes at a price of more difficult impulse responses.

For lifted  $\tilde{G}$  and lifted Hold  $\tilde{H}$  see Section 2.3.1 of Chapter 2.

## 4.5 Discrete lifting in frequency domain

Similar to the lifted  $z$ -transform of analog signals in Section 2.3.2, we can define the lifted  $z$ -transform for discrete signals.

**Definition 4.5.1.** The (discrete)  $z$ -transform  $\vec{f}(z; m)$  of a lifted discrete signal  $\vec{f}$  is defined as

$$\mathfrak{Z}(\vec{f}) = \vec{f}(z; m) := \sum_{k \in \mathbb{Z}} \vec{f}[k; m] z^{-k} = \sum_{k \in \mathbb{Z}} \vec{f}[Mk + m] z^{-k} \quad (4.5)$$

where  $m \in \mathbb{M}$ .  $\mathfrak{Z}(\vec{f})$  is also called (discrete) lifted  $z$ -transform of the signal  $\vec{f}$ .

It is clear from the above definition that if a discrete signal  $\vec{f}$  is generated by sampling an analog signal with sampling period  $h$  then its  $z$ -transform is the same as its lifted  $z$ -transform. For this reason the lifted  $z$ -transform of such a signal is represented by  $\vec{f}(z)$ . Similarly the (lifted) Fourier transform is defined as:

**Definition 4.5.2.** The (discrete) Fourier transform  $\vec{f}(e^{j\theta}; m)$  of a lifted discrete signal  $\vec{f}$  is defined as

$$\mathfrak{F}(\vec{f}) = \vec{f}(e^{j\theta}; m) := \sum_{k \in \mathbb{Z}} \vec{f}[k; m] e^{-j\theta k} = \sum_{k \in \mathbb{Z}} \vec{f}[Mk + m] e^{-j\theta k} \quad (4.6)$$

where  $m \in \mathbb{M}$ .  $\mathfrak{F}(\vec{f})$  is also called (discrete) lifted Fourier transform of the signal  $\vec{f}$ .

Clearly, for a given  $m$ ,  $\vec{f}(e^{j\theta}; m)$  is the discrete time Fourier transform (see [27]) of the sequence  $\vec{f}[Mk + m]$  and  $\vec{f}(e^{j\theta}; m)$  is periodic in  $\theta$  with period  $2\pi$  as  $\vec{f}(e^{j(\theta+2\pi)}; m) = \vec{f}(e^{j\theta}; m)$ .

Since lifting and  $z$ -transform are invertible processes, the inverse lifted  $z$ -transform can be defined as a combination of the inverse  $z$ -transform and inverse of the lifting process. The following example illustrates calculation of the (lifted) Fourier transform.

**Example 4.5.3.** Consider the analog signal  $f(t) = \mathbb{1}_{[0, h)}(t)$ . Ideally sampling  $f$  at interval  $h'$ , we have

$$\vec{f}[k] = \begin{cases} 1 & k \in \mathbb{M} \\ 0 & \text{elsewhere} \end{cases}$$

By Definition 4.4.2, the lifted signal  $\vec{f}$  of  $f$  is given by

$$\vec{f}[k; m] = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

for all  $m \in \mathbb{M}$ . Hence, the lifted Fourier transform of  $\vec{f}$  is  $\vec{f}(e^{j\theta}; m) = 1$  for all  $m \in \mathbb{M}$ .

Following is a key result which relates the lifted  $z$ -transform of a discrete signal  $\vec{f}$  with time interval  $h'$  and its the classic  $z$ -transform.

**Theorem 4.5.4** (Discrete key lifting formula). *Consider a discrete signal  $\vec{f} : \mathbb{Z} \rightarrow \mathbb{C}$  with time interval  $h'$  and let  $\bar{f}(z)$  be its classic  $z$ -transform. If  $\bar{f}(z)$  exists then the lifted  $z$ -transform  $\vec{f}(z; m)$  of  $\vec{f}[n]$  follows from its classic  $z$ -transform as*

$$\vec{f}(z; m) = \frac{1}{M} \sum_{i=0}^{M-1} \bar{f}(\tilde{z}_i) \tilde{z}_i^m \quad m \in \mathbb{M} \quad (4.7)$$

where  $\tilde{z}_i = |z|^{\frac{1}{M}} e^{j \frac{\text{Arg} z}{M}} e^{j \frac{2\pi i}{M}}$  ( $\text{Arg}$  is the principle argument function used in (3.7)) are the  $M$  complex roots of the equation  $\tilde{z}^M = z$  for a given  $z$ . Conversely the classic  $z$ -transform of  $\bar{f}$  follows from the lifted  $z$ -transform as

$$\bar{f}(\tilde{z}) = \sum_{m=0}^{M-1} \vec{f}(\tilde{z}^M; m) \tilde{z}^{-m} \quad (4.8)$$

for a given  $\tilde{z}$ .

*Proof.* See Appendix 4.A (page 100). □

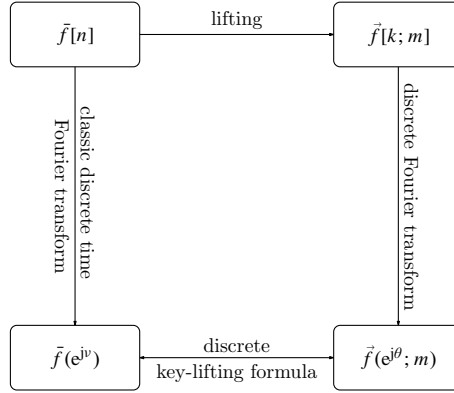
The following corollary is immediate.

**Corollary 4.5.5.** *Consider a discrete signal  $\vec{f} : \mathbb{Z} \rightarrow \mathbb{C}$  with time interval  $h'$  and let  $\bar{f}(e^{j\nu})$ ,  $\nu \in [-\pi, \pi]$  represents its classic discrete time Fourier transform [27, chap. 10]. If  $\bar{f}(e^{j\nu})$  exists then the lifted Fourier transform  $\vec{f}(e^{j\theta}; m)$  of  $\vec{f}[n]$  follows from  $\bar{f}(e^{j\nu})$  as*

$$\vec{f}(e^{j\theta}; m) = \frac{1}{M} \sum_{i=0}^{M-1} \bar{f}(e^{j\omega_i h'}) e^{j\omega_i h' m} \quad (4.9)$$

where as always  $\omega_i = \frac{\theta + 2\pi i}{h}$ . Conversely  $\bar{f}(e^{j\nu})$  follows from the lifted fourier transform as

$$\bar{f}(e^{j\nu}) = \sum_{m=0}^{M-1} \vec{f}(e^{j\nu M}; m) e^{-j\nu m} \quad (4.10)$$



**Figure 4.6:** Relation between various transforms and discrete key lifting formula.

*Proof.* Set  $z = e^{j\theta}$  in (4.7) and  $\tilde{z} = e^{j\nu}$  in (4.8). □

**Remark 4.5.6.** It is interesting to note that  $e^{j\omega_0 h' m}, \dots, e^{j(\omega_0 M - 1) h' m}$  (used in (4.9)) are mutually orthogonal. Let us define  $\vec{e}_i(e^{j\theta}) \in \mathbb{C}^M$  whose  $m$ 'th component is given by  $\vec{e}_i(e^{j\theta}; m) := \frac{1}{\sqrt{M}} e^{j\omega_i h' m}$ . Now,  $\{\vec{e}_i(e^{j\theta})\}_{i \in \mathbb{M}}$  form an orthonormal basis of the space  $\mathbb{C}^M$  as  $\langle \vec{e}_k(e^{j\theta}), \vec{e}_i(e^{j\theta}) \rangle = \sum_{m=0}^{M-1} \vec{e}_i(e^{j\theta}; m) \vec{e}_k(e^{-j\theta}; m) = \delta[k - i]$ . The above fact will help later in obtaining the SVD of a sampler.

Figure 4.6 explains the relation between the various transforms and the discrete key lifting formula.

The following examples illustrates the use of (4.9).

**Example 4.5.7.** Consider an analog signal  $f = \text{sinc}(\frac{t}{h})$ . Sampling  $f$  at interval  $h' = \frac{h}{M}$ , we have discrete signal  $\tilde{f}[n] = \text{sinc}(n/M)$ . The discrete time Fourier transform of  $\tilde{f}$  is given by

$$\tilde{f}(e^{j\nu}) = \begin{cases} 1 & \nu \in [-\frac{\pi}{M}, \frac{\pi}{M}] \\ 0 & \text{elsewhere} \end{cases}$$

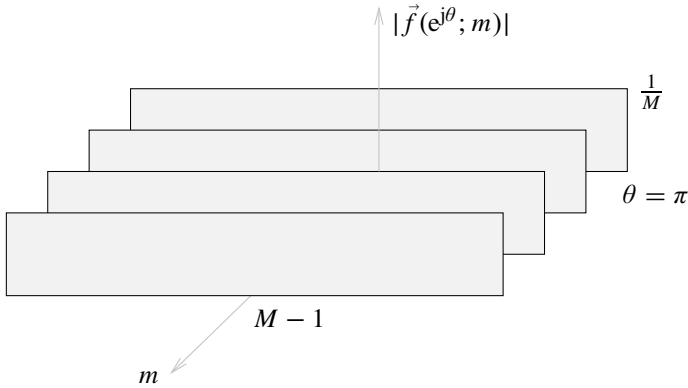
Therefore for  $\theta \in [-\pi, \pi]$ ,

$$\tilde{f}(e^{j\omega_i h'}) = \tilde{f}(e^{j(\frac{\theta + 2\pi i}{M})}) = \begin{cases} 1 & i = kM, k \in \mathbb{Z} \\ 0 & \text{elsewhere} \end{cases}.$$

Now, (4.9) gives the lifted Fourier transform

$$\vec{f}(e^{j\theta}; m) = \frac{1}{M} \tilde{f}(e^{j\omega_0 h'}) e^{j\omega_0 h' m} = \frac{1}{M} e^{j\theta \frac{m}{M}}.$$

See Figure 4.7 (page 68).



**Figure 4.7:** Magnitude plot of the lifted fourier transform  $\vec{f}(e^{j\theta}; m)$  of  $\tilde{f}[n] = \text{sinc}(n/M)$ . Shown here for  $\theta \in [-\pi, \pi]$  only.

### 4.5.1 Transfer function in lifted frequency domain

The lifted transfer function of  $\mathcal{G}$  and hold is already discussed in Section 2.3.2. In this section, the transfer functions of lifted downsamplers and samplers (with sampling period  $h'$ ) are defined.

As in Section 2.3.2, the transfer function of a linear  $h$ -time shift invariant system is defined as the  $z$ -transform of its lifted impulse response system.

Taking  $z$ -transform of the output  $\vec{y}$  and the input  $\vec{y}_h$  of a lifted downsampler  $\hat{S}_h$  in (4.3) results in *lifted frequency domain downsampler*

$$\mathfrak{Z}(\vec{y}) = \hat{S}_h \mathfrak{Z}(\vec{y}_h) : \quad \vec{y}(z) = \hat{S}_h(z) \vec{y}_h(z)$$

where  $\hat{S}_h(z) : \{\mathbb{M} \rightarrow \mathbb{C}\} \rightarrow \mathbb{C}^r$  (it is the  $z$ -transform of the lifted impulse response system of the lifted sampler  $\hat{S}_h$ ) is given by

$$\vec{y}(z) = \hat{S}_h(z) \vec{y}_h(z) : \quad \vec{y}(z) = \sum_{m \in \mathbb{M}} \vec{\chi}(z; -m) \vec{y}_h(z; m). \quad (4.11)$$

Here  $\vec{\chi}(z)$  is the lifted  $z$ -transform (with lifting interval  $h$ ) of the downsampling function  $\chi[n]$ .  $\hat{S}_h(z)$  is called the *transfer function* of the lifted system  $\hat{S}_h$  (or *lifted transfer function* of  $\tilde{S}_h$ ). By the above equation, the transfer function  $\hat{S}_h(z)$  is an operator whose kernel is given by  $\vec{\chi}(z, -m)$ . The derivation of (4.11) is given in Appendix 4.A (page 102).

Similarly, taking  $z$ -transform of the output  $\vec{y}_{h'}$  and the input  $\vec{y}$  of a lifted sampler  $\hat{S}_{h'}$  in (4.4) results in *lifted frequency domain sampler*

$$\mathfrak{Z}(\vec{y}_{h'}) = \hat{S}_{h'} \mathfrak{Z}(\vec{y}) : \quad \vec{y}_{h'}(z) = \hat{S}_{h'}(z) \vec{y}(z)$$

where  $\hat{S}_{h'}(z) : \{[0, h) \rightarrow \mathbb{C}\} \rightarrow \{\mathbb{M} \rightarrow \mathbb{C}\}$  (it is the  $z$ -transform of the lifted

impulse response system of the lifted sampler  $\hat{S}_{h'}$  is given by

$$\begin{aligned} \vec{y}_{h'}(z) = \hat{S}_{h'}(z)\check{y}(z) : \vec{y}_{h'}(z; m) &= \int_0^h \left( \sum_{n \in \mathbb{Z}} \check{\psi}[n](mh' - \tau) \right) z^{-n} \check{y}(z; \tau) d\tau \\ &= \int_0^h \check{\psi}(z; mh' - \tau) \check{y}(z; \tau) d\tau \end{aligned} \quad (4.12)$$

Here  $\check{\psi}(z)$  is the lifted  $z$ -transform (with lifting interval  $h$ ) of the sampling function  $\psi(t)$ .  $\hat{S}_{h'}(z)$  is called the *transfer function* of the lifted system  $\hat{S}_{h'}$  (or *lifted transfer function* of  $S_{h'}$ ). By the above equation, the transfer function  $\hat{S}_{h'}(z)$  is an operator whose kernel is given by  $\check{\psi}(z; mh' - \tau)$ .

## 4.6 Signal and System norms

In this section we define different norms for discrete lifted signals and systems. This section closely follows Section 2.4.

Given a signal  $f \in \ell_{h'}^2$  (the samples in the discrete signal are separated by fixed  $h'$  time). Consider the norm of a discrete signal  $f \in \ell_{h'}^2(\mathbb{Z}, \mathbb{C}^n)$ ,

$$\begin{aligned} \|f\|_2^2 &= \sum_{i \in \mathbb{Z}} \|f_i\|_2^2 = \sum_{k \in \mathbb{Z}} \sum_{m \in \mathbb{M}} \|f(kh + m)\|_2^2 \\ &= \sum_{k \in \mathbb{Z}} \sum_{m \in \mathbb{M}} \|\vec{f}[k; m]\|_2^2 \\ &= \sum_{k \in \mathbb{Z}} \|\vec{f}[k]\|_{\ell^2(\mathbb{M})}^2 =: \|\vec{f}\|_2^2. \end{aligned} \quad (4.13)$$

This defines the norm of a discrete lifted signal  $\vec{f}$ . This also shows that lifting a discrete signal  $f \in \ell_{h'}^2$ , by definition results in a lifted signal  $\vec{f}$  with the same norm in the Hilbert space  $\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))$ . Here  $\ell^2(\mathbb{M}) := \ell^2(\mathbb{M}, \mathbb{C}^n)$ . The space  $\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))$  has inner product

$$\langle \check{x}, \check{y} \rangle = \sum_{k \in \mathbb{Z}} \sum_{m \in \mathbb{M}} \langle \vec{x}[k; m], \vec{y}[k; m] \rangle_{\mathbb{C}^n}.$$

Equation (4.13) also implies that  $\ell_{h'}^2$  is isomorphic to  $\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))$  [67, §7.4]. Therefore, the norms in both of the spaces can be denoted by  $\|\cdot\|_2$ .

If a discrete lifted signal  $\vec{f}$  is in  $\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))$  then its Fourier transform  $\mathfrak{F}(\vec{f})$  belongs to  $L^2(\mathbb{T}, \ell^2(\mathbb{M}))$  because

$$\begin{aligned} \|\mathfrak{F}(\vec{f})\|_2^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\vec{f}(e^{j\theta})\|_{\ell^2(\mathbb{M})}^2 d\theta = \sum_{k \in \mathbb{Z}} \|\vec{f}[k]\|_{\ell^2(\mathbb{M})}^2 \\ &= \|\vec{f}\|_{\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))}^2 = \|\vec{f}\|_{\ell_{h'}^2}^2 < \infty \end{aligned}$$



The equivalence of  $\|\check{\mathfrak{F}}(\check{f})\|_2 = \|\check{f}\|_{\ell^2(\mathbb{Z}, L^2[0, h])}$  is known as *Parseval identity* which says that  $\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))$  is isomorphic to  $L^2(\mathbb{T}, \ell^2(\mathbb{M}))$ .

**Remark 4.6.1.** *The space  $\ell^2(\mathbb{M}, \mathbb{C}^n)$  is isomorphic to space  $\mathbb{C}^{n \times M}$  with Hilbert-Schmidt norm. Therefore, sometimes we write  $\ell^2(\mathbb{M}, \mathbb{C})$  as  $\mathbb{C}^M$ .*

### 4.6.1 System norms

To include systems whose domain/range belongs to  $\ell_{h'}^2$  (e.g. downsampler  $\bar{S}_h$  and fast sampler  $S_{h'}$ ) in the definitions of the system norms  $L^\infty$ ,  $z^l H^\infty$ ,  $L^2$  and  $z^l H^2$ , we have to change  $\bar{S}$  in Section 2.4.3 to mean  $L^2[0, h)$  or  $\mathbb{C}^n$  or  $\ell^2(\mathbb{M}, \mathbb{C}^n)$ . Consequently, the meaning of  $\bar{S}_i$ ,  $\bar{S}_o$ ,  $\bar{\mathbb{R}}$  and  $\bar{\mathbb{R}}_{HS}$  is modified. For example, the  $L^\infty$  norm of the downsampler  $\bar{S}_h$  equals

$$\|\bar{S}_h\|_{L^\infty} = \operatorname{ess\,sup}_{\theta \in [-\pi, \pi]} \|\hat{S}_h(e^{j\theta})\| < \infty \quad (4.14)$$

where  $\hat{S}_h(e^{j\theta})$  is the (lifted) transfer function of  $S_h$  and  $\|\hat{S}_h(e^{j\theta})\|$  is given by

$$\|\hat{S}_h(e^{j\theta})\| = \sup_{\|x\|_2=1} \|\hat{S}_h(e^{j\theta})x\|_2 \quad x \in \ell^2(\mathbb{M}, \mathbb{C}^n)$$

A result similar to Lemma 2.4.7 can be stated for downsamplers also.

**Lemma 4.6.2.** *Consider a downsampler given by (4.2). If the downsampling function  $\chi$  of the downsampler belongs to  $\ell_{h'}^2$ , then the downsampler belongs to  $L^\infty \cap L^2$ .*

*Proof.* The proof is similar to the proof of Lemma 2.4.7. □

## 4.7 Rank theorem

In time domain, it is not so clear when a given interpolator, is a hybrid interpolator of a given order  $r$ . However, it is somewhat clear in the (lifted) frequency domain as shown in this section.

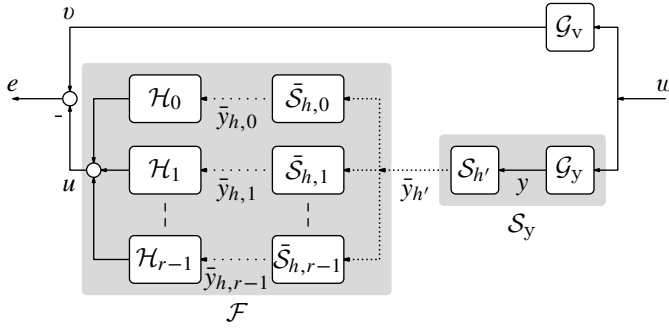
First, we write the hybrid interpolator  $\mathcal{F}$  in the lifted frequency domain.

**Lemma 4.7.1.** *The hybrid interpolator  $\mathcal{F} := \mathcal{H}\bar{S}_h$  in lifted frequency domain is an operator*

$$\check{u} = \check{F}\check{y}_{h'} : \quad \check{u}(e^{j\theta}; \tau) = \sum_{m=0}^{M-1} \check{\varrho}(e^{j\theta}; \tau; m) \check{y}_{h'}(e^{j\theta}; m)$$

where the kernel  $\check{\varrho}$  can be expressed in terms of sampling and hold functions (see Section 4.3) as

$$\check{\varrho}(e^{j\theta}; \tau; m) = \check{\varphi}(e^{j\theta}; \tau) \check{\chi}(e^{j\theta}; -m). \quad (4.15)$$



**Figure 4.8:** Sampled data setup with hybrid interpolator of order- $r$ .

*Proof.*

$$\begin{aligned} \check{u}(e^{j\theta}; \tau) &= \check{\phi}(e^{j\theta}; \tau) \bar{y}_h(e^{j\theta}) = \check{\phi}(e^{j\theta}; \tau) \sum_{m=0}^{M-1} \bar{\chi}(e^{j\theta}; -m) \bar{y}_{h'}(e^{j\theta}; m) \\ &= \sum_{m=0}^{M-1} \check{\phi}(e^{j\theta}; \tau) \bar{\chi}(e^{j\theta}; -m) \bar{y}_{h'}(e^{j\theta}; m) \end{aligned}$$

□

The above lemma is useful if we have an interpolator  $\hat{F}$  and we want to represent it as a cascade of a downsampler and a hold i.e. as a hybrid interpolator.

Now we concentrate on obtaining a condition for an interpolator such that it is a hybrid interpolator. Consider an order  $r$  hybrid interpolator  $\mathcal{F}$  shown in Figure 4.8. In this case the downsampling function  $\bar{\chi}[n]$  of downsampler  $\hat{S}_h$  has  $r$  rows. Clearly the lifted fourier transform  $\bar{y}_h(e^{j\theta})$  of the output  $\bar{y}_h$  of the downsampler has  $r$  rows. Therefore at each  $\theta \in [-\pi, \pi]$ , we have  $\bar{y}_h(e^{j\theta}) \in \mathbb{C}^r$ . Hence,  $\text{rank } \hat{S}_h(e^{j\theta}) \leq r$  at a given  $\theta$ . We can obtain such an upper bound for hybrid interpolator  $\mathcal{F} := \mathcal{H}\hat{S}_h$  also. Since  $\hat{F} = \hat{H}\hat{S}_h$  and  $\text{rank } \hat{S}_h(e^{j\theta}) \leq r$  at a given  $\theta$ , there must exist an upper bound on  $\text{rank } \hat{F}$  at each  $\theta$ . This upper bound is  $r$  as shown in the following lemma.

**Theorem 4.7.2** (Rank theorem). *Given an interpolator  $\mathcal{F}$  and assume that the transfer function  $\hat{F}(e^{j\theta})$  exists. If  $\mathcal{F}$  is a hybrid interpolator of order  $r$  then*

$$\text{rank } \hat{F} \leq r \quad \forall \theta \in [-\pi, \pi].$$

*Proof.* See Appendix 4.A (page 102). □

By Theorem 4.7.2, an interpolator  $\mathcal{F}$  is a hybrid interpolator of order  $r$  only if  $\text{rank}(\hat{F})$  is at most  $r$  for all  $\theta \in [-\pi, \pi]$ .

Since the input of the downsampler is a discrete signal, we can expect that the rank of the downsampler (and hence the rank of hybrid interpolator) in the lifted domain must have an upper bound. Indeed.

**Corollary 4.7.3.** *The rank of a hybrid interpolator  $\mathring{F}(e^{j\theta})$  at a given  $\theta \in [-\pi, \pi]$  is less than or equal to  $M$ .*

*Proof.* Since the input  $\vec{y}_{h'}(e^{j\theta})$  of  $\mathring{S}_h(e^{j\theta})$  in lifted frequency domain is in  $\mathbb{C}^M$  at a fixed  $\theta$ , we have  $(\text{Ker } \mathring{S}_h(e^{j\theta}))^\perp \subseteq \mathbb{C}^M$  at each  $\theta$ . Hence,

$$(\text{Ker } \mathring{F}(e^{j\theta}))^\perp \subseteq (\text{Ker } \mathring{S}_h(e^{j\theta}))^\perp \subseteq \mathbb{C}^M$$

and  $\text{rank}(\mathring{F}(e^{j\theta})) \leq M$  at a given  $\theta$ .  $\square$

Corollary 4.7.3 says that increasing the order  $r$  of a hybrid interpolator beyond  $M$  is unnecessary. *Therefore, without loss of generality we take  $r \leq M$  in this chapter.*

## 4.8 Non-causal downsampling problem

In this section, we define the downsampling problem and simplify it as much as possible using lifting.  $\mathbb{K}$  represents either  $L^2$  or  $L^\infty$  and  $\mathbb{K}$  represents the Hilbert-Schmidt ( $HS$ ) norm if  $\mathbb{K} = L^2$  or induced 2-norm if  $\mathbb{K} = L^\infty$ , in this section. Throughout this section we use short-hand  $\mathcal{S}_y := \mathcal{S}_{h'}\mathcal{G}_y$ . Clearly,  $\mathcal{S}_y$  is a sampler.

We also assume that the output dimension of the downsampler and the input dimension of the hold is given and it is equal to  $r$ .

Now, we state the downsampling problem more precisely.

**Problem  $\mathcal{P}_1$  (Downsampling problem)** : *Given  $\mathcal{G}_v, \mathcal{G}_y \in L^2 \cap L^\infty$ , find a downsampler  $\bar{\mathcal{S}}_h \in L^\infty$  (with output dimension  $r \in \mathbb{Z}^+$ ) and a hold  $\mathcal{H} \in L^\infty$  (with input dimension  $r$ ) such that  $\|\mathcal{G}_v - \mathcal{H}\bar{\mathcal{S}}_h\mathcal{S}_{h'}\mathcal{G}_y\|_{\mathbb{K}}$  is minimized.*

The optimal downsampler and hold in Problem  $\mathcal{P}_1$  can be non-causal. Stability of all systems in the downsampling setup (see Figure 4.3) is an important criteria. This is because we never want to have an unbounded output of a system due to some bounded input or noise at any stage of signal processing.

**Remark 4.8.1.** *As explained in Section 2.6, stability of  $\mathcal{G}_v$  and  $\mathcal{G}_y$  implies that they are operators which maps  $L^2(\mathbb{R})$  to  $L^2(\mathbb{R})$ . Similarly, we need a stable hold  $\mathcal{H}$  and downsampler  $\bar{\mathcal{S}}_h$ , thus  $\mathcal{H}$  is an operator which maps  $\ell^2(\mathbb{Z})$  to  $L^2(\mathbb{R})$  and  $\bar{\mathcal{S}}_h$  is an operator which maps  $\ell^2_{h'}(\mathbb{Z})$  to  $\ell^2(\mathbb{Z})$ .*

Clearly, if we have a solution of the downsampling problem then the hybrid interpolator  $\mathcal{F} := \mathcal{H}\bar{\mathcal{S}}_h$  belongs to  $L^\infty$ . Therefore, in order to solve the downsampling problem, we consider the following problem

**Problem  $\mathcal{P}_2$**  : *Given  $\mathcal{G}_v, \mathcal{G}_y \in L^2 \cap L^\infty$ , find a hybrid interpolator  $\mathcal{F} \in L^\infty$  of at most order  $r \in \mathbb{Z}^+$  such that  $\|\mathcal{G}_v - \mathcal{F}\mathcal{S}_{h'}\mathcal{G}_y\|_{\mathbb{K}}$  is minimized.*

To solve Problem  $\mathcal{P}_1$ , we first obtain a solution to Problem  $\mathcal{P}_2$  and then write the resulting optimal hybrid interpolator as a cascade of a *stable* downsampler

and hold. We will see later that this approach is easier than solving Problem  $\mathcal{P}_1$  directly.

A simplification of Problem  $\mathcal{P}_2$  is possible if  $\mathcal{S}_y := \mathcal{S}_{h'}\mathcal{G}_y \in L^\infty \cap L^2$ , which is indeed the case under some weak assumptions as shown in following Lemma.

**Lemma 4.8.2.** *Given LCTI system  $\mathcal{G}_y \in L^2$  and a sampler  $\mathcal{S}_{h'}$  with sampling function  $\psi(t)$  such that  $|\psi(j\omega)|$  is uniformly bounded in  $\omega$ . Define sampler  $\mathcal{S}_y := \mathcal{S}_{h'}\mathcal{G}_y$  and let  $\psi_y$  be its sampling function. Then,*

1.  $\psi_y(t) := \psi(t) * g_y(t)$  where  $*$  represents the analog convolution operator and  $g_y$  is the impulse response of system  $\mathcal{G}_y$ .
2.  $\psi_y \in L^2$ .
3. For almost all given  $\theta \in [-\pi, \pi]$ , the lifted transfer function  $\hat{S}_y(e^{j\theta})$  of sampler  $\mathcal{S}_y$  is a mapping from  $L^2[0, h)$  to  $\mathbb{C}^M$  and it has an SVD (modulo ordering) of the form

$$\hat{S}_y(e^{j\theta})\check{w} = \sqrt{\frac{M}{h}} \sum_{k \in \mathbb{M}} \alpha_k(e^{j\theta}) \left\langle \check{w}, \check{p}_k(e^{j\theta}) \right\rangle \vec{e}_k(e^{j\theta}), \quad \check{w} \in L^2[0, h) \quad (4.16)$$

where for fixed  $\theta$ ,  $\alpha_k(e^{j\theta}) \in \mathbb{C}$ ,  $\check{e}_k(e^{j\theta}) \in L^2[0, h)$ ,  $\check{p}_k(e^{j\theta}) \in L^2[0, h)$ , and  $\vec{e}_k(e^{j\theta}) \in \mathbb{C}^M$  such that

$$\begin{aligned} \alpha_k(e^{j\theta}) &:= \sqrt{\sum_{i \in \mathbb{Z}} |\psi_y(j\omega_{k+Mi})|^2}, \\ \check{e}_k(e^{j\theta}; \tau) &:= \frac{1}{\sqrt{h}} e^{j\omega_k \tau}, \quad \tau \in [0, h) \\ \check{p}_k(e^{j\theta}; \sigma) &:= \sum_{i \in \mathbb{Z}} \frac{\psi_y^*(j\omega_{k+Mi})}{\alpha_k} \check{e}_{k+Mi}(e^{j\theta}; \sigma), \quad \sigma \in [0, h) \\ \vec{e}_k(e^{j\theta}; m) &:= \frac{1}{\sqrt{M}} e^{j\omega_k m h'}, \quad m \in \mathbb{M}. \end{aligned}$$

Here  $\omega_k = \frac{\theta + 2\pi k}{h}$ ,  $k \in \mathbb{Z}$ .

4.  $\mathcal{S}_y \in L^\infty \cap L^2$ .

*Proof.* See Appendix 4.A (page 102). □

Although the ideal sampler is not stable, the above lemma allows us to take  $\mathcal{S}_{h'}$  to be an ideal sampler as long as the LCTI system  $\mathcal{G}_y \in L^2$ .

**Remark 4.8.3.** *As explained in Section 2.6, stability of sampler  $\mathcal{S}_y$  implies that it is an operator which maps  $L^2(\mathbb{R})$  to  $\ell^2(\mathbb{Z})$ .*

Theorem 4.7.2 says that an interpolator  $\mathcal{F}$  is a hybrid interpolator of order  $r$  only if  $\text{rank}(\tilde{F}) \leq r$  at all  $\theta \in [-\pi, \pi]$ . This result is used to solve Problem  $\mathcal{P}_2$  by translating it to an equivalent lifted frequency domain problem as shown below.

**Theorem 4.8.4.** *Given  $\mathcal{G}_v, \mathcal{G}_y \in L^\infty \cap L^2$ . Define  $\tilde{F}_{\text{opt}}$  at almost every  $\theta \in [-\pi, \pi]$  as*

$$\tilde{F}_{\text{opt}}(e^{j\theta}) := \arg \min_{\tilde{F}(e^{j\theta})} \|\check{\mathcal{G}}_v(e^{j\theta}) - \tilde{F}(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_{\mathbb{K}} \quad (4.17)$$

with constraint  $\text{rank } \tilde{F}_{\text{opt}} \leq r$  at each  $\theta \in [-\pi, \pi]$ . If  $\tilde{F}_{\text{opt}}$  is well-defined and bounded, then  $\mathcal{F}_{\text{opt}}$  is an interpolator that minimizes  $\|\mathcal{G}_v - \mathcal{F}\mathcal{S}_y\|_{\mathbb{K}}$  over all interpolators  $\mathcal{F}$  with  $\text{rank } \tilde{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ .

*Proof.* See Appendix 4.A (page 103). □

If interpolator  $\mathcal{F}_{\text{opt}}$  in the above theorem is also a hybrid interpolator then the above theorem says that doing point-wise minimization in lifted Fourier domain is sufficient to solve Problem  $\mathcal{P}_2$ . Moreover, if interpolator  $\mathcal{F}_{\text{opt}}$  in the above theorem is also a hybrid interpolator with *stable* downsampler and hold then we have a solution of Problem  $\mathcal{P}_1$ . Therefore, in order to solve the downsampling problem, we consider the following problem.

**Problem  $\mathcal{P}_3$  :** *Given  $\mathcal{G}_v, \mathcal{G}_y \in L^2 \cap L^\infty$ , find a well-defined and bounded  $\tilde{F}(e^{j\theta})$  with  $\text{rank } \tilde{F}(e^{j\theta}) \leq r$  such that  $\|\check{\mathcal{G}}_v(e^{j\theta}) - \tilde{F}(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_{\mathbb{K}}$  is minimized at almost each  $\theta \in [-\pi, \pi]$ .*

To solve Problem  $\mathcal{P}_1$ , we first obtain a solution to Problem  $\mathcal{P}_3$  and then write the resulting optimal interpolator as a cascade of a stable downsampler and hold. This factorization of the optimal interpolator is straight-forward as we will see later.

*In the rest of the chapter, the transfer function  $\check{\mathcal{G}}(e^{j\theta})$  of a system  $\mathcal{G}$  is abbreviated as  $\check{\mathcal{G}}$  at given  $\theta$ . Unless necessary, we do this for all signals too.*

## 4.9 $L^2$ optimal downsampling

The  $L^2$  downsampling problem is the Problem  $\mathcal{P}_1$  with  $L^2$  system norm. As explained in Section 4.8, this problem can be solved by obtaining a solution of Problem  $\mathcal{P}_3$  with the Hilbert-Schmidt norm first and then writing the resulting optimal interpolator as a cascade of a *stable* downsampler and hold. Section 4.9.1 contains some basic results that are important in the solution of Problem  $\mathcal{P}_3$  with the Hilbert-Schmidt norm. The solutions of Problem  $\mathcal{P}_3$  and the  $L^2$  downsampling problem are presented in Section 4.9.2.

### 4.9.1 Preliminaries

This section describes some basic results which are useful in obtaining a solution of Problem  $\mathcal{P}_3$  with Hilbert-Schmidt norm i.e. obtaining

$$\hat{F}_{\text{opt}} := \arg \min_{\hat{F}} \|\check{G}_v - \hat{F} \hat{S}_y\|_{HS}$$

with constraint  $\text{rank } \hat{F}_{\text{opt}} \leq r$  at each  $\theta \in [-\pi, \pi]$ . Here the minimization is point-wise in  $\theta$ . At almost all  $\theta$ , the operator  $\check{G}_v : L^2[0, h) \rightarrow L^2[0, h)$  is a Hilbert-Schmidt operator as  $\mathcal{G}_v \in \mathcal{L}^2$  (see Lemma 2.4.5). Similarly, because of finite rank, the operator  $\hat{S}_y$  is a Hilbert-Schmidt operator at almost all  $\theta$  (see Lemma 4.8.2). Since  $\hat{F}_{\text{opt}}$  is bounded at almost all  $\theta$  by Lemma 2.4.4, Problem  $\mathcal{P}_3$  can be thought of as a special case of a generic problem of obtaining

$$F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|_{HS}$$

where  $A : \mathcal{H} \rightarrow \mathcal{H}_0$  and  $B : \mathcal{H} \rightarrow \mathcal{H}_1$  are Hilbert-Schmidt operators,  $F : \mathcal{H}_1 \rightarrow \mathcal{H}_0$  is a linear operator (not necessarily bounded) and  $\mathbb{F}_r$  denotes the set of all bounded linear operators mapping the space  $\mathcal{H}_1$  to  $\mathcal{H}_0$  and that have rank at most  $r$ . The spaces  $\mathcal{H}$ ,  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are separable Hilbert spaces so that these spaces have an orthonormal basis [47, theorem 3.52]. In our special case of Problem  $\mathcal{P}_3$ ,  $\mathcal{H} = \mathcal{H}_0 = L^2[0, h)$  and  $\mathcal{H}_1 = \mathbb{C}^M$ .

Also, let  $P : \mathcal{H} \rightarrow \mathcal{H}$  represent the orthogonal projection on  $(\text{Ker } B)^\perp$  and

$$A_N := A(I - P),$$

throughout this section. We also define the set for a linear operator  $G$

$$\mathbb{N}_G := \begin{cases} \{0, 1, \dots, \text{rank } G - 1\} & \text{if rank } G \text{ is finite} \\ \mathbb{N} & \text{if rank } G \text{ is infinite} \end{cases}. \quad (4.18)$$

Since the image of a Hilbert-Schmidt (hence compact) operator is not necessarily closed,  $\text{Im } B$  is not necessarily closed [63, theorem 2.14]. An example of such a case is discussed in [31, theorem 5.1]. However, in most of the downsampling problems later,  $B$  is a finite dimensional operator, therefore  $\text{Im } B$  is closed. For generality, in this preliminaries section we assume that  $\text{Im } B$  is not necessarily closed.

An algorithm for obtaining  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|_{HS}$  is presented at the end of this section.

Since  $(\text{Ker } AP)^\perp \subseteq (\text{Ker } B)^\perp$ , the orthogonal projection  $P$  helps us in identifying the space where  $F$  could play a role. Using  $B(I - P) = 0$ , we have the following result.

**Lemma 4.9.1.**

$$\|A - FB\|_{HS}^2 = \|A_N\|_{HS}^2 + \|AP - FB\|_{HS}^2. \quad (4.19)$$

where  $A_N := A(I - P)$ .

*Proof.* See Appendix 4.B (page 104).  $\square$

From (4.19) it is clear that an  $F \in \mathbb{F}_r$  that minimizes  $\|A - FB\|_{HS}$  also minimizes  $\|AP - FB\|_{HS}$ . If a rank- $r$   $F$  is such that  $FB$  cancels out the dominant  $r$  singular values in an SVD of  $AP$ , then this  $F$  is optimal. This fact is used in the following theorem to obtain  $F_{\text{opt}}$ .

**Theorem 4.9.2.** *Suppose an SVD of  $AP$  is given by*

$$APx = \sum_{i \in \mathbb{N}_{AP}} \sigma_i \langle x, e_i \rangle f_i, \quad x \in \mathcal{H} \quad (4.20)$$

where  $\{e_i\}$  and  $\{f_i\}$  are orthonormal sequences in  $\mathcal{H}$  and  $\mathcal{H}_0$  respectively, and  $\{\sigma_i\}$  is a real non-increasing non-zero sequence.

Define a space  $\mathbb{W} := \overline{\text{span}}\{e_0, e_1, \dots, e_{r-1}\}$ . Now, for a given  $r \leq \text{rank}(AP)$

$$F_{\text{opt}}x := \sum_{i=0}^{r-1} \sigma_i \langle B^+x, e_i \rangle f_i, \quad x \in \mathcal{H}_1 \quad (4.21a)$$

$$= \sum_{i=0}^{r-1} \sigma_i \langle QB^+x, e_i \rangle f_i \quad (4.21b)$$

is bounded iff  $\mathbb{W} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$ . Here  $B^+ : \mathcal{H}_1 \rightarrow \mathcal{H}$  is the pseudo-inverse of the operator  $B$  [13, Definition 2.2] and  $Q : \mathcal{H} \rightarrow \mathbb{W}$  is the orthogonal projection on the space  $\mathbb{W}$ . Now, if the  $F_{\text{opt}}$  is bounded then it minimizes  $\|AP - FB\|_{HS}$  over all  $F \in \mathbb{F}_r$ .

*Proof.* See Appendix 4.B (page 104).  $\square$

The uniqueness of  $F_{\text{opt}}$  in the above theorem depends upon the singular values of  $AP$ . For example, if  $r = 1$  and all singular values of  $AP$  are same then we have multiple solutions.

**Remark 4.9.3.** *The dependence of the space  $\mathbb{W}$  on  $r$  forced us to write the Theorem 4.9.2 for  $r \leq \text{rank } AP$  only. Because if  $r > \text{rank}(AP)$  then  $\mathbb{W}$  is the space  $\overline{\text{span}}\{e_0, e_1, \dots, e_{\text{rank } AP-1}\}$  for all  $r > \text{rank}(AP)$  and  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|_{HS}$  is given by (4.21a) with  $r = \text{rank } AP$  as  $AP - F_{\text{opt}}B = 0$ . Taking  $\sigma_i = 0$  for all  $i \geq \text{rank}(AP)$ , we can write  $F_{\text{opt}}$ , for  $r$  less than or equal to minimum of  $\dim \mathcal{H}$ ,  $\dim \mathcal{H}_0$  and  $\dim \mathcal{H}_1$  (denoted by  $\min(\dim \mathcal{H}, \dim \mathcal{H}_0, \dim \mathcal{H}_1)$ ) as*

$$F_{\text{opt}}x = \sum_{i=0}^{r-1} \sigma_i \langle B^+x, e_i \rangle f_i, \quad x \in \mathcal{H}_1. \quad (4.22)$$

Note that  $\text{rank}(AP) \leq \min(\dim \mathcal{H}, \dim \mathcal{H}_0, \dim \mathcal{H}_1)$ .

Note that, if  $\mathbb{W} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$ , then  $B^+$  may still be unbounded in (4.21a) whereas the operator  $QB^+$  is always bounded.

The condition that  $\mathbb{W} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  in Theorem 4.9.2 is met if the  $e_i$ 's (see (4.20)) belong to the set of right singular vectors of  $B$ . This is proved in following corollary.

**Corollary 4.9.4.** *Suppose an SVD of  $AP$  is given by (4.20) and that operator  $B : \mathcal{H} \rightarrow \mathcal{H}_1$  is defined by*

$$Bx = \sum_{i \in \mathbb{N}_B} \beta_i \langle x, g_i \rangle h_i \quad x \in \mathcal{H} \quad (4.23)$$

where  $\{g_i\}_{i \in \mathbb{N}_B}$  and  $\{h_i\}_{i \in \mathbb{N}_B}$  are orthonormal sequences in  $\mathcal{H}$  and  $\mathcal{H}_1$  respectively, and  $\beta_i \neq 0$ . If  $e_i = g_{n(i)}$  for some  $n(i)$  — in other words  $\{e_i\} \subset \{g_k\}$  — then

$$F_{\text{opt}}x := \sum_{i=0}^{r-1} \frac{\sigma_i}{\beta_{n(i)}} \langle x, h_{n(i)} \rangle f_i \quad x \in \mathcal{H}_1. \quad (4.24)$$

minimizes  $\|AP - FB\|_{HS}$  over all  $F \in \mathbb{F}_r$ .

*Proof.* See Appendix 4.B (page 106). □

Having obtained the  $F_{\text{opt}}$ , it is natural to seek for the minimal error norm  $\|A - F_{\text{opt}}B\|_{HS}$ .

**Corollary 4.9.5.** *Let  $F_{\text{opt}}$  as in Theorem 4.9.2. Then,*

$$\|A - F_{\text{opt}}B\|_{HS}^2 = \|A\|_{HS}^2 - \|F_{\text{opt}}B\|_{HS}^2.$$

*Proof.* See Appendix 4.B (page 106). □

This may be surprising because  $\mathbb{F}_r$  is not a subspace of the space of bounded linear operators. Yet we still have this Pythagoras type result. Now we can summarize the steps for obtaining  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|_{HS}$  in the form of Algorithm 1.

## 4.9.2 $L^2$ optimal downsampling solution

Recall that  $S_y := S_h \mathcal{G}_y$ . The Problem  $\mathcal{P}_3$  with Hilbert-Schmidt norm can be solved by application of Algorithm 1. Among the important parameters to find in application of Algorithm 1 are the orthogonal projection  $\check{P}$  on the space  $(\text{Ker } \check{S}_y)^\perp$ , and an SVD of the operator  $\check{G}_v \check{P}$  at almost every  $\theta \in [-\pi, \pi]$ . Having obtained an SVD of the operator  $\check{S}_y$  in Lemma 4.8.2, we can immediately write down an orthonormal basis  $\{\check{p}_k\}_{k \in \mathbb{M}, \alpha_k \neq 0}$  of the space  $(\text{Ker } \check{S}_y)^\perp$ , hence  $\check{P}$  at almost every  $\theta \in [-\pi, \pi]$  as

$$\check{P}x = \sum_{k \in \mathbb{M}, \alpha_k \neq 0} \langle x, \check{p}_k \rangle \check{p}_k, \quad x \in L^2[0, h). \quad (4.25)$$



---

**Algorithm 1** For  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|_{HS}$

---

**Require:**  $A$  and  $B$  are Hilbert-Schmidt operators

Obtain the orthogonal projection  $P$  on  $(\text{Ker } B)^\perp$

**if**  $AP = 0$  **then**

$F_{\text{opt}} = 0$

**else**

Obtain SVD of  $AP$  and  $e_i$  defined in (4.20).

**if**  $e_i \in \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  **then**

Obtain  $B^+$

$F_{\text{opt}}$  is given by (4.21a).

**else**

$F_{\text{opt}} \in \mathbb{F}_r$  does not exist.

**return**

**end if**

**end if**

Calculate  $\|A - F_{\text{opt}}B\|_{HS}$  using Corollary 4.9.5

**return**  $F_{\text{opt}}, \|A - F_{\text{opt}}B\|_{HS}$

---

The next step in Algorithm 1 is to obtain an SVD of the operator  $\check{G}_v \check{P}$  at almost every  $\theta \in [-\pi, \pi]$ .

**Lemma 4.9.6.** *Given LCTI system  $G_v \in L^\infty \cap L^2$ . Using all the notations and conditions of Lemma 4.8.2 and (4.25), an SVD (modulo ordering) of the operator  $\check{G}_v \check{P} : L^2[0, h) \rightarrow L^2[0, h)$  at almost all  $\theta \in [-\pi, \pi]$  is given by*

$$\check{G}_v \check{P} \check{w} = \sum_{n \in \mathbb{M}, \alpha_n \neq 0} \sigma_n \langle \check{w}, \check{p}_n \rangle \check{q}_n \quad (4.26)$$

where

$$\sigma_n = \frac{\sqrt{\sum_{i \in \mathbb{Z}} |G_v(j\omega_{n+Mi}) \psi_y(j\omega_{n+Mi})|^2}}{\alpha_n},$$

$$\check{q}_n(\tau) = \frac{1}{\sigma_n \alpha_n} \sum_{i \in \mathbb{Z}} G_v(j\omega_{n+Mi}) \psi_y^*(j\omega_{n+Mi}) \check{e}_{n+Mi}(e^{j\theta}; \tau)$$

and  $\tau \in [0, h)$ .

*Proof.* See Appendix 4.B (page 107). □

Clearly  $\text{rank}(\check{G}_v \check{P}) \leq \text{rank}(\check{P})$  at a given  $\theta$ . Since  $\check{S}_y$  and  $\check{G}_v \check{P}$  share right singular vectors, we can invoke Corollary 4.9.4. Therefore, according to the Algorithm 1, we have all the basic tools to obtain a solution of Problem  $\mathcal{P}_3$ . For properly writing the results we need the definition of *Dominant index set*:

**Definition 4.9.7.** The dominant index set is defined as an ordered set of the indices of the dominant  $r$  elements of a bounded sequence  $\mathfrak{S} := \{\sigma_0, \sigma_1, \dots\}$ . It is denoted by  $D_r(\mathfrak{S})$ .

For example if  $\mathfrak{S} := \{10, 2, 9, 4, 1\}$  then  $D_3(\mathfrak{S}) = \{0, 2, 3\}$  and  $D_4(\mathfrak{S}) = \{0, 2, 3, 1\}$ . Note that  $r$  in  $D_r(\mathfrak{S})$  cannot be greater than cardinality of the  $\mathfrak{S}$ .

The following theorem, which is one of the main results of this chapter, describes a solution of Problem  $\mathcal{P}_3$  (with the Hilbert-Schmidt norm).

**Theorem 4.9.8.** Given LCTI system  $\mathcal{G}_v \in L^\infty \cap L^2$ . Let  $\mathfrak{S}$  be the set of finite singular values of the operator  $\check{G}_v \check{P}$  at almost every  $\theta \in [-\pi, \pi]$  i.e.  $\mathfrak{S} := \{\sigma_k\}_{k \in \mathbb{M}, \alpha_k \neq 0}$ . Given  $r \leq M$ , define  $\check{F}_{\text{opt}}$  as

$$\check{F}_{\text{opt}} \vec{y}_{h'} := \sum_{k \in D_r(\mathfrak{S})} \Gamma_k \left\langle \vec{y}_{h'}, \frac{1}{\sqrt{M}} \vec{e}_k \right\rangle, \quad (4.27)$$

where

$$\Gamma_k := \frac{\sum_{i \in \mathbb{Z}} G_v(j\omega_{k+Mi}) \psi_y^*(j\omega_{k+Mi}) e^{j(\omega_{k+Mi})\tau}}{\sum_{i \in \mathbb{Z}} |\psi_y(j\omega_{k+Mi})|^2}, \quad (4.28)$$

whenever  $\theta \in \mathfrak{A} := \{\theta \in [-\pi, \pi] : \check{G}_y(e^{j\theta}) \neq 0\}$  and  $\tau \in [0, h)$ . For  $\theta \in [-\pi, \pi] \setminus \mathfrak{A}$ , we can take  $\check{F}_{\text{opt}} = 0$ . If this interpolator  $\mathcal{F}_{\text{opt}}$  is well-defined and stable, then it minimizes  $\|\mathcal{G}_v - \mathcal{F}\mathcal{S}_y\|_{L^2}$  over all interpolators  $\mathcal{F}$  with  $\text{rank } \check{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ .

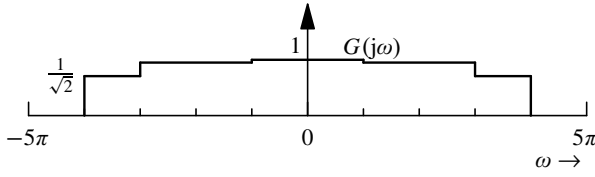
*Proof.* Note that since  $\mathcal{G}_v \in L^2$ , the  $\check{G}_v$  is a Hilbert-Schmidt operator for almost all  $\theta$ . Also, because of finite rank, the operator  $\check{S}_y$  is a Hilbert-Schmidt operator for almost all  $\theta$  and its image is closed. Now, the result follows from Corollary 4.9.4 and Remark 4.9.3.  $\square$

Note that the set  $D_r(\mathfrak{S})$  can change with  $\theta \in [-\pi, \pi]$ . Theorem 4.9.8 solves the Problem  $\mathcal{P}_3$  with the Hilbert-Schmidt norm and gives the kernel of the interpolator  $\mathcal{F}_{\text{opt}}$  in the lifted frequency domain. However, to solve the  $L^2$  downsampling problem  $\mathcal{P}_1$ , we have to also write the optimal interpolator  $\mathcal{F}_{\text{opt}}$  as a cascade of a stable downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$ . To this end, (4.15) is very useful as shown in the following result.

**Theorem 4.9.9.** Let  $\mathcal{F}_{\text{opt}}$ ,  $\Gamma_k$ ,  $\mathfrak{S}$  be as in Theorem 4.9.8. Define the lifted downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$  such that their transfer functions have kernel

$$\vec{\chi}_{\text{opt}}(e^{j\theta}; -m) := \frac{1}{M} \begin{bmatrix} e^{-j\omega_{k_0} m h'} \\ \vdots \\ e^{-j\omega_{k_{r-1}} m h'} \end{bmatrix}, \quad (4.29a)$$

$$\check{\phi}_{\text{opt}}(e^{j\theta}; \tau) := [\Gamma_{k_0}, \dots, \Gamma_{k_{r-1}}] \quad (4.29b)$$



**Figure 4.9:**  $G(j\omega)$  of Example 4.9.11

at each  $\theta \in [-\pi, \pi]$  respectively. Here  $\{k_0, k_1, \dots, k_{r-1}\} := D_r(\mathfrak{S})$ . If the downsampler  $\tilde{\mathcal{S}}_{h,\text{opt}}$  and hold  $\mathcal{H}_{\text{opt}}$  are well-defined and stable, then they solve the  $L^2$  downsampling problem.

*Proof.* Note that the inner product equals

$$\left\langle \check{y}_{h'}(e^{j\theta}), \vec{e}_k(e^{j\theta}) \right\rangle_{\mathbb{C}^M} = \sum_{m \in \mathbb{M}} \check{y}_{h'}(e^{j\theta}; m) e^{-j\frac{\theta+2\pi k}{h}mh'}.$$

The rest of the proof follows from (4.15).  $\square$

Note that the  $k_i$  in Theorem 4.9.9 can change with  $\theta \in [-\pi, \pi]$ . Also note that if the optimal downsampler  $\tilde{\mathcal{S}}_{h,\text{opt}}$  and hold  $\mathcal{H}_{\text{opt}}$  are well-defined then interpolator  $\tilde{\mathcal{F}}_{\text{opt}}$  given by (4.27) is a hybrid interpolator by definition.

**Remark 4.9.10.** In general, for any bistable invertible mapping  $J$ , the  $\tilde{\mathcal{H}}_{\text{opt}}J$  and  $J^{-1}\tilde{\mathcal{S}}_{h,\text{opt}}$  are also optimal hold and downsampler.

For any  $k \in \mathbb{M}$ , we call the set of frequencies

$$\{\theta/h + 2\omega_N(k + Mi)\}_{i \in \mathbb{N}}$$

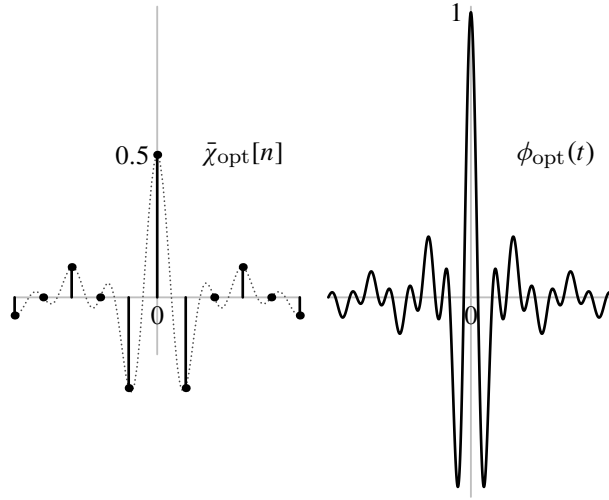
the  $h'$ -aliased frequencies of the frequency  $\theta/h + 2\omega_N k$ . Theorem 4.9.8 clearly states that these  $h'$ -aliased frequencies play a crucial role in deciding the optimal downsampler (see the construction of  $\mathfrak{S}$  in Theorem 4.9.8) for any finite  $M$ . This fact is illustrated with an example given below.

**Example 4.9.11.** Let  $\mathcal{G}_v$  and  $\mathcal{G}_y$  be LCTI systems with classic frequency response  $G_v(j\omega)$  and  $G_y(j\omega)$  respectively. Also, let  $G_v(j\omega) = G_y(j\omega) = G(j\omega)$  where

$$G(j\omega) = \begin{cases} 1 & \omega \in [-\pi, \pi] \\ 0.95 & \omega \in [-3\pi, -\pi] \cup [1\pi, 3\pi] \\ \frac{1}{\sqrt{2}} & \omega \in [-4\pi, -3\pi] \cup [3\pi, 4\pi] \\ 0 & \text{elsewhere} \end{cases}.$$

$G(j\omega)$  is shown in Figure 4.9. Also, suppose that the sampler  $\mathcal{S}_{h'}$  has sampling function  $\psi(t)$  whose classic frequency response is given by

$$\psi(j\omega) = \begin{cases} 1 & \omega \in [-4\pi, 4\pi] \\ 0 & \text{elsewhere} \end{cases}$$



**Figure 4.10:** Discrete sampling function  $\bar{\chi}_{opt}[n]$  of the  $L^2$  optimal downsampler and the Hold function  $\phi_{opt}(t)$  of  $L^2$  optimal hold (right) in Example 4.9.11. The left figure also shows  $\frac{1}{2} \cos(\pi t) \operatorname{sinc}(\frac{1}{2})$  (dotted).

and  $M = 2$ ,  $h = 1$ . The Nyquist frequency is  $\omega_N = \pi/h = \pi$ .

Now, the input of the sampler  $\mathcal{S}_{h'}$  is bandlimited to  $\omega_B := 4\pi$  rad/sec as  $G(j\omega)$  is bandlimited to  $4\pi$  rad/sec. Since the sampling interval of  $\mathcal{S}_{h'}$  i.e.  $h' (= \frac{1}{2})$  is greater than the Nyquist interval (defined as  $\frac{\pi}{\omega_B}$ ) required for its input i.e.  $\frac{1}{4}$ , there are  $h'$ -aliased frequencies at the output of sampler  $\mathcal{S}_{h'}$  (see Figure 4.3) [27, §5.1]. Calculating the singular values  $\sigma_i$  (see Lemma 4.9.6) in the presence of  $h'$ -aliased frequencies, we find that

$$\sigma_0 = \sqrt{5/6} \approx 0.913, \quad \sigma_1 = 0.95 \quad \forall \theta \in [-\pi, \pi]$$

Note that  $\sigma_1 > \sigma_0$  for all  $\theta \in [-\pi, \pi]$ . Now, it follows from Theorem 4.9.8 that if  $r = 1$  then the  $L^2$  optimal interpolator in lifted frequency domain is given by

$$\hat{F}_{opt}x = \frac{e^{j\omega_1\tau} + e^{j\omega_{-1}\tau}}{2} \left\langle x, \frac{1}{2} e^{j\omega_1 m h'} \right\rangle \quad m \in \{0, 1\}$$

at almost all  $\theta \in [-\pi, \pi]$ . Using (4.29) and the inverse lifted transform, we can write the discrete sampling function  $\bar{\chi}_{opt}[n]$  and hold function  $\phi_{opt}(t)$  of the optimal downsampler and hold as (see also Figure 4.10)

$$\begin{aligned} \bar{\chi}_{opt}[n] &= \frac{1}{2} \operatorname{sinc}\left(\frac{n}{2}\right) (-1)^n \\ \phi_{opt}(t) &= \frac{1}{2} \left( e^{j2\omega_N t} \operatorname{sinc}(t) + e^{-j2\omega_N t} \operatorname{sinc}(t) \right) = \cos(2\omega_N t) \operatorname{sinc}(t). \end{aligned}$$

Since the optimal downsampler and hold are well-defined,  $\mathcal{F}_{\text{opt}}$  is an order-1 hybrid interpolator. Also, since  $\text{sinc}[n] \in \ell^2$ ,  $\text{sinc}(t) \in L^2$  and  $|\cos(t)|$  is bounded, it follows from lemma's 2.4.7 and 4.6.2 that the optimal downsampler and hold are stable (i.e. in  $L^\infty$ ).

### 4.9.3 $L^2$ Error Norm

To quantify the reconstruction error (in the optimal case i.e.  $\mathcal{F} = \mathcal{F}_{\text{opt}}$ ), we need the error system norm  $\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}$ . The following theorem states how to obtain it.

**Theorem 4.9.12.** *Let  $\mathcal{F}_{\text{opt}}$ , set  $\mathfrak{S}$  and set  $\mathfrak{A}$  be as in Theorem 4.9.8. The  $L^2$  system norm of the operator  $\mathcal{F}_{\text{opt}}\mathcal{G}_y$  is given by*

$$\|\mathcal{F}_{\text{opt}}\mathcal{G}_y\|_{L^2}^2 = \frac{1}{2\pi h} \int_{\mathfrak{A}} \sum_{k \in D_r(\mathfrak{S})} \sigma_k^2 d\theta \quad (4.30)$$

where  $\sigma_k$ 's are defined in Lemma 4.9.6. Now the  $L^2$  system norm of the system  $\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{G}_y$  is given by

$$\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{G}_y\|_{L^2}^2 = \|\mathcal{G}_v\|_{L^2}^2 - \|\mathcal{F}_{\text{opt}}\mathcal{G}_y\|_{L^2}^2 \quad (4.31)$$

where  $\mathcal{F}_{\text{opt}}$  is defined in Theorem 4.9.8.

*Proof.* For  $\theta \in \mathfrak{A}$ ,  $\|\hat{F}_{\text{opt}}\hat{S}_y\|_{HS}^2 = \sum_{k \in D_r(\mathfrak{S})} \sigma_k^2$  as the Hilbert-Schmidt norm of the operator  $\hat{F}_{\text{opt}}\hat{S}_y$  is given by the sum of squares of singular values. For  $\theta \in \mathfrak{A}^c$ ,  $\hat{F}_{\text{opt}}\hat{S}_y = 0$  for any bounded  $\hat{F}_{\text{opt}}$ , which implies  $\|\hat{F}_{\text{opt}}\hat{S}_y\|_{HS} = 0$  at those  $\theta$ . Now, (4.30) and (4.31) follow from (2.27) and Corollary 4.9.5.  $\square$

**Example 4.9.13.** *For the system of Example 4.9.11, we find that*

$$\|\mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} 0.95^2 d\theta$$

and

$$\begin{aligned} \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}^2 &= \|\mathcal{G}_v\|_{L^2}^2 - \|\mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}^2 \\ &= \frac{1}{2\pi h} \int_{-\pi}^{\pi} \left(1 + \frac{1}{2} + 2 \times 0.95^2 - 0.95^2\right) d\theta = 2.403. \end{aligned}$$

Therefore, the reconstructed power ( $\|\mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}^2 / \|\check{\mathcal{G}}_v\|_{L^2}^2$ ) is 27.29%. If we do full order downsampling (in this case order  $M = 2$ ) then the  $\hat{F}_{\text{opt}}$  will be of rank-2 and

$$\|\mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} 0.95^2 + 5/6 d\theta = 1.736$$

Then, the reconstructed power is 52.53%. This is the theoretical limit, i.e. the maximum that can be obtained by any downsampler and hold of any order, given  $\mathcal{G}_v$  and  $\mathcal{S}_y$ .

## 4.10 $L^\infty$ optimal downsampling

The  $L^\infty$  downsampling problem is the Problem  $\mathcal{P}_1$  with  $L^\infty$  system norm. As explained in Section 4.8, this problem can be solved by obtaining a solution of Problem  $\mathcal{P}_3$  with the induced 2-norm first and then writing the resulting optimal interpolator as a cascade of a *stable* downsampler and hold. Section 4.10.1 contains some basic results that are important in the solution of Problem  $\mathcal{P}_3$  with the induced 2-norm. The solutions of Problem  $\mathcal{P}_3$  and the  $L^\infty$  downsampling problem are presented in Section 4.10.2.

### 4.10.1 Preliminaries

This section describes some basic results which are useful in obtaining a solution of Problem  $\mathcal{P}_3$  with induced 2-norm i.e. obtaining

$$\check{F}_{\text{opt}} := \arg \min_{\check{F}} \|\check{G}_v - \check{F}\check{S}_y\|$$

with constraint  $\text{rank } \check{F}_{\text{opt}} \leq r$  at each  $\theta \in [-\pi, \pi]$ . Here the minimization is point-wise in  $\theta$ . At almost all  $\theta$ , the operator  $\check{G}_v : L^2[0, h) \rightarrow L^2[0, h)$  is a compact operator as  $\mathcal{G}_v \in L^2$  (see Lemma 2.4.5). Similarly, because of finite rank, the operator  $\check{S}_y$  is a compact operator at almost all given  $\theta$  (see Lemma 4.8.2). Since  $\check{F}_{\text{opt}}$  is bounded at almost all  $\theta$  by Lemma 2.4.4, Problem  $\mathcal{P}_3$  can be thought as a special case of a generic problem of obtaining

$$F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|$$

where  $A : \mathcal{H} \rightarrow \mathcal{H}_0$  and  $B : \mathcal{H} \rightarrow \mathcal{H}_1$  are compact operators,  $F : \mathcal{H}_1 \rightarrow \mathcal{H}_0$  is a linear operator (not necessarily bounded) and  $\mathbb{F}_r$  denotes the set of all bounded linear operators mapping the space  $\mathcal{H}_1$  to  $\mathcal{H}_0$  and that have rank at most  $r$ . The spaces  $\mathcal{H}$ ,  $\mathcal{H}_0$  and  $\mathcal{H}_1$  are separable Hilbert spaces so that these spaces have an orthonormal basis [47, theorem 3.52]. Also,  $P : \mathcal{H} \rightarrow \mathcal{H}$  represents the orthogonal projection on  $(\text{Ker } B)^\perp$ , and we define  $A_N$  and  $T_\gamma$  as

$$\begin{aligned} A_N &:= A(I - P), \\ T_\gamma &:= I - \gamma^{-2} A_N A_N^* \quad \forall \gamma \in (\|A_N\|, \infty), \end{aligned}$$

where  $\|\cdot\|$  represents the induced 2-norm of the operator throughout this section.  $\mathbb{N}_G$  for an operator  $G$  has the same meaning as in Section 4.9.1 (see (4.18)). Similar to Section 4.9.1, we assume that  $\text{Im } B$  is not necessarily closed in this section.

An algorithm for obtaining  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|$  is presented at the end of this section.

Similar to (4.19), the orthogonal projection  $P$  provides some clue about the lower bound of  $\|A - FB\|$  for all bounded  $F$ . Indeed,

**Lemma 4.10.1.**  $\inf_F \|A - FB\| \geq \|A_N\|.$

*Proof.* Since  $B(I - P) = 0$ , we have  $A_N = (A - FB)(I - P)$ . Therefore,

$$\|A_N\| \leq \|A - FB\| \|I - P\| = \|A - FB\|.$$

□

The following is a standard but important step in  $L^\infty$  optimization [71, 31]:

**Lemma 4.10.2.** *If  $\gamma > \|A_N\|$  then*

$$\|A - FB\| \leq \gamma \Leftrightarrow \|T_\gamma^{-\frac{1}{2}}(AP - FB)\| \leq \gamma$$

*Proof.* See Appendix 4.C (page 108). □

It is also clear from Lemma 4.10.2 that if  $\gamma > \|A_N\|$  then the singular values of  $T_\gamma^{-\frac{1}{2}}AP$  (or singular values of  $(AP)^*T_\gamma^{-1}AP$ ) provide a clue about optimal  $F_{\text{opt}}$  at the given  $\gamma$ .

We start with an SVD of  $AA^*$  and other simplifications that are later useful in obtaining singular values of  $(AP)^*T_\gamma^{-1}AP$  for a given  $\gamma$ .

**Lemma 4.10.3.** *Let an SVD of  $AA^* : \mathcal{H}_0 \rightarrow \mathcal{H}_0$  be given by*

$$AA^*x = \sum_{k \in \mathbb{N}_A} \alpha_k^2 \langle x, v_k \rangle v_k,$$

where  $\{\alpha_k\}$  is a non-increasing non-zero sequence. Define for a given  $f_i, i \in \mathbb{N}_{AP}$  (see (4.20)) the subspace

$$\mathcal{V}_i := \overline{\text{span}}\{v_k | k \in \mathbb{N}_A \& \langle v_k, f_i \rangle \neq 0\} \quad (4.32)$$

and the orthogonal projection  $P_i$  onto the subspace  $\mathcal{V}_i$ . Then  $f_i \in \mathcal{V}_i$  and an SVD (modulo ordering) of  $P_i AA^* P_i : \mathcal{H}_0 \rightarrow \mathcal{V}_i$  exists and it has the form

$$P_i AA^* P_i x = \sum_{\substack{k \in \mathbb{N}_A \\ \langle v_k, f_i \rangle \neq 0}} \alpha_k^2 \langle x, v_k \rangle v_k, \quad x \in \mathcal{H}_0. \quad (4.33)$$

*Proof.* Since  $A$  is assumed to be a compact operator, its SVD exists [69]. Hence SVD of  $AA^*$  exists. As  $\text{Im } A = \text{Im } AA^* = \overline{\text{span}}\{v_1, v_2, \dots\}$  and  $f_i \in \text{Im } AP \subseteq \text{Im } A$ , therefore we have for each  $i$  that  $f_i \in \mathcal{V}_i$ . Now, (4.33) follows from  $P_i x = \sum_{\substack{k \in \mathbb{N}_A \\ \langle v_k, f_i \rangle \neq 0}} \langle x, v_k \rangle v_k$ . □

In Lemma 4.10.3, there is some degree of freedom in selection of the singular vectors  $v_k$  of  $AA^*$ . However, once the singular vectors  $v_k$  of  $AA^*$  is fixed then the space is  $\mathcal{V}_i$  is fixed.

The following corollary is immediate.

**Corollary 4.10.4.** *The spaces  $\mathcal{V}_i$  (defined in Lemma 4.10.3) are invariant under the operator  $AA^*$ .*

*Proof.* The proof follows from the fact that the  $v_k$  are eigenvectors of  $AA^*$ .  $\square$

Without loss of generality an SVD of  $P_i AA^* P_i$  can be written by rearranging (4.33) as

$$P_i AA^* P_i x = \sum_{n \in \mathbb{N}_{P_i A}} \alpha_{in}^2 \langle x, v_{in} \rangle v_{in}, \quad x \in \mathcal{H}_0 \quad (4.34)$$

where  $\{\alpha_{in}^2\}_{n \in \mathbb{N}}$  are in descending order,  $\alpha_{in}^2$  are non-zero singular values of  $AA^*$  and  $v_{in}$  are singular vectors of  $AA^*$  such that  $\langle v_{in}, f_i \rangle \neq 0$ .

In general, obtaining singular values of  $(AP)^* T_\gamma^{-1} AP$  for a given  $\gamma$  can be very cumbersome but in some special cases it is relatively easy. One of the special case happens if the following assumption is satisfied.

**Assumption  $\mathcal{A}_1$**  : *The subspaces  $\{\mathcal{V}_i\}_{i \in \mathbb{N}_{AP}}$  defined in (4.32) are mutually orthogonal.*

The above is a strong assumption. However in case of the downsampling this is true as we will see later (this assumption is also satisfied in [31, theorem 5.1]).

Since  $f_i \in \mathcal{V}_i$  (see Lemma 4.10.3), therefore if Assumption  $\mathcal{A}_1$  is satisfied then  $f_i \perp \mathcal{V}_j$ ,  $j \neq i$ . Hence, the orthogonal projection  $P_i$  on the space  $\mathcal{V}_i$  satisfies  $P_i f_j = \delta[i - j] f_j$ .

Under Assumption  $\mathcal{A}_1$ , an SVD of the operator  $(AP)^* T_\gamma^{-\frac{1}{2}} AP$  can be obtained, which eventually helps in obtaining  $F_{\text{opt}}$ , as follows:

**Lemma 4.10.5.** *Suppose an SVD of  $AP$  exists and is given by (4.20). Assume that for each  $i$ , an SVD of  $P_i AA^* P_i : \mathcal{H}_0 \rightarrow \mathcal{V}_i$  exists and has the form given in (4.34). If  $\gamma > \|A_N\|$  and Assumption  $\mathcal{A}_1$  is satisfied for the spaces  $\{\mathcal{V}_i\}_{i \in \mathbb{N}_{AP}}$ , then an SVD of  $(AP)^* T_\gamma^{-1} AP$  exists and it is given by (modulo ordering)*

$$(AP)^* T_\gamma^{-1} AP x = \sum_{i \in \mathbb{N}_{AP}} \eta_i^2(\gamma) \langle x, e_i \rangle e_i$$

where

$$\eta_i^2(\gamma) = \left( \gamma^{-2} + \left( \sigma_i^2 \sum_k \frac{|\langle f_i, v_{ik} \rangle|^2}{1 - \gamma^{-2} \alpha_{ik}^2} \right)^{-1} \right)^{-1}, \quad (4.35)$$

and  $e_i$  and  $f_i$  are as defined in (4.20).

*Proof.* See Appendix 4.C (page 108).  $\square$

If Assumption  $\mathcal{A}_1$  is satisfied for the spaces  $\{\mathcal{V}_i\}_{i \in \mathbb{N}_{AP}}$ , then the above Lemma also says that if the right hand side of (4.35) is negative for some value of  $\gamma$  then that  $\gamma$  must be less than or equal to  $\|A_N\|$ . Also, the singular values of  $T_\gamma^{-\frac{1}{2}} AP$  are  $\eta_i(\gamma)$  and the right singular vectors are  $e_i$  (for all  $i$ ).



Even after obtaining an SVD of  $(AP)^* T_\gamma^{-\frac{1}{2}} AP$ , our original problem of obtaining  $\arg \min_{F \in \mathbb{F}_r} \|A - FB\|$  is still two-fold as both  $\min_{F \in \mathbb{F}_r} \|A - FB\|$  and  $F_{\text{opt}}$  are still unknown. However, if we restrict  $F$  to a particular class then we can obtain both.

**Lemma 4.10.6.** *Suppose that an SVD of  $AP$  is given by (4.20). Given  $r \leq \text{rank } AP$ , define the set  $\mathfrak{F}$  that consists of all rank- $r$   $F$  given by*

$$Fx = \sum_{i \in \mathfrak{C}} \sigma_i \langle B^+ x, e_i \rangle f_i, \quad x \in \mathcal{H}_1 \quad (4.36)$$

where  $\mathfrak{C}$  is any set of different  $r$  non-negative integers in the set  $\mathbb{N}_{AP}$  and  $e_i$  and  $f_i$  are as defined in (4.20). For convenience, we define  $F_0 \in \mathfrak{F}$  as that  $F$  which has  $\mathfrak{C} = D_r(\mathfrak{S})$  (see Definition 4.9.7) where  $\mathfrak{S} := \{\alpha_{i0}\}_{i \in \mathbb{N}_{AP}}$ . If Assumption  $\mathcal{A}_1$  is satisfied for the spaces  $\{\mathcal{V}_i\}_{i \in \mathbb{N}_{AP}}$ , then  $F_0 = \arg \min_{F \in \mathfrak{F}} \|A - FB\|$  and

$$\|A - F_0 B\| = \max_{\substack{i \notin D_r(\mathfrak{S}) \\ i \in \mathbb{N}}} \{\alpha_{i0}, \|A_N\|\}. \quad (4.37)$$

*Proof.* See Appendix 4.C (page 110). □

Note that  $\mathfrak{F}$  might contains unbounded operators, therefore  $\mathfrak{F} \not\subseteq \mathbb{F}_r$ .

In the following theorem we show that  $F_0$  defined in Lemma 4.10.6 not only minimizes  $\|A - FB\|$  over the set  $\mathfrak{F}$  but also over  $\mathbb{F}_r$  under certain conditions.

**Theorem 4.10.7.** *For a given  $r \leq \text{rank } AP$ , let the set  $D_r(\mathfrak{S})$  be as in Lemma 4.10.6. Also suppose that an SVD of  $AP$  has the form given in (4.20). If space  $\overline{\text{span}}\{e_i\}_{i \in D_r(\mathfrak{S})} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  and Assumption  $\mathcal{A}_1$  is satisfied for the spaces  $\{\mathcal{V}_i\}_{i \in \mathbb{N}}$  then a rank- $r$  bounded  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|$  exists and can be chosen as  $F_0$  (see Lemma 4.10.6). Also,  $\|A - F_{\text{opt}} B\| = \|A - F_0 B\|$  (see (4.37)).*

*Proof.* See Appendix 4.C (page 111). □

**Remark 4.10.8.** *Theorem 4.10.7 is for  $r \leq \text{rank } AP$  only. However, if  $r > \text{rank}(AP)$  then  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|_{HS}$  is given by (4.21a) with  $r = \text{rank } AP$  as  $AP - F_{\text{opt}} B = 0$  and  $\|A - F_0 B\| = \|A_N\|$  in this case. Taking  $\sigma_i = 0$  for all  $i \geq \text{rank}(AP)$ , we can write  $F_{\text{opt}}$ , for  $r$  less than or equal to  $\dim \mathcal{H}_0$  or  $\dim \mathcal{H}_1$  whichever is smaller, as (4.22).*

In Theorem 4.10.7, we need  $\overline{\text{span}}\{e_i\}_{i \in D_r(\mathfrak{S})} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  for the  $F_0$ , defined in Lemma 4.10.6, to be bounded. However, if  $F_0$  is unbounded then a bounded  $F_{\text{opt}} \in \mathbb{F}_r$  may still exists such that  $\|A - F_{\text{opt}} B\| = \|A - F_0 B\|$ . However for the application we are considering in this chapter (the downsampling problem), the condition  $\overline{\text{span}}\{e_i\}_{i \in D_r(\mathfrak{S})} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  is satisfied (see Theorem 4.10.11).

**Remark 4.10.9.** The construction of  $\mathcal{V}_i$  given in Lemma 4.10.3 is done in such a way that it suits the downsampling application. However, if we can construct subspaces  $\mathcal{V}_i \subseteq \mathcal{H}_0$  such that

1. For all  $i \in \mathbb{N}_{AP}$ ,  $f_i \in \mathcal{V}_i$  and  $f_i \perp \mathcal{V}_j, j \neq i$ .
2.  $\mathcal{V}_i$  are invariant subspaces of  $AA^*$ .
3. The subspaces  $\{\mathcal{V}_i\}_{i \in \mathbb{N}_{AP}}$  are mutually orthogonal.
4. For all  $i \in \mathbb{N}_{AP}$ , there exists an SVD of  $P_i AA^* P_i$  of the form

$$P_i AA^* P_i x = \sum_{n \in \mathbb{N}_{P_i A}} \alpha_{in}^2 \langle x, v_{in} \rangle v_{in}, \quad x \in \mathcal{H}_0 \quad (4.38)$$

where  $P_i$  is the orthogonal projection onto the subspace  $\mathcal{V}_i$  and  $\langle v_{in}, f_i \rangle \neq 0, \forall n$ .

then it can be proved that the Lemma 4.10.5, Lemma 4.10.6 and Theorem 4.10.7 remain unchanged for this newly defined  $\mathcal{V}_i$ .

Now we can summarize all the steps for obtaining  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|$  in a form of Algorithm 2.

---

**Algorithm 2** For  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|$

---

**Require:**  $A$  and  $B$  are linear bounded operators

Obtain the orthogonal projection  $P$  on  $(\text{Ker } B)^\perp$

Obtain SVD of  $AP$  and  $e_i, f_i$  defined in (4.20).

Obtain  $\mathcal{V}_i$  defined in Lemma 4.10.3 and  $\mathfrak{S}$  defined in Lemma 4.10.6

**if** the  $\mathcal{V}_i$ 's satisfy Assumption  $\mathcal{A}_1$  &  $\overline{\text{span}}\{e_i\}_{i \in D_r(\mathfrak{S})} \in \text{Im } B^* \oplus (\text{Im } B^*)^\perp$

**then**

Obtain  $B^+$

$F_{\text{opt}} = F_0$  (see Lemma 4.10.6).

**else**

Procedure limitation

**return nil**

**end if**

Calculate  $\|A - F_{\text{opt}}B\|$  by using (4.37)

**return**  $F_{\text{opt}}, \|A - F_{\text{opt}}B\|$

---

## 4.10.2 $L^\infty$ optimal downsampling solution

Problem  $\mathcal{P}_3$  with induced 2-norm can be solved by application of the Algorithm 2 (page 87). As in Section 4.9, we define  $\mathcal{S}_y := \mathcal{S}_h' \mathcal{G}_y$ . The orthogonal projection operator  $\check{P}$  on the space  $(\text{Ker } \check{S}_y)^\perp$ , and an SVD of the operator  $\check{G}_v \check{P}$  at almost

every  $\theta \in [-\pi, \pi]$ , required in Algorithm 2, are given by (4.25) and Lemma 4.9.6 respectively.

The next thing in Algorithm 2 is to find the invariant spaces  $\mathcal{V}_i$  based on an SVD of  $\check{G}_v \check{G}_v^*$  and left singular vectors  $\check{q}_i$  of  $\check{G}_v \check{P}$  (see Lemma 4.9.6). An SVD (modulo ordering) is given by [31]

$$\check{G}_v \check{G}_v^* \check{w} = \sum_{i \in \mathbb{Z}} G_v(j\omega_i)^2 \langle \check{w}, \check{e}_i \rangle \check{e}_i.$$

where  $\check{e}_i$  are defined in Lemma 4.8.2. Using Lemma 4.9.6, an SVD (modulo ordering) of the operator  $\check{G}_v \check{P} : L^2[0, h) \rightarrow L^2[0, h)$  at almost all  $\theta \in [-\pi, \pi]$  is given by

$$\check{G}_v \check{P} \check{w} = \sum_{i \in \mathbb{M}, \alpha_i \neq 0} \sigma_i \langle \check{w}, \check{p}_i \rangle \check{q}_i.$$

Hence, the  $\mathcal{V}_i$  defined in Lemma 4.10.3 in this case equal

$$\mathcal{V}_i := \overline{\text{span}}\{\check{e}_{i+Mk}\}_{k \in \mathbb{Q}_i}, \quad (4.39)$$

where  $\mathbb{Q}_i := \{k \in \mathbb{N} : \langle \check{q}_i, \check{e}_{i+Mk} \rangle \neq 0\}$  and  $\check{q}_i$  are defined in Lemma 4.9.6.

Assumption  $\mathcal{A}_1$  is satisfied in the downsampling problem as shown in the following lemma.

**Lemma 4.10.10.**  $\mathcal{V}_i$  defined in (4.39) are mutually orthogonal.

*Proof.* For  $i \neq j$ ,  $\mathcal{V}_i \perp \mathcal{V}_j$  as  $\check{e}_k \forall k \in \mathbb{Z}$  are orthogonal to each other at each  $\theta \in [-\pi, \pi]$ .  $\square$

By the above Lemma the  $\mathcal{V}_i$ 's are (mutually) orthogonal (Assumption  $\mathcal{A}_1$ ), therefore we can proceed further according to Algorithm 2. Suppose  $\check{P}_i$  represents the orthogonal projection onto the space  $\mathcal{V}_i$  at each  $\theta \in [-\pi, \pi]$ . Since, the  $\check{e}_k$  are eigenvectors of the operator  $\check{G}_v \check{G}_v^*$ , we have for every  $\check{w} \in L^2[0, h)$  the

$$\check{P}_i \check{G}_v \check{G}_v^* \check{P}_i \check{w} = \sum_{k \in \mathbb{Q}_i} G_v(j\omega_{i+Mk})^2 \langle \check{w}, \check{e}_{i+Mk} \rangle \check{e}_{i+Mk}$$

Now, we give a solution of Problem  $\mathcal{P}_3$  with induced 2-norm using Theorem 4.10.7.

**Theorem 4.10.11.** Given LCTI systems  $\check{G}_v, \check{G}_y \in L^\infty \cap L^2$ . Define  $\check{G}_N := \check{G}_v(I - \check{P})$ ,  $|G_{v\max, i}| := \max_{l \in \mathbb{Q}_i} \{|G_v(j\omega_{i+Ml})|\}$ , and  $\mathfrak{S} := \{|G_{v\max, i}|\}_{i \in \mathbb{M}}$  at almost every  $\theta \in [-\pi, \pi]$ . Given  $r \leq M$ , define  $\check{F}_{\text{opt}}$  as

$$\check{F}_{\text{opt}} \vec{y}_{h'} := \sum_{k \in D_r(\mathfrak{S})} \Gamma_k \left\langle \vec{y}_{h'}, \frac{1}{\sqrt{M}} \vec{e}_k \right\rangle, \quad (4.40)$$

where  $D_r(\mathfrak{S})$  is defined in Definition 4.9.7 and  $\Gamma_k$  is defined in (4.28), whenever  $\theta \in \mathfrak{A} := \{\theta \in [-\pi, \pi] : \check{G}_y(e^{j\theta}) \neq 0\}$  and  $\tau \in [0, h)$ . For  $\theta \in [-\pi, \pi] \setminus \mathfrak{A}$ , we can take  $\check{F}_{\text{opt}} = 0$ . If this interpolator  $\mathcal{F}_{\text{opt}}$  is well-defined and stable, then it minimizes  $\|\mathcal{G}_v - \mathcal{F}\mathcal{S}_y\|_{L^\infty}$  over all interpolators  $\mathcal{F}$  with  $\text{rank } \check{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ . The optimal norm is given by

$$\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty} = \text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_v - \check{F}_{\text{opt}}\check{S}_y\| \quad (4.41)$$

where

$$\|\check{G}_v - \check{F}_{\text{opt}}\check{S}_y\| = \max_{i \notin D_r(\mathfrak{S}), i \in \mathbb{M}} (|G_{v\text{max}, i}|, \|\check{G}_N\|)$$

*Proof.* At each  $\theta \in [-\pi, \pi]$ ,  $\check{S}_y(e^{j\theta})$  is a finite dimensional operator. Therefore at almost each  $\theta \in [-\pi, \pi]$ ,  $\text{Im } \check{S}_y(e^{j\theta})$  is closed. The rest of the proof follows from Lemma 4.10.10, Theorem 4.10.7 and Remark 4.10.8.  $\square$

The quantity  $\text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_N\|$  is known as *Parrott lower bound* [45]. Note that in (4.40) the set  $\mathfrak{S}$  is defined as a set of all singular values of  $\check{G}_v\check{P}$  at a given  $\theta$  whereas in (4.27) the set  $\mathfrak{S}$  is defined as a set of maximal singular values of  $\check{P}_i\check{G}_v$ .

**Remark 4.10.12.** To solve the  $L^\infty$  downsampling problem, we have to also write the optimal interpolator  $\mathcal{F}_{\text{opt}}$  (defined in (4.40)) as a cascade of a stable downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$ . This can be done by using the same technique adopted in Theorem 4.9.9. Define  $\check{S}_{h,\text{opt}}(e^{j\theta})$  and hold  $\check{H}_{\text{opt}}(e^{j\theta})$  with kernels given in (4.29) using the set  $\mathfrak{S}$  defined in Theorem 4.10.11. If the downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\mathcal{H}_{\text{opt}}$  are well-defined and stable, then they solve the  $L^\infty$  downsampling problem. Note that if the optimal downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\mathcal{H}_{\text{opt}}$  are well-defined then interpolator  $\check{F}_{\text{opt}}$  given by (4.40) is a hybrid interpolator by definition.

In general, calculation of the *Parrott lower bound* can be very tricky but if  $G_v(j\omega)$  has finite support then, the calculation can be done by using matrices. The matrix associated with the operator  $\check{G}_N : L^2[0, h) \rightarrow L^2[0, h)$  with respect to the orthonormal bases  $\{\check{e}_i\}$  is the array  $[a_{kl}]_{k,l \in \mathbb{Z}}$  where  $a_{kl} = \langle \check{G}_N \check{e}_l, \check{e}_k \rangle$  [69, §7.6]. It can be shown that

$$a_{kl} = \begin{cases} G_v(j\omega_k) \left( \bar{\delta}_{kl} - \frac{\psi_y^*(j\omega_k)\psi_y(j\omega_l)}{\alpha_{\text{rem}(k,M)}^2} \right) & \text{rem}(k-l, M) = 0 \\ 0 & \text{rem}(k-l, M) \neq 0 \end{cases} \quad (4.42)$$

where  $\bar{\delta}_{kl} := \bar{\delta}[k-l]$  and  $\text{rem}(k, M) := k - \lfloor k/M \rfloor M$  where  $\lfloor k/M \rfloor$  denotes the largest integer less than or equal to  $k/M$ . This technique is used in the following example.

**Example 4.10.13.** Consider the Example 4.9.11 but here we find the  $L^\infty$  optimal downsampling solution instead of the  $L^2$  optimal downsampling solution.

Since  $M = 2$ , we have  $\mathbb{M} = \{0, 1\}$ . For  $i \in \mathbb{M}$ ,  $q_i$  defined in Lemma 4.9.6 becomes,

$$\check{q}_0 = \begin{cases} \frac{2}{\sqrt{5}}\check{e}_0 + \frac{1}{\sqrt{5}}\check{e}_2, & \theta \in [-\pi, 0] \\ \frac{2}{\sqrt{5}}\check{e}_0 + \frac{1}{\sqrt{5}}\check{e}_{-2}, & \theta \in [0, \pi] \end{cases}$$

$$\check{q}_1 = \frac{1}{\sqrt{2}}(\check{e}_{-1} + \check{e}_1), \quad \theta \in [-\pi, \pi]$$

Then for every  $\theta \in [-\pi, \pi]$ ,  $G_{\text{vmax},i}$  defined in Theorem 4.10.11 becomes

$$|G_{\text{vmax},i}| = \begin{cases} 1 & i = 0 \\ 0.95 & i = 1 \end{cases}$$

Clearly  $D_1(\mathfrak{S}) = \{0\}$  in the Theorem 4.10.11. Using Theorem 4.10.11, if  $r = 1$  then an  $L^\infty$  optimal interpolator in lifted frequency domain is given by,

$$\hat{F}_{\text{opt}x} = \begin{cases} \left( \frac{2}{3}e^{j\omega_0\tau} + \frac{1}{3}e^{j\omega_2\tau} \right) \left\langle x, \frac{1}{2}e^{j\omega_0mh'} \right\rangle & \theta \in [-\pi, 0] \\ \left( \frac{2}{3}e^{j\omega_0\tau} + \frac{1}{3}e^{j\omega_{-2}\tau} \right) \left\langle x, \frac{1}{2}e^{j\omega_0mh'} \right\rangle & \theta \in [0, \pi] \end{cases}$$

where  $m \in \{0, 1\}$ . Using (4.15) and the inverse lifted transform, we can write the discrete sampling function  $\bar{\chi}_{\text{opt}}[n]$  and the hold function  $\phi_{\text{opt}}(t)$  of the optimal downsampler and hold as

$$\bar{\chi}_{\text{opt}}[n] = \frac{1}{2} \text{sinc}\left(\frac{n}{2}\right)$$

$$\phi_{\text{opt}}(t) = \frac{2}{3} \text{sinc}(t) + \frac{1}{6} \cos\left(\frac{7\pi t}{2}\right) \text{sinc}\left(\frac{t}{2}\right)$$

See Figure 4.11. Since the optimal downsampler and hold are well-defined,  $\mathcal{F}_{\text{opt}}$  is an order-1 hybrid interpolator. Also, since  $\text{sinc}[n] \in \ell^2$ ,  $\text{sinc}(t) \in L^2$  and  $|\cos(t)|$  is bounded, it follows from lemma's 2.4.7 and 4.6.2 that the optimal downsampler and hold are stable (i.e. in  $L^\infty$ ).

Since  $G_v(j\omega)$  has finite support, we can obtain an equivalent finite dimensional matrix representation of the operator  $\check{G}_N$  for norm calculation (see [69, §7.2]). This yields,

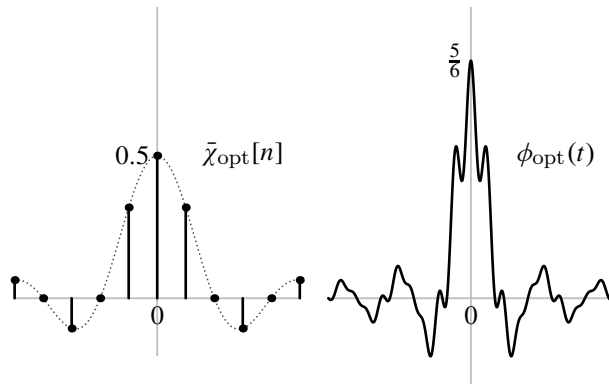
$$\|\check{G}_N\| = 0.95 \quad \forall \theta \in [-\pi, \pi]$$

and

$$\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty} = \text{ess sup}_{\theta \in [-\pi, \pi]} \max(0.95, \|\check{G}_N\|) = 0.95.$$

Comparing with Example 4.9.11, we can say that the  $L^2$  and  $L^\infty$  downsampling problem may have entirely different solutions. However, if we do full order downsampling (in this case order 2), then new  $\check{F}_{\text{opt}}$  is of rank-2 and will be the same as  $L^2$  optimal one. In this case,  $D_2(\mathfrak{S}) = \{0, 1\}$  and

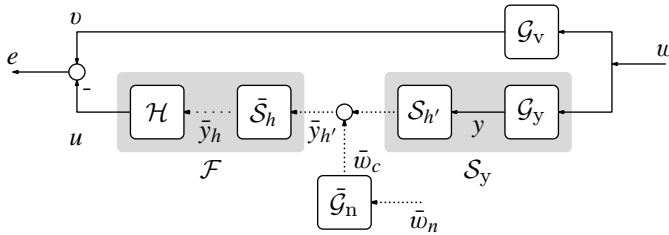
$$\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty} = \text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_N\| = 0.95$$



**Figure 4.11:** Discrete sampling function  $\bar{\chi}_{opt}[n]$  of the  $L^\infty$  optimal downsampler and Hold function  $\phi_{opt}(t)$  of the  $L^\infty$  optimal hold (right). The left figure also shows  $\frac{1}{2} \text{sinc}(\frac{t}{2})$  (dotted).

Comparing again with Example 4.9.13, increasing rank of  $\bar{F}$  from 1 to 2, decreases the  $L^2$  reconstruction error but not the  $L^\infty$  reconstruction error.

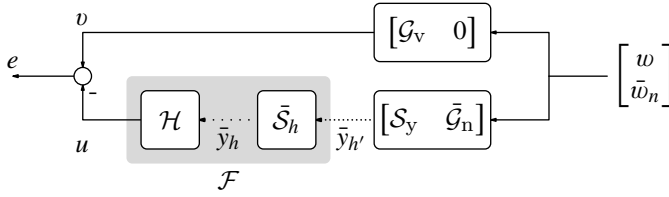
### 4.11 Downsampling in the presence of noise



**Figure 4.12:** Sampled-data setup for downsampling in the presence of noise

In this section, we will see the effect of colored noise on the  $L^2$  and  $L^\infty$  downsampling problem. Noise analysis in this section is just an application of the theory we discussed in the previous sections. The setup for noise analysis is shown in Figure 4.12. Here  $\bar{w}_c : \mathbb{Z} \rightarrow \mathbb{C}$  is colored noise modeled by the  $h'$ -time invariant system  $\bar{\mathcal{G}}_n$  with input signal  $\bar{w}_n : \mathbb{Z} \rightarrow \mathbb{C}$  which is the white noise. The  $h'$ -time invariant system  $\bar{\mathcal{G}}_n \in L^\infty$  is defined as

$$\bar{w}_c = \bar{\mathcal{G}}_n \bar{w}_n \quad : \quad \bar{w}_c[k] = \sum_{i \in \mathbb{Z}} \bar{g}_n[k - i] \bar{w}_n[i] \tag{4.43}$$



**Figure 4.13:** Modified Sampled-data set-up for noise

where  $\bar{g}_n$  is the impulse response of the discrete system  $\bar{G}_n$ .

For noise analysis, using the sampled-data system theory, we redraw Figure 4.12 as Figure 4.13. Figure 4.13 is similar to Figure 4.3 except for the signal models and the input signal. Now, the error system in the presence of noise is defined as the mapping from  $\begin{bmatrix} w \\ \bar{w}_n \end{bmatrix}$  to  $e$  and it is of the form

$$\mathcal{G}_e := \mathcal{G}_A - \mathcal{F}\mathcal{G}_B$$

where

$$\mathcal{G}_A := \begin{bmatrix} \mathcal{G}_v & 0 \end{bmatrix} \quad (4.44)$$

$$\mathcal{G}_B := \begin{bmatrix} \mathcal{S}_y & \bar{G}_n \end{bmatrix} \quad (4.45)$$

and  $\mathcal{S}_y := \mathcal{S}_{h'}\mathcal{G}_y$  is stable sampler with sampling function  $\psi_y$  (see Lemma 4.8.2).

Our aim is to obtain an  $\mathcal{F} := \mathcal{H}\bar{S}_h$  of order at most  $r$  such that  $\|\mathcal{G}_e\|_{L^2}$  or  $\|\mathcal{G}_e\|_{L^\infty}$  is minimized. Note that it is not necessary that a hybrid interpolator  $\bar{\mathcal{F}}$  (of given order) that minimizes  $\|\mathcal{G}_e\|_{L^2}$  also minimizes  $\|\mathcal{G}_e\|_{L^\infty}$ .

Similar to Section 4.9 and Section 4.10, Algorithm 1 and 2 will be used to obtain the  $L^2$  and  $L^\infty$  optimal interpolator in the presence of noise. In both Algorithm 1 and 2, we start with obtaining the projection operator  $\check{P}_B$  on the space  $(\text{Ker } \check{G}_B)^\perp$  at each  $\theta$ . Having an SVD of  $\check{G}_B$ , we can immediately get an orthonormal basis of  $(\text{Ker } \check{G}_B)^\perp$ , hence the orthogonal projector  $\check{P}_B$ . We start with an SVD of  $\check{G}_n$  at each  $\theta \in [-\pi, \pi]$ . This can be obtained by following lemma.

**Lemma 4.11.1.** *Given is an  $h'$ -time invariant system  $\bar{G}_n \in L^\infty$  as defined in (4.43) with  $\bar{G}_n(z)$  representing the classic  $z$ -transform of the kernel  $g_n[i]$  of the operator  $\bar{G}_n$ . Then, at each  $\theta \in [-\pi, \pi]$ , an SVD (modulo ordering) of the lifted operator  $\check{G}_n$  is given by*

$$\check{G}_n \bar{w}_n = \sum_{k=0}^{M-1} \bar{G}_n(e^{j\omega_k h'}) \langle \bar{w}_n, \bar{e}_k \rangle \bar{e}_k$$

where the  $\bar{e}_k$  are defined in Lemma 4.8.2.

*Proof.* See Appendix 4.D (page 113). □

**Remark 4.11.2.** Note that since  $\text{rank } \check{G}_n(e^{j\theta}) \leq M$  at almost each  $\theta \in [-\pi, \pi]$ ,  $\bar{G}_n \in L^\infty$  implies  $\check{G}_n \in L^2$  (follows from Lemma 2.4.6). Therefore  $\bar{G}_n \in L^\infty$  is same as  $\check{G}_n \in L^2 \cap L^\infty$ .

Since  $\mathcal{G}_v, \mathcal{S}_y$  and  $\bar{G}_n$  are stable (see also Lemma 4.8.2), we have that  $\mathcal{G}_A$  and  $\mathcal{G}_B$  are stable. Now, the stability of  $\mathcal{G}_A$  and  $\mathcal{G}_B$  implies that they are operators mapping  $L^2(\mathbb{R}) \times \ell^2(\mathbb{Z})$  to  $L^2(\mathbb{R}) \times \ell^2(\mathbb{Z})$ . Lifting the input signal  $[w \ \bar{w}_n]^T =: w_s$  results in the lifted signal  $\check{w}_s$  where  $\check{w}_s[k]$  belongs to space  $L^2[0, h) \times \mathbb{C}^M$ . The inner product of the space  $L^2[0, h) \times \mathbb{C}^M$  is defined as

$$\left\langle \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_3 \\ w_4 \end{bmatrix} \right\rangle = \langle w_1, w_3 \rangle_{L^2[0, h)} + \langle w_2, w_4 \rangle_{\mathbb{C}^M}$$

where  $w_1, w_3 \in L^2[0, h)$  and  $w_2, w_4 \in \mathbb{C}^M$ .

Now, an SVD of the operator  $\check{G}_B$  can be obtained by noting that  $\check{S}_y = \check{S}_h' \check{G}_y$  and  $\check{G}_n$  share the same left singular vectors.

**Lemma 4.11.3.** Let  $\mathcal{G}_v, \mathcal{G}_y$  and  $\bar{G}_n$  are in  $L^2 \cap L^\infty$ . Also let  $\alpha_k, \check{p}_k$  and  $\vec{e}_k$  be as in Lemma 4.8.2. At almost all  $\theta \in [-\pi, \pi]$ , the lifted operator  $\check{G}_B$  can be expressed as

$$\check{G}_B \check{w}_s = \sum_{k \in \mathbb{M}, \mu_k \neq 0} \mu_k \langle \check{w}_s, \check{p}_{n,k} \rangle \vec{e}_k, \quad \check{w}_s \in L^2[0, h) \times \mathbb{C}^M \quad (4.46)$$

where  $\check{w}_s := \begin{bmatrix} \check{w} \\ \vec{w}_n \end{bmatrix}$ ,

$$\mu_k := \sqrt{\frac{M}{h} \alpha_k^2 + |\bar{G}_n(e^{j\omega_k h'})|^2}, \text{ and } \check{p}_{n,k} := \frac{1}{\mu_k} \left[ \sqrt{\frac{M}{h}} \alpha_k \check{p}_k \quad \bar{G}_n(e^{j\omega_k h'})^* \vec{e}_k \right]^T$$

where  $*$  is complex conjugate operator.

*Proof.* See Appendix 4.D (page 113).  $\square$

Now the orthogonal projection operator  $\check{P}_B$  onto the space of  $(\text{Ker } \check{G}_B)^\perp$  is straight-forward because we have an orthonormal basis of the space  $(\text{Ker } \check{G}_B)^\perp$ . In fact,  $\check{P}_B$  is given by

$$\check{P}_B \check{x} = \sum_{k \in \mathbb{M}, \mu_k \neq 0} \langle \check{x}, \check{p}_{n,k} \rangle \check{p}_{n,k}, \quad \check{x} \in L^2[0, h) \times \mathbb{C}^M \quad (4.47)$$

The next step in Algorithm 1 and 2 is to obtain the SVD of the operator  $\check{G}_A \check{P}_B$ .

**Lemma 4.11.4.** Let  $\mathcal{G}_v, \mathcal{G}_y$  and  $\bar{G}_n$  are in  $L^2 \cap L^\infty$ . Also let  $\check{p}_{n,k}$ , and  $\mu_k$  be as in Lemma 4.11.3. If  $\check{P}_B : L^2[0, h) \times \mathbb{C}^M \rightarrow L^2[0, h) \times \mathbb{C}^M$  denotes the orthogonal projection into the space  $(\text{Ker } \check{G}_B)^\perp$ , then an SVD (modulo ordering) of the operator  $\check{G}_A \check{P}_B$  at almost every  $\theta \in [-\pi, \pi]$  is given by

$$\check{G}_A \check{P}_B \check{w}_s = \sum_{k \in \mathbb{M}, \mu_k \neq 0} \rho_k \langle \check{w}_s, \check{p}_{n,k} \rangle \check{q}_k, \quad \check{w}_s \in L^2[0, h) \times \mathbb{C}^M \quad (4.48)$$



where the  $\check{q}_k$  are defined in Lemma 4.9.6,  $\check{p}_{n,k}$  are defined in Lemma 4.11.3 and

$$\rho_k = \sqrt{\frac{M}{h} \frac{\sqrt{\sum_{i \in \mathbb{Z}} |G_v(j\omega_{k+Mi}) \psi_y^*(j\omega_{k+Mi})|^2}}{\mu_k}}$$

*Proof.* See Appendix 4.D (page 114).  $\square$

From this point onwards Algorithm 1 and 2 differ, so  $L^2$  and  $L^\infty$  optimal downsampling in the presence of noise is treated in two different subsections given next.

### 4.11.1 $L^2$ optimal downsampling in the presence of noise

This section describes the  $L^2$  optimal downsampling problem in the presence of noise. Since  $\check{G}_B$  and  $\check{G}_A \check{P}_B$  share right singular vectors, we can use Corollary 4.9.4. Therefore, according to the Algorithm 1, we have all the basic tools to obtain an interpolator that minimizes  $\|\check{G}_A - \mathcal{F} \check{G}_B\|_{L^2}$  over all interpolators  $\mathcal{F}$  with  $\text{rank } \check{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ . This eventually helps in solving the  $L^2$  optimal downsampling problem in the presence of noise.

**Theorem 4.11.5.** *Let  $\mathcal{G}_v$ ,  $\mathcal{G}_y$  and  $\check{G}_n$  are in  $L^2 \cap L^\infty$ . Also let  $\alpha_k$ ,  $\psi_y(j\omega)$  and  $\check{e}_k$  be as in Lemma 4.8.2. Let  $\mathfrak{S}$  be the set of finite singular values  $\rho_k$  of the operator  $\check{G}_A \check{P}_B$  defined in Lemma 4.11.4 at almost every  $\theta \in [-\pi, \pi]$  i.e.  $\mathfrak{S} := \{\rho_k\}_{k \in \mathbb{M}, \mu_k \neq 0}$ . Given  $r \leq M$ , define  $\check{F}_{n,\text{opt}}$  as*

$$\check{F}_{n,\text{opt}} \check{y}_{h'} = \sum_{k \in D_r(\mathfrak{S})} \Gamma_{n,k} \left\langle \check{y}_{h'}, \frac{1}{\sqrt{M}} \check{e}_k \right\rangle, \quad (4.49)$$

where

$$\Gamma_{n,k} := \frac{\sum_{i \in \mathbb{Z}} G_v(j\omega_{k+Mi}) \psi_y^*(j\omega_{k+Mi}) e^{j(\omega_{k+Mi})\tau}}{\alpha_k^2 + \frac{h}{M} |\check{G}_n(e^{j\omega_k h'})|^2}, \quad (4.50)$$

whenever  $\theta \in \mathfrak{A} := \{\theta \in [-\pi, \pi] : \check{G}_B(e^{j\theta}) \neq 0\}$  and  $\tau \in [0, h)$ . For  $\theta \in [-\pi, \pi] \setminus \mathfrak{A}$ , we can take  $\check{F}_{n,\text{opt}} = 0$ . If this interpolator  $\mathcal{F}_{n,\text{opt}}$  is well-defined and stable, then it minimizes  $\|\check{G}_A - \mathcal{F} \check{G}_B\|_{L^2}$  over all interpolators  $\mathcal{F}$  with  $\text{rank } \check{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ .

*Proof.* Similar to the proof of the Theorem 4.9.8.  $\square$

Note that set  $D_r(\mathfrak{S})$  can change at each  $\theta$ .

**Remark 4.11.6.** *In order to write the optimal interpolator  $\check{F}_{n,\text{opt}}$  (defined in (4.49)) as a cascade of a stable downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$ , we can use the same*

technique adopted in Theorem 4.9.9. Define lifted downsampler  $\hat{S}_{h,\text{opt}}$  and hold  $\hat{H}_{\text{opt}}$  such that their transfer functions have kernel

$$\vec{\chi}_{\text{opt}}(e^{j\theta}; -m) := \frac{1}{M} \begin{bmatrix} e^{-j\omega_{k_0} m h'} \\ \vdots \\ e^{-j\omega_{k_{r-1}} m h'} \end{bmatrix}, \quad (4.51a)$$

$$\check{\phi}_{\text{opt}}(e^{j\theta}; \tau) := [\Gamma_{n,k_0}, \dots, \Gamma_{n,k_{r-1}}]. \quad (4.51b)$$

at each  $\theta \in [-\pi, \pi]$  respectively. Here  $\{k_0, k_1, \dots, k_{r-1}\} := D_r(\mathfrak{S})$  where the ordered set  $D_r(\mathfrak{S})$  is defined in Theorem 4.11.5. If the downsampler  $\bar{S}_{h,\text{opt}}$  and hold  $\mathcal{H}_{\text{opt}}$  are well-defined and stable, then they minimize  $\|\mathcal{G}_A - \mathcal{H}\bar{S}_h\mathcal{G}_B\|_{\mathbb{L}^2}$  over all downsamplers and holds. Note the effect of the noise on  $\mathfrak{S}$ ,  $k_i$  and  $\Gamma_{n,k_i}$ . Also note that if the optimal downsampler  $\bar{S}_{h,\text{opt}}$  and hold  $\mathcal{H}_{\text{opt}}$  are well-defined then interpolator  $\bar{F}_{n,\text{opt}}$  given by (4.49) is a hybrid interpolator by definition.

The optimal error norm  $\|\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2}$  can be found using following lemma.

**Lemma 4.11.7.** *The squared  $\mathbb{L}^2$  system norm of the operator  $\mathcal{F}_{n,\text{opt}}\mathcal{G}_B$  is given by*

$$\|\mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2}^2 = \frac{1}{2\pi h} \int_{\mathfrak{A}} \sum_{k \in D_r(\mathfrak{S})} \rho_k^2 d\theta \quad (4.52)$$

where  $\rho_k$  is defined in Lemma 4.11.4 and  $\mathcal{F}_{n,\text{opt}}$  is given by (4.49). The  $\mathbb{L}^2$  system norm of the operator  $\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B$  is given by

$$\|\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2}^2 = \|\mathcal{G}_v\|_{\mathbb{L}^2}^2 - \|\mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2}^2 \quad (4.53)$$

*Proof.* The proof is similar to the proof of Theorem 4.9.12.  $\square$

Intuitively, non-zero noise increases the reconstruction error. Indeed.

**Corollary 4.11.8.**  $\|\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2} \geq \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{\mathbb{L}^2}$

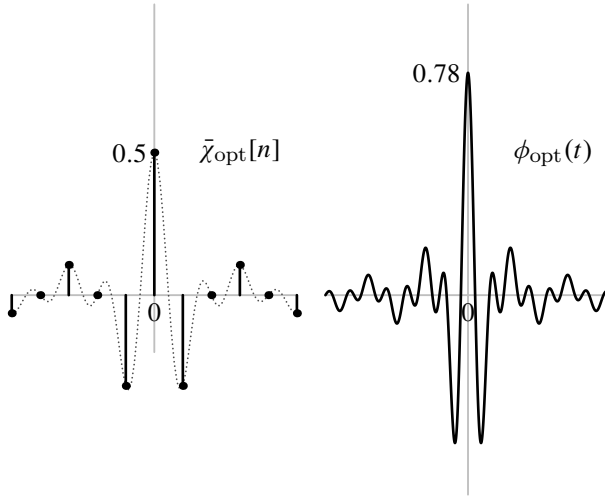
*Proof.* Since  $\sqrt{\frac{h}{M}}\mu_k \geq \alpha_k$ , we have that  $\frac{\rho_k}{\sigma_k} = \frac{\alpha_k}{\sqrt{\frac{h}{M}}\mu_k} \leq 1$ . This further implies  $\|\mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2} \leq \|\mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{\mathbb{L}^2}$ . It means  $\|\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{\mathbb{L}^2} \geq \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{\mathbb{L}^2}$  by Lemma 4.11.7 and Theorem 4.9.12.  $\square$

This fact is also illustrated with the following example.

**Example 4.11.9.** *Consider Example 4.9.11 but here we find the  $\mathbb{L}^2$  optimal downsampling solution in the presence of noise. Also, we assume that  $\bar{G}_n$  is the identity. Calculating the singular values  $\rho_i$  as defined in Lemma 4.11.4, we find that*

$$\rho_0 = 0.79$$

$$\theta \in [-\pi, \pi]$$



**Figure 4.14:** Discrete sampling function  $\bar{\chi}_{\text{opt}}[n]$  of the  $L^2$  optimal downsampler and Hold function  $\phi_{\text{opt}}(t)$  of the  $L^2$  optimal hold (right) in the presence of noise in Example 4.11.9. The left figure also shows  $\frac{1}{2} \cos(\pi t) \text{sinc}(\frac{t}{2})$  (dotted).

$$\rho_1 = 0.84$$

$$\theta \in [-\pi, \pi]$$

Note that  $\rho_1 > \rho_0$  for all  $\theta \in [-\pi, \pi]$ . Now, it follows from Theorem 4.11.5 that if  $r = 1$  then the  $L^2$  optimal interpolator in lifted frequency domain is given by

$$\check{F}_{n,\text{opt}}x = 0.78 \frac{e^{j\omega_1 \tau} + e^{j\omega_{-1} \tau}}{2} \left\langle x, \frac{1}{2} e^{j\omega_1 m h'} \right\rangle$$

for almost all  $\theta \in [-\pi, \pi]$  and  $m \in \{0, 1\}$ .

Using (4.51) and the inverse lifted transform, we can write the discrete sampling function  $\bar{\chi}_{\text{opt}}[n]$  and hold function  $\phi_{\text{opt}}(t)$  of the optimal downsampler and hold as

$$\begin{aligned} \bar{\chi}_{\text{opt}}[n] &= \frac{1}{2} \text{sinc}\left(\frac{n}{2}\right) (-1)^n \\ \phi_{\text{opt}}(t) &= 0.78 \cos(2\omega_N t) \text{sinc}(t). \end{aligned}$$

See Figure 4.14. Since the optimal downsampler and hold are well-defined,  $\mathcal{F}_{\text{opt}}$  is an order-1 hybrid interpolator. Also, since  $\text{sinc}[n] \in \ell^2$ ,  $\text{sinc}(t) \in L^2$  and  $|\cos(t)|$  is bounded, it follows from lemma's 2.4.7 and 4.6.2 that the optimal downsampler and hold are stable (i.e. in  $L^\infty$ ).

Also, using (4.52) and (4.53), we obtain

$$\|\mathcal{G}_v\|_{L^2}^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} (1 + \frac{1}{2} + 2(0.95^2)) d\theta = 3.305$$

$$\|\mathcal{F}_{n,\text{opt}}\mathcal{S}_{h'}\mathcal{G}_y\|_{L^2}^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} \rho_1^2 d\theta = 0.707$$

Hence,

$$\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_{h'}\mathcal{G}_y\|_{L^2}^2 = 2.5983.$$

Therefore, the reconstructed power ( $\|\mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^2}^2/\|\check{\mathcal{G}}_v\|_{L^2}^2$ ) is 21.38%. If we do full order downsampling (in this case order 2), then  $\mathcal{F}_{\text{opt}}$  is of order-2 and

$$\|\mathcal{F}_{\text{opt}}\mathcal{S}_{h'}\mathcal{G}_y\|_{L^2}^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} \rho_1^2 + \rho_0^2 d\theta = 1.3317$$

Then, the reconstructed power is 40.29%. Comparing with Example 4.9.13, the recovery is reduced in the presence of noise as expected.

### 4.11.2 $L^\infty$ optimal downsampling in the presence of noise

This section describes how discrete noise affects the  $L^\infty$  optimal downsampling problem. Similar to Section 4.10, obtaining a general  $L^\infty$  solution can be tricky but if Assumption  $\mathcal{A}_1$  is satisfied then the problem can be solved. Since  $\check{\mathcal{G}}_A\check{\mathcal{G}}_A^* = \check{\mathcal{G}}_v\check{\mathcal{G}}_v^*$  and the left singular vectors  $\check{q}_k$  of  $\check{\mathcal{G}}_A\check{P}_B$  (see (4.48)) are the same as the left singular vectors of  $\check{\mathcal{G}}_v\check{P}$  (see (4.26)), we have that the subspaces  $\mathcal{V}_i$  in the presence of noise are the same as in the case of without noise (see (4.39)). It is already proved in Lemma 4.10.10 that these subspaces  $\mathcal{V}_i$  are mutually orthogonal i.e. Assumption  $\mathcal{A}_1$  holds in this case.

Now, we obtain an interpolator that minimizes  $\|\mathcal{G}_A - \mathcal{F}\mathcal{G}_B\|_{L^\infty}$  over all interpolators  $\mathcal{F}$  with rank  $\check{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ . This eventually helps in solving the  $L^\infty$  optimal downsampling problem in the presence of noise.

**Theorem 4.11.10.** *Let  $\mathcal{G}_v$ ,  $\mathcal{G}_y$  and  $\check{\mathcal{G}}_n$  are in  $L^2 \cap L^\infty$ . Also let  $\alpha_k$ ,  $\psi_y(j\omega)$  and  $\vec{e}_k$  be as in Lemma 4.8.2. Let  $|\check{G}_{v,\max,i}|$  and  $D_r(\mathfrak{S})$  be as in Theorem 4.10.11. Given  $r \leq M$ , define  $\check{F}_{n,\text{opt}}$  as*

$$\check{F}_{n,\text{opt}}\vec{y}_{h'} = \sum_{k \in D_r(\mathfrak{S})} \Gamma_{n,k} \left\langle \vec{y}_{h'}, \frac{1}{\sqrt{M}} \vec{e}_k \right\rangle, \quad (4.54)$$

where  $\Gamma_{n,k}$  is defined in (4.50), whenever  $\theta \in \mathfrak{A} := \{\theta \in [-\pi, \pi] : \check{\mathcal{G}}_B(e^{j\theta}) \neq 0\}$  and  $\tau \in [0, h)$ . For  $\theta \in [-\pi, \pi] \setminus \mathfrak{A}$ , we can take  $\check{F}_{n,\text{opt}} = 0$ . If this interpolator  $\mathcal{F}_{n,\text{opt}}$  is well-defined and stable, then it minimizes  $\|\mathcal{G}_A - \mathcal{F}\mathcal{G}_B\|_{L^\infty}$  over all interpolators  $\mathcal{F}$  with rank  $\check{F} \leq r$  at each  $\theta \in [-\pi, \pi]$ . The optimal norm is given by

$$\|\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{L^\infty} = \text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{\mathcal{G}}_A - \check{F}_{n,\text{opt}}\check{\mathcal{G}}_B\| \quad (4.55)$$

where

$$\|\check{G}_A - \check{F}_{n,\text{opt}}\check{G}_B\| = \max_{\substack{i \notin D_r(\mathfrak{S}) \\ i \in \mathbb{M}}} \left( |G_{v\text{max},i}|, \|\check{G}_A(I - \check{P}_B)\| \right).$$

Here  $\|\cdot\|$  is the induced 2-norm.

*Proof.* Proof follows from Theorem 4.10.7.  $\square$

The quantity  $\text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_A(I - \check{P}_B)\|$  is known as the *Parrott lower bound* in the presence of noise.

**Remark 4.11.11.** In order to write the optimal interpolator  $\check{F}_{n,\text{opt}}$  (defined in (4.54)) as a cascade of a stable downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$ , we can use the same technique adopted in Theorem 4.9.9. Define  $\check{S}_{h,\text{opt}}(e^{j\theta})$  and hold  $\check{H}_{\text{opt}}(e^{j\theta})$  with kernels given in (4.51) using the set  $\mathfrak{S}$  defined in Theorem 4.11.10. If the downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$  are well-defined and stable, then they minimize  $\|\check{G}_A - \mathcal{H}\check{S}_h\check{G}_B\|_{L^\infty}$  over all downsamplers and holds. Note that if the optimal downsampler  $\check{S}_{h,\text{opt}}$  and hold  $\check{H}_{\text{opt}}$  are well-defined then interpolator  $\check{F}_{n,\text{opt}}$  given by (4.54) is a hybrid interpolator by definition.

Intuitively, noise should increase the Parrott lower bound as well as the reconstruction error, which can be seen analytically as follows.

**Lemma 4.11.12.** Let  $\check{G}_v$  and  $\check{S}_y, \check{P}$  be as in Theorem 4.10.11. Then,

1.  $\|\check{G}_A - \mathcal{F}_{n,\text{opt}}\check{G}_B\|_{L^\infty} \geq \|\check{G}_v - \mathcal{F}_{\text{opt}}\check{S}_y\|_{L^\infty}$
2.  $\text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_A(I - \check{P}_B)\| \geq \text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_v(I - \check{P})\|$

where  $\mathcal{F}_{\text{opt}}$  is the optimal interpolator in the noise-free case defined in Theorem 4.10.11 and  $\mathcal{F}_{n,\text{opt}}$  is the optimal interpolator in the noise case defined in Theorem 4.11.10.

*Proof.* See Appendix 4.D (page 114).  $\square$

Similar to Section 4.10.2, if  $G_v(j\omega)$  has finite support then calculation of the Parrott lower bound in the presence of noise can be done using matrices. The matrix associated with the operator  $\check{G}_A(I - \check{P}_B)$ , which maps from  $L^2[0, h) \times \mathbb{C}^M$  to  $L^2[0, h)$ , with respect to the orthonormal bases  $\left\{ \left[ \begin{array}{c} \check{e}_i \\ 0 \end{array} \right] \right\}_{i \in \mathbb{N}}, \left\{ \left[ \begin{array}{c} 0 \\ \check{e}_k \end{array} \right] \right\}_{k \in \mathbb{M}}$  and  $\{\check{e}_k\}$  is the matrix  $[[b_{kl}]_{k \in \mathbb{Z}, l \in \mathbb{M}} \quad [a_{kl}]_{k, l \in \mathbb{Z}}]$  where (see [69, §7.6])

$$b_{kl} := \left\langle \check{G}_A(I - \check{P}_B) \left[ \begin{array}{c} 0 \\ \check{e}_l \end{array} \right], \check{e}_k \right\rangle \quad \text{and} \quad a_{kl} := \left\langle \check{G}_A(I - \check{P}_B) \left[ \begin{array}{c} \check{e}_l \\ 0 \end{array} \right], \check{e}_k \right\rangle.$$

This equal to

$$b_{kl} = \begin{cases} -\frac{\sqrt{M}\check{G}_n(e^{j\omega_l h'})G_v(j\omega_k)\psi_v^*(j\omega_k)}{\sqrt{h}\mu_l^2} & \text{rem}(k, M) = l \\ 0 & \text{rem}(k, M) \neq l \end{cases} \quad (4.56a)$$

$$a_{kl} = \begin{cases} G_v(j\omega_k) \left( \bar{\delta}_{kl} - \frac{\psi_y^*(j\omega_k)\psi_y(j\omega_l)}{M^2 \mu_{\text{rem}(k,M)}} \right) & \text{rem}(k-l, M) = 0 \\ 0 & \text{rem}(k-l, M) \neq 0 \end{cases} \quad (4.56b)$$

where  $\bar{\delta}_{kl} = \bar{\delta}[k-l]$  and  $\text{rem}(k, M)$  is defined in (4.42).

Although, the Parrott lower bound may increase in the presence of noise, it does not mean that the reconstruction error has to increase. There are cases where the noise does not increase the reconstruction error at all. This happens if the Parrott lower bound in the presence of noise is smaller than or equal to

$$\text{ess sup}_{\theta \in [-\pi, \pi]} \max_{\substack{i \in D_r(\mathfrak{S}) \\ i \in \mathbb{M}}} |G_{v_{\max}, i}|.$$

This is contrary to the  $L^2$  case where non-zero noise will always increase the error norm as seen in the following example.

**Example 4.11.13.** Consider Example 4.10.13 but here we determine an  $L^\infty$  optimal downsampling solution in the presence of noise. Also, we assume that  $\check{G}_n$  is the identity. Similar to Example 4.10.13, we find that  $D_1(\mathfrak{S}) = \{0\}$  in this example also. If  $r = 1$  then an  $L^\infty$  optimal interpolator (4.54) in lifted frequency domain is given by,

$$\check{F}_{n, \text{opt}} x = \begin{cases} \left( \frac{1}{2} e^{j\omega_0 \tau} + \frac{1}{4} e^{j\omega_2 \tau} \right) \left\langle x, \frac{1}{2} e^{j\omega_0 m h'} \right\rangle & \theta \in [-\pi, 0] \\ \left( \frac{1}{2} e^{j\omega_0 \tau} + \frac{1}{4} e^{j\omega_{-2} \tau} \right) \left\langle x, \frac{1}{2} e^{j\omega_0 m h'} \right\rangle & \theta \in [0, \pi] \end{cases}$$

Using (4.15) and the inverse lifted transform, we can write the discrete sampling function  $\check{\chi}_{\text{opt}}[n]$  and hold function  $\phi_{\text{opt}}(t)$  of the optimal downsampler and hold as

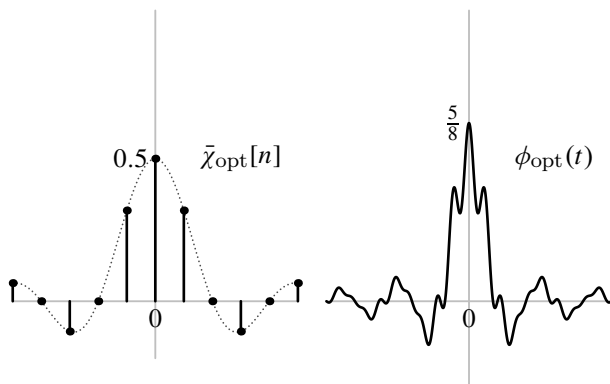
$$\begin{aligned} \check{\chi}_{n, \text{opt}}[n] &= \frac{1}{2} \text{sinc}\left(\frac{n}{2}\right) \\ \phi_{n, \text{opt}}(t) &= \frac{1}{2} \text{sinc}(t) + \frac{1}{8} \cos\left(\frac{7\pi t}{2}\right) \text{sinc}\left(\frac{t}{2}\right). \end{aligned}$$

See Figure 4.15 (page 100). Since the optimal downsampler and hold are well-defined,  $\mathcal{F}_{\text{opt}}$  is an order-1 hybrid interpolator. Also, since  $\text{sinc}[n] \in \ell^2$ ,  $\text{sinc}(t) \in L^2$  and  $|\cos(t)|$  is bounded, it follows from lemma's 2.4.7 and 4.6.2 that the optimal downsampler and hold are stable (i.e. in  $L^\infty$ ). Since  $G_v(j\omega)$  has finite support, we can obtain an equivalent finite dimensional matrix representation of the operator  $\check{G}_A(I - \check{P}_B)$  for norm calculation (see (4.56)). This yields,

$$\|\check{G}_A(I - \check{P}_B)\| = 0.95 \quad \forall \theta \in [-\pi, \pi]$$

Hence the Parrott lower bound in the presence of noise is  $\text{ess sup}_{\theta \in [-\pi, \pi]} \|\check{G}_A(I - \check{P}_B)\| = 0.95$ . Also,

$$\|\check{G}_A - \mathcal{F}_{\text{opt}} \mathcal{G}_B\|_{L^\infty} = \text{ess sup}_{\theta \in [-\pi, \pi]} (0.95, \|\check{G}_A(I - \check{P}_B)\|) = 0.95.$$



**Figure 4.15:** Discrete sampling function  $\bar{\chi}_{\text{opt}}[n]$  of the  $L^\infty$  optimal downsampler and Hold function  $\phi_{\text{opt}}(t)$  of the  $L^\infty$  optimal hold (right) in the presence of noise. The left figure also shows  $\frac{1}{2} \text{sinc}(\frac{t}{2})$  (dotted).

Although  $\check{G}_A(I - \check{P}_B)$  in this example does not equal to the  $\check{G}_N$  of Example 4.10.13, but  $\|\mathcal{G}_A - \mathcal{F}_{\text{opt}}\mathcal{G}_B\|_{L^\infty}$  is equal to  $\|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty}$ . So in this example, the unit power noise does not increase the reconstruction error.

## 4.12 Concluding Remarks

This chapter describes the optimal non-causal solution of the downsampling problem of minimizing  $\|\mathcal{G}_v - \mathcal{H}\bar{\mathcal{S}}_h\mathcal{S}_{h'}\mathcal{G}_y\|_{L^2}$  and  $\|\mathcal{G}_v - \mathcal{H}\bar{\mathcal{S}}_h\mathcal{S}_{h'}\mathcal{G}_y\|_{L^\infty}$  over all stable holds (with input dimension  $r$ ) and downsamplers (with output dimension  $r$ ). The result is generic in the sense that the spectrum of input signals need not be bandlimited. Obtaining the optimal downsampler becomes more intricate because of aliased frequencies. Although the solution is provided for LCTI  $\mathcal{G}_v$ , by proceeding in similar manner, the downsampling problem can be solved for any linear  $h$ -time shift invariant system  $\mathcal{G}_v$ , using Theorem 4.9.2 and Theorem 4.10.7. The effect of noise is also studied in this chapter. As expected noise increases the reconstruction error  $L^2$ , but it is not necessary that noise increases the  $L^\infty$  reconstruction error norm.

## 4.A Proofs of the results in sections 4.5, 4.7 and 4.8

*Proof of Theorem 4.5.4.* Assume that the classic  $z$ -transform  $\bar{f}(z)$  of  $\bar{f}$  is given. The  $z$ -transform (4.5) is defined for a lifted signal spaced at interval  $h$  whereas the discrete signal  $\bar{f}(n)$  is spaced at interval  $h' = h/M$ . Therefore, we need a fractional  $z$ -transform, in other words  $z^{\frac{1}{M}}$ -transform. But for any  $z = |z|e^{j\text{Arg}z} \in$

$\mathbb{C}$  where  $\text{Arg}$  is the principle argument function [48, Chapter 3], the  $z^{\frac{1}{M}}$  not unique and has  $M$  roots given by  $z^{\frac{1}{M}} = |z|^{\frac{1}{M}} e^{j\frac{\text{Arg } z}{M}} e^{j\frac{2\pi i}{M}}$  for all  $i \in \mathbb{M}$ . We denote the roots as  $\tilde{z}_i = |z|^{\frac{1}{M}} e^{j\frac{\text{Arg } z}{M}} e^{j\frac{2\pi i}{M}}$ . The classic  $z$ -transform of  $f[n]$  at  $\tilde{z}_i$  now becomes

$$\begin{aligned}\bar{f}(\tilde{z}_i) &= \sum_{n \in \mathbb{Z}} \bar{f}[n] \tilde{z}_i^{-n} = \sum_{k \in \mathbb{Z}} \sum_{m=0}^{M-1} \bar{f}[Mk + m] \tilde{z}_i^{-(Mk+m)} \\ &= \sum_{m=0}^{M-1} \tilde{z}_i^{-m} \sum_{k \in \mathbb{Z}} \bar{f}[Mk + m] \tilde{z}_i^{-Mk} \\ &= \sum_{m=0}^{M-1} \tilde{z}_i^{-m} \sum_{k \in \mathbb{Z}} \bar{f}[Mk + m] z^{-k} \\ &= \sum_{m=0}^{M-1} \vec{f}(z; m) |z|^{-\frac{1}{M}m} e^{-j\frac{\text{Arg } z}{M}m} e^{-j\frac{2\pi i}{M}m}\end{aligned}$$

Define  $s_0 := \frac{1}{M}(\log(|z|) + j \text{Arg } z)$ , then  $e^{s_0} = |z|^{\frac{1}{M}} e^{j\frac{\text{Arg } z}{M}} = \tilde{z}_i e^{-j\frac{2\pi i}{M}}$  and  $e^{s_0 M} = |z| e^{j \text{Arg } z} = z$ . This implies

$$\frac{1}{M} \bar{f}(\tilde{z}_i) = \frac{1}{M} \sum_{m=0}^{M-1} \left( \vec{f}(z; m) e^{-s_0 m} \right) e^{-j\frac{2\pi i}{M}m}$$

Since  $\vec{f}(z; m) e^{-s_0 m}$  is a periodic discrete signal as a function of  $m$  with period  $M$ . Hence,  $\frac{1}{M} \bar{f}(\tilde{z}_i)$  is  $i$ th discrete Fourier series coefficient of sequence  $\vec{f}(z; m) e^{-s_0 m}$  in  $m$ . Therefore,

$$\begin{aligned}\vec{f}(z; m) e^{-s_0 m} &= \frac{1}{M} \sum_{i=0}^{M-1} \bar{f}(\tilde{z}_i) e^{j\frac{2\pi m}{M}i} \\ \vec{f}(z; m) &= \frac{1}{M} \sum_{i=0}^{M-1} \bar{f}(\tilde{z}_i) \tilde{z}_i^m\end{aligned}$$

Conversely, assume that the discrete lifted  $z$ -transform  $\vec{f}(z)$  of  $\bar{f}$  is given. Then,

$$\begin{aligned}\sum_{m=0}^{M-1} \vec{f}(\tilde{z}^M; m) \tilde{z}^{-m} &= \sum_{m=0}^{M-1} \tilde{z}^{-m} \sum_{k \in \mathbb{Z}} \bar{f}[Mk + m] \tilde{z}^{-Mk} \\ &= \sum_{k \in \mathbb{Z}} \sum_{m=0}^{M-1} \bar{f}[Mk + m] \tilde{z}^{-(Mk+m)} = \sum_{n \in \mathbb{Z}} \bar{f}[n] \tilde{z}^{-n} \\ &= \bar{f}(\tilde{z})\end{aligned}$$

□



*Proof of (4.11).* Consider the  $z$ -transform of the output of the lifted downsampler (4.3):

$$\begin{aligned}
 \bar{y}(z) &= \sum_{n \in \mathbb{Z}} \bar{y}_h z^{-n} = \sum_{n \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \acute{S}_h[n-i] \vec{y}_{h'}[i] z^{-n} \\
 &= \sum_{i \in \mathbb{Z}} \left( \sum_{n \in \mathbb{Z}} \acute{S}_h[n-i] z^{-(n-i)} \right) \vec{y}_{h'}[i] z^{-i} \\
 &= \left( \sum_{n \in \mathbb{Z}} \acute{S}_h[n] z^{-n} \right) \sum_{i \in \mathbb{Z}} \vec{y}_{h'}[i] z^{-i} \\
 &= \acute{S}_h(z) \vec{y}_{h'}(z)
 \end{aligned}$$

Now it follows from (4.3) that

$$\begin{aligned}
 \bar{y}(z) &= \sum_{m \in \mathbb{M}} \left( \sum_{k \in \mathbb{Z}} \vec{\chi}[k; -m] z^{-k} \right) \vec{y}_{h'}(z; m) \\
 &= \sum_{m \in \mathbb{M}} \vec{\chi}(z; -m) \vec{y}_{h'}(z; m).
 \end{aligned}$$

By the above equation, the transfer function  $\acute{S}_h(z)$  is an operator whose kernel is given by  $\vec{\chi}(z, -m)$ .  $\square$

*Proof of Theorem 4.7.2.* Given that the output  $\bar{y}_h$  of the downsampler has  $r$  rows. Therefore at each  $\theta \in [-\pi, \pi]$ , we have  $\bar{y}_h(e^{j\theta}) \in \mathbb{C}^r$ . In addition to the above, we have

$$\check{u}(e^{j\theta}; \tau) = \check{F}(e^{j\theta}) \bar{y}_h(e^{j\theta}) = \check{\phi}(e^{j\theta}; \tau) \bar{y}_h(e^{j\theta})$$

Therefore,  $\dim(\text{Im } \check{F}(e^{j\theta})) \leq \dim(\text{Im } \acute{S}_h(e^{j\theta})) \leq r$  at almost every  $\theta \in [-\pi, \pi]$ . Hence  $\text{rank}(\check{F}) \leq r$ .  $\square$

*Proof of Lemma 4.8.2.* In the proof we drop  $(e^{j\theta})$  from all signals or systems unless necessary. *Part 1:* Trivial.

*Part 2:* Since  $\psi_y(j\omega) = \psi(j\omega)G_y(j\omega)$  and  $G_y \in L^2$ , we have

$$\int_{-\infty}^{\infty} |\psi_y(j\omega)|^2 d\omega \leq \|\psi\|_{L^\infty} \|G_y\|_{L^2(\mathbb{T})} < \infty,$$

where  $\|\psi\|_{L^\infty} := \text{ess sup}_{\omega \in \mathbb{R}} |\psi(j\omega)|$ . This implies  $\psi_y \in L^2$  by Parseval's theorem.

*Part 3:* Similar to the derivation of (4.12), the system  $\vec{y}_{h'} = \acute{S}_y \check{w}$  equals

$$\vec{y}_{h'}(e^{j\theta}; m) = \int_0^h \check{\psi}_y(e^{j\theta}; mh' - \tau) \check{w}(\tau) d\tau$$

where  $m \in \{0, \dots, M-1\}$ . Using the Key lifting formula (2.18), we have

$$\psi_y(e^{j\theta}; mh' - \tau) = \sum_k \psi_y(j\omega_k) e^{j\omega_k(mh' - \tau)}.$$

Therefore  $\vec{y}_{h'}$  can be written as

$$\begin{aligned} \vec{y}_{h'}(e^{j\theta}; m) &= \frac{1}{h} \sum_k \psi_y(j\omega_k) \left( \int_0^h \check{w}(\tau) e^{-j\omega_k \tau} d\tau \right) e^{j\omega_k mh'} \\ &= \frac{1}{\sqrt{h}} \sum_k \psi_y(j\omega_k) \langle \check{w}, \check{e}_k \rangle e^{j\omega_k mh'} \\ &= \sqrt{\frac{M}{h}} \sum_{k \in \mathbb{Z}} \psi_y(j\omega_k) \langle \check{w}, \check{e}_k \rangle \vec{e}_k \end{aligned} \quad (4.57)$$

Note that, for any given  $\theta$ , the  $\{\vec{e}_k\}_{k \in \mathbb{M}}$  forms an orthonormal basis in  $\mathbb{C}^M$  and that  $\vec{e}_k = \vec{e}_{k+Mi} \quad \forall i \in \mathbb{Z}$  (see Remark 4.5.6). So, the infinite sum in (4.57) equals a finite sum:

$$\begin{aligned} \vec{y}_{h'} &= \sqrt{\frac{M}{h}} \sum_{k \in \mathbb{Z}} \psi_y(j\omega_k) \langle \check{w}, \check{e}_k \rangle \vec{e}_k \\ &= \sqrt{\frac{M}{h}} \sum_{k=0}^{M-1} \sum_{i \in \mathbb{Z}} \psi_y(j\omega_{k+Mi}) \langle \check{w}, \check{e}_{k+Mi} \rangle \vec{e}_{k+Mi} \\ &= \sqrt{\frac{M}{h}} \sum_{k=0}^{M-1} \alpha_k \left\langle \check{w}, \sum_{i \in \mathbb{Z}} \frac{\psi_y^*(j\omega_{k+Mi})}{\alpha_k} \check{e}_{k+Mi} \right\rangle \vec{e}_k \\ &= \sqrt{\frac{M}{h}} \sum_{k=0}^{M-1} \alpha_k \langle \check{w}, \check{p}_k \rangle \vec{e}_k \end{aligned} \quad (4.58)$$

Since  $\psi_y \in L^2$ , we have  $\alpha_k < \infty$  for almost all  $\theta$ , this again implies  $\check{p}_k \in L^2[0, h)$  for almost all  $\theta$ . Note that still, we have for almost all  $\theta$  that  $\langle \check{p}_i, \check{p}_k \rangle = \delta[i - k]$ .

*Part 4:* Follows from Lemma 2.4.7. □

*Proof of Theorem 4.8.4.* In Problem  $\mathcal{P}_2$ , we need  $\mathcal{F} \in L^\infty$ . By Lemma 2.4.4 this means implies a bounded  $\hat{F}$  at almost all  $\theta$ .

Now, the  $L^2$  norm of the error system is given by

$$\|\mathcal{G}_v - \mathcal{F}\mathcal{S}_y\|_{L^2}^2 = \frac{1}{2\pi h} \int_{-\pi}^{\pi} \|\check{G}_v(e^{j\theta}) - \hat{F}(e^{j\theta})\check{S}_y(e^{j\theta})\|_{HS}^2 d\theta$$

Since,  $\|\check{G}_v(e^{j\theta}) - \hat{F}(e^{j\theta})\check{S}_y(e^{j\theta})\|_{HS} \geq 0$ , therefore minimizing at each  $\theta$  with constraint of  $\text{rank}(\hat{F}(e^{j\theta})) \leq r$ , defines  $\hat{F}_{\text{opt}}$  for all  $\theta$ .

The  $L^\infty$  norm of the error system is given by

$$\|\mathcal{G}_v - \mathcal{F}\mathcal{S}_y\|_{L^\infty} = \operatorname{ess\,sup}_{\theta \in [-\pi, \pi]} \|\check{\mathcal{G}}_v(e^{j\theta}) - \check{F}(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_\infty$$

Lets us assume  $\gamma_{\text{opt}} := \min_{\mathcal{F}} \|\mathcal{G}_v - \mathcal{F}\mathcal{S}_y\|_{L^\infty}$  obtained by a  $\mathcal{F}_1$  such that  $\mathcal{F}_1 \neq \mathcal{F}_{\text{opt}}$  and  $\operatorname{ess\,sup}_{\theta \in [-\pi, \pi]} \operatorname{rank}(\check{F}_1(e^{j\theta})) \leq r$ . But it follows from definition of  $\check{F}_{\text{opt}}$  that  $\|\check{\mathcal{G}}_v(e^{j\theta}) - \check{F}_1(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_\infty \geq \|\check{\mathcal{G}}_v(e^{j\theta}) - \check{F}_{\text{opt}}(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_\infty$  at almost all  $\theta \in [-\pi, \pi]$ , therefore

$$\begin{aligned} & \operatorname{ess\,sup}_{\theta \in [-\pi, \pi]} \|\check{\mathcal{G}}_v(e^{j\theta}) - \check{F}_1(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_\infty \\ & \geq \operatorname{ess\,sup}_{\theta \in [-\pi, \pi]} \|\check{\mathcal{G}}_v(e^{j\theta}) - \check{F}_{\text{opt}}(e^{j\theta})\check{\mathcal{S}}_y(e^{j\theta})\|_\infty \\ & \Rightarrow \|\mathcal{G}_v - \mathcal{F}_1\mathcal{S}_y\|_{L^\infty} \geq \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty} \end{aligned}$$

But we assumed  $\mathcal{F}_1$  is optimal, this means  $\|\mathcal{G}_v - \mathcal{F}_1\mathcal{S}_y\|_{L^\infty} \leq \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty}$ . Therefore,  $\|\mathcal{G}_v - \mathcal{F}_1\mathcal{S}_y\|_{L^\infty} = \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty}$ . This implies  $\mathcal{F}_{\text{opt}}$  is also optimal.  $\square$

## 4.B Proofs of the results in Section 4.9

*Proof of Lemma 4.9.1.* Let  $\{b_0, b_1, \dots\}$  be an orthonormal basis of  $\operatorname{Ker} B$ . Also, let  $\{b_0^\perp, b_1^\perp, \dots\}$  be an orthonormal basis of  $(\operatorname{Ker} B)^\perp$ . For all  $i \in \mathbb{N}$ , we have  $Bb_i = 0$  and  $Pb_i = 0$  (as  $b_i \in \operatorname{Ker} B$ ) and  $A(I - P)b_i^\perp = 0$  (as  $b_i^\perp \in (\operatorname{Ker} B)^\perp$ ). Since  $\{b_0, b_1, \dots\} \cup \{b_0^\perp, b_1^\perp, \dots\}$  forms an orthonormal basis of  $\mathcal{H}$ , we have

$$\begin{aligned} \|A - FB\|_{HS}^2 &= \sum_i \|(A - FB)b_i\|^2 + \sum_i \|(A - FB)b_i^\perp\|^2 \\ &= \sum_i \|A_N b_i\|^2 + \sum_i \|(AP - FB)b_i^\perp\|^2 \\ &= \|A_N\|_{HS}^2 + \|AP - FB\|_{HS}^2 \end{aligned}$$

Finally, since  $A$  and  $B$  are Hilbert-Schmidt operators, therefore  $A(I - P)$ ,  $AP - FB$  and  $A - FB$  are Hilbert-Schmidt operators (this can be proved easily using [47, Thm 6.16]).  $\square$

*Proof of Theorem 4.9.2.* Since  $A$  is compact (as  $A$  is Hilbert-Schmidt),  $AP$  is also compact, therefore SVD of  $AP$  exists [69]. Now, first it will be shown that the  $F_{\text{opt}}$  given in (4.21a) is a bounded operator iff  $\mathbb{W} \subseteq \operatorname{Im} B^* \oplus (\operatorname{Im} B^*)^\perp$ . Two situations arise here:  $\operatorname{Im} B$  is closed or not closed.

- *Suppose  $\operatorname{Im} B$  is closed.*

Since  $\operatorname{Im} B$  is closed, a bounded  $B^+ : \mathcal{H}_1 \rightarrow (\operatorname{Ker} B)^\perp$  exists and  $B^+B = P$  [13, Proposition 2.3]. Boundedness of the operator  $F_{\text{opt}}$  follows from [47, Exercise 6.11] and the fact that  $B^+x \in \mathcal{H}$  for all  $x \in \mathcal{H}_1$ .

- Suppose  $\text{Im } B$  is not closed.

Note that  $B$  is a closed and dense operator because it is a bounded operator on the domain  $\mathcal{H}_1$ . Since  $\text{Im } B$  is not closed, the pseudo-inverse  $B^+$  exists but on domain  $D(B^+) = \text{Im } B \oplus (\text{Im } B)^\perp$  [13, Definition 2.2]. This makes  $B^+$  a dense operator, hence this proves the existence of  $(B^+)^*$ . Also,  $\text{Im } B$  is not closed, implies  $\text{Im } B^*$  is not closed [67, §7.5], therefore  $(B^*)^+$  has domain  $D((B^*)^+) = \text{Im } B^* \oplus (\text{Im } B^*)^\perp$ .

Although  $B^+$  is unbounded for domain  $\mathcal{H}_1$ , the boundedness of  $F_{\text{opt}}$  depends upon  $e_i$  as shown below. Using  $(B^+)^* = (B^*)^+$  [21, thm 9.3.2], for any  $x \in \mathcal{H}_1$  we have

$$F_{\text{opt}}x = \sum_{i=0}^{r-1} \sigma_i \langle x, (B^+)^* e_i \rangle f_i = \sum_{i=0}^{r-1} \sigma_i \langle x, (B^*)^+ e_i \rangle f_i.$$

Now, if for all  $i \in \{0, \dots, r-1\}$ ,  $e_i \in \text{Im } B^* \oplus (\text{Im } B^*)^\perp$ , then  $\|(B^*)^+ e_i\| < \infty$  because the domain of  $(B^*)^+$  is  $\text{Im } B^* \oplus (\text{Im } B^*)^\perp$ . Hence, by the Pythagoras theorem and the Cauchy-Schwartz inequality

$$\begin{aligned} \frac{\|F_{\text{opt}}x\|^2}{\|x\|^2} &= \frac{1}{\|x\|^2} \sum_{i=0}^{r-1} \sigma_i^2 |\langle x, (B^*)^+ e_i \rangle|^2 \\ &\leq \sum_{i=0}^{r-1} \sigma_i^2 \|(B^*)^+ e_i\|^2 < \infty \end{aligned}$$

Hence,  $F_{\text{opt}}$  is a bounded operator on domain  $\mathcal{H}_1$ .

Now, let us assume  $F_{\text{opt}}$  is a bounded operator on domain  $\mathcal{H}_1$  but  $e_n \notin \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  for some  $n \in \{0, \dots, r-1\}$ . Let  $g_n(x) := \langle x, (B^*)^+ e_n \rangle$  for all  $x \in \mathcal{H}_1$ . Since  $\sigma_i \neq 0$  and  $g_i(x) = \frac{1}{\sigma_i} \langle F_{\text{opt}}x, f_i \rangle$ , we have using Cauchy-Schwartz

$$|g_i(x)| \leq \frac{1}{\sigma_i} \|F_{\text{opt}}x\| \leq \frac{1}{\sigma_i} \|F_{\text{opt}}\| \|x\|.$$

Now, boundedness of  $F_{\text{opt}}$  implies  $\frac{|g_i(x)|}{\|x\|} \leq \frac{1}{\sigma_i} \|F_{\text{opt}}\| < \infty \forall i \in [0, r-1]$ . Hence  $g_n : \mathcal{H}_1 \rightarrow \mathbb{C}$  is a bounded functional in the dual space of  $\mathcal{H}_1$ . According to the Riesz-Frechet theorem [47, thm 4.31], this is only possible when  $(B^*)^+ e_n$  is defined and belongs to  $\mathcal{H}_1$ . This implies  $e_n \in \text{Im } B^* \oplus (\text{Im } B^*)^\perp$ , which is a contradiction. It means when  $e_n \notin \text{Im } B^* \oplus (\text{Im } B^*)^\perp$  for some  $n \in \{0, \dots, r-1\}$  then  $F_{\text{opt}}$  is not a bounded operator on domain  $\mathcal{H}_1$ .

Now it can be shown that  $F_{\text{opt}}$  is optimal also. For any  $x \in \mathcal{H}$ , we have

$$F_{\text{opt}}Bx = \sum_{i=0}^{r-1} \sigma_i \langle B^+ Bx, e_i \rangle f_i = \sum_{i=0}^{r-1} \sigma_i \langle Px, e_i \rangle f_i$$

$$= \sum_{i=0}^{r-1} \sigma_i \langle x, P e_i \rangle f_i = \sum_{i=0}^{r-1} \sigma_i \langle x, e_i \rangle f_i \quad (4.59)$$

which is the best rank- $r$  approximation of  $AP$ . Clearly if  $r = \text{rank } AP$ , then  $AP - FB = 0$ . Finally, the rank- $r$  minimizing solution of problem  $\min_F \|AP - FB\|_{HS}$  is also the solution of the problem  $\min_F \|A - FB\|_{HS}$  (see (4.19)).

Since,

$$F_{\text{opt}} x = \sum_{i=0}^{r-1} \sigma_i \langle B^+ x, Q e_i \rangle f_i = \sum_{i=0}^{r-1} \sigma_i \langle Q B^+ x, e_i \rangle f_i \quad (4.60)$$

Note that

$$Q B^+ x = \sum_{i=0}^{r-1} \langle B^+ x, e_i \rangle e_i \quad x \in \mathcal{H}_1$$

Now, the proof of boundedness of the operator  $Q B^+$  is similar to the proof of boundedness of the operator  $F_{\text{opt}}$ .  $\square$

*Proof of Corollary 4.9.4.* The rank- $r$  minimizing solution of problem  $\min_F \|A - FB\|_{HS}$  is also the solution of the problem  $\min_F \|AP - FB\|_{HS}$  (see (4.19)). Using (4.23) and (4.24), for any  $x \in \mathcal{H}$ , we have

$$F_{\text{opt}} B x = \sum_{i=0}^{r-1} \sigma_i \langle x, g_{n(i)} \rangle f_i = \sum_{i=0}^{r-1} \sigma_i \langle x, e_i \rangle f_i$$

which is the best rank- $r$  approximation of  $AP$ . Since  $\beta_i \neq 0$ ,  $F_{\text{opt}}$  is a bounded operator.  $\square$

*Proof of Corollary 4.9.5.* Using an SVD of  $AP$  given in (4.20) and SVD of  $F_{\text{opt}} B$  given in (4.59), we get

$$\begin{aligned} \|AP - F_{\text{opt}} B\|_{HS}^2 &= \sum_{i \geq r} \sigma_i^2 = \sum_{i \in \mathbb{N}_{AP}} \sigma_i^2 - \sum_{i=0}^{r-1} \sigma_i^2 \\ &= \|AP\|_{HS}^2 - \|F_{\text{opt}} B\|_{HS}^2 \end{aligned}$$

for any  $r < \text{rank } AP$ . Also, we have  $\|A\|_{HS}^2 = \|A_N\|_{HS}^2 + \|AP\|_{HS}^2$  which follows from (4.19) when we take  $F = 0$ . Again from (4.19),

$$\begin{aligned} \|A - F_{\text{opt}} B\|_{HS}^2 &= \|A_N\|_{HS}^2 + \|AP - F_{\text{opt}} B\|_{HS}^2 \\ &= \|A_N\|_{HS}^2 + \|AP\|_{HS}^2 - \|F_{\text{opt}} B\|_{HS}^2 \\ &= \|A\|_{HS}^2 - \|F_{\text{opt}} B\|_{HS}^2 \end{aligned}$$

$\square$

*Proof of Lemma 4.9.6.* An SVD (modulo ordering) of the operator  $\check{G}_v$  mapping  $L^2[0, h)$  to  $L^2[0, h)$ , at almost all  $\theta$  is given by (see [31]).

$$\check{G}_v \check{w} = \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \langle \check{w}, \check{e}_i \rangle \check{e}_i$$

Using (4.25), for any  $\check{w} \in L^2[0, h)$  we have

$$\begin{aligned} \check{G}_v \check{P} \check{w} &= \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \langle \check{P} \check{w}, \check{e}_i \rangle \check{e}_i \\ &= \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \langle \check{w}, \check{P} \check{e}_i \rangle \check{e}_i \\ &= \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \left\langle \check{w}, \sum_{k \in \mathbb{M}, \alpha_k \neq 0} \langle \check{e}_i, \check{p}_k \rangle \check{p}_k \right\rangle \check{e}_i \\ &= \sum_{i \in \mathbb{Z}} \sum_{n=0}^{M-1} G_v(j\omega_{n+Mi}) \left\langle \check{w}, \sum_{k \in \mathbb{M}, \alpha_k \neq 0} \langle \check{e}_{n+Mi}, \check{p}_k \rangle \check{p}_k \right\rangle \check{e}_{n+Mi} \\ &= \sum_{i \in \mathbb{Z}} \sum_{n \in \mathbb{M}, \alpha_n \neq 0} G_v(j\omega_{n+Mi}) \langle \check{e}_{n+Mi}, \check{p}_n \rangle^* \langle \check{w}, \check{p}_n \rangle \check{e}_{n+Mi} \end{aligned}$$

Since  $\langle \check{e}_{n+Mi}, \check{p}_n \rangle^* = \frac{\psi_y^*(j\omega_{n+Mi})}{\alpha_n}$ , we have

$$\begin{aligned} \check{G}_v \check{P} \check{w} &= \sum_{i \in \mathbb{Z}} \sum_{n \in \mathbb{M}, \alpha_n \neq 0} \frac{G_v(j\omega_{n+Mi}) \psi_y^*(j\omega_{n+Mi})}{\alpha_n} \langle \check{w}, \check{p}_n \rangle \check{e}_{n+Mi} \\ &= \sum_{n \in \mathbb{M}, \alpha_n \neq 0} \langle \check{w}, \check{p}_n \rangle \sum_{i \in \mathbb{Z}} \frac{G_v(j\omega_{n+Mi}) \psi_y^*(j\omega_{n+Mi})}{\alpha_n} \check{e}_{n+Mi} \\ &= \sum_{n \in \mathbb{M}, \alpha_n \neq 0} \sigma_n \langle \check{w}, \check{p}_n \rangle \check{q}_n \end{aligned} \tag{4.61}$$

Since  $\psi_y(t) \in L^2$  and  $G_v \in L^2$ , we have  $\alpha_k, \sigma_n < \infty$  for almost all  $\theta$ , this again implies  $\check{q}_n \in L^2[0, h)$  for almost all  $\theta$ . Note that still, for almost all  $\theta$   $\langle \check{q}_i, \check{q}_k \rangle = \delta[i - k]$ .  $\square$

## 4.C Proofs of the results in Section 4.10

Throughout this section, we will use

$$A_N := A(I - P), \quad T_\gamma := I - \gamma^{-2} A_N A_N^* \quad \forall \gamma \in (\|A_N\|, \infty),$$

and  $\|\cdot\|$  means the induced 2-norm of the operator.

*Proof of Lemma 4.10.2.* Since  $(I - P)P = P(I - P) = 0$ , we have  $A_N(AP - FB)^* = A(I - P)((AP - FB)P)^* = A(I - P)P(AP - FB)^* = 0$  and  $(AP - FB)A_N^* = (A_N(AP - FB))^* = 0$ . Therefore

$$\begin{aligned} \|A - FB\| \leq \gamma &\Leftrightarrow (A - FB)(A - FB)^* \leq \gamma^2 I \\ &\Leftrightarrow (A_N + AP - FB)(A_N + AP - FB)^* \leq \gamma^2 I \\ &\Leftrightarrow A_N A_N^* + (AP - FB)(AP - FB)^* \leq \gamma^2 I \end{aligned}$$

Since operator  $A_N A_N^*$  and  $(AP - FB)(AP - FB)^*$  are positive, we have

$$\|A - FB\| \leq \gamma \Leftrightarrow (AP - FB)(AP - FB)^* \leq \gamma^2 T_\gamma$$

Here operator  $T_\gamma$  is positive (as  $\|A_N\| < \gamma$ ), therefore  $T_\gamma^{\frac{1}{2}}$  exists [47, Theorem 5.58]. Given  $\gamma^2 > \|A_N\|^2 = \|A_N A_N^*\|$ , then the operator  $T_\gamma$  is invertible [47, Theorem 4.40], which again implies  $T_\gamma^{\frac{1}{2}}$  is invertible. Now,

$$\begin{aligned} \|A - FB\| \leq \gamma &\Leftrightarrow T_\gamma^{-\frac{1}{2}}(AP - FB)(T_\gamma^{-\frac{1}{2}}(AP - FB))^* \leq \gamma^2 I \\ &\Leftrightarrow \|T_\gamma^{-\frac{1}{2}}(AP - FB)\| \leq \gamma \end{aligned}$$

□

*Proof of Lemma 4.10.5.* By Corollary 4.10.4 for a given  $i$ , the space  $\mathcal{V}_i$  is invariant under operator  $AA^*$ , hence  $P_i AA^* P_i = AA^* P_i$ . Also,  $\mathcal{V}_i$  is invariant under operator  $AP(AP)^*$  because if  $x \in \mathcal{V}_i$ , then

$$\begin{aligned} AP(AP)^* x &= \sum_{n \in \mathbb{N}_{AP}} \sigma_n^2 \langle x, f_n \rangle f_n = \sum_{n \in \mathbb{N}_{AP}} \sigma_n^2 \langle P_i x, f_n \rangle f_n \\ &= \sum_{n \in \mathbb{N}_{AP}} \sigma_n^2 \langle x, P_i f_n \rangle f_n = \sigma_i^2 \langle x, f_i \rangle f_i \end{aligned}$$

where last equation follows from the Assumption  $\mathcal{A}_1$  which says  $f_i \perp \mathcal{V}_j$  for all  $j \neq i$ . Since the space  $\mathcal{V}_i$  is invariant under operator  $AA^*$  and  $AP(AP)^*$ , this implies it is invariant under operator  $A_N A_N^* = AA^* - AP(AP)^*$ . Hence, if  $\gamma > \|A_N\|$ , then  $\mathcal{V}_i$  is invariant under operator  $T_\gamma$  and  $T_\gamma^{-1}$  (using Neumann series).

Note that  $(AP)^* T_\gamma^{-\frac{1}{2}} AP$  is a compact operator as  $AP$  is compact operator [47, Theorem 6.3]. Hence an SVD of the operator  $(AP)^* T_\gamma^{-\frac{1}{2}} AP$  exists [69, Eqn. 16.1]. First it will be shown that the eigenvectors of the operator  $(AP)^* T_\gamma^{-1} AP$  are in the set  $\{e_i\}_{i \in \mathbb{N}_{AP}}$ . To prove this, let us define  $y_i := T_\gamma^{-1} f_i$  where  $y_i \in \mathcal{V}_i$  (as  $\mathcal{V}_i$  is invariant under  $T_\gamma^{-1}$ ). Now considering,

$$T_\gamma^{-1} AP e_i = \sigma_i T_\gamma^{-1} f_i = \sigma_i y_i = \sigma_i (\langle y_i, f_i \rangle f_i + f_i^\perp)$$

where  $f_i^\perp \perp f_i$ . Since  $y_i$  and  $f_i$  belong to  $\mathcal{V}_i$ ,  $f_i^\perp \in \mathcal{V}_i$ . This implies  $f_i^\perp \perp f_j$  also for any  $j \neq i$ , as  $f_j \in \mathcal{V}_i^\perp$ . This means,

$$(AP)^* T_\gamma^{-1} AP e_i = \sigma_i^2 \langle y_i, f_i \rangle e_i$$

This implies  $e_i$  is an eigenvector of the operator  $(AP)^* T_\gamma^{-1} AP$ . Now it can be proved by contradiction that the eigenvectors of  $(AP)^* T_\gamma^{-1} AP$  with non-zero eigenvalues are in the set  $\{e_i\}_{i \in \mathbb{N}_{AP}}$ . Suppose there exist an eigenvector  $\hat{e} \notin \overline{\text{span}}\{e_i\}_{i \in \mathbb{N}_{AP}}$  of  $(AP)^* T_\gamma^{-1} AP$  with non-zero eigenvalue. This implies  $\hat{e} \in \overline{\text{Im}}((AP)^* T_\gamma^{-1} AP) = (\text{Ker}((AP)^* T_\gamma^{-1} AP))^\perp$  as  $(AP)^* T_\gamma^{-1} AP$  is self-adjoint ( $\overline{\text{Im}}$  means closure of image) [47, Equation 6.8]. However,

$$(\text{Ker}((AP)^* T_\gamma^{-1} AP))^\perp \subseteq (\text{Ker} AP)^\perp = \overline{\text{span}}\{e_i\}_{i \in \mathbb{N}_{AP}}$$

as  $\text{Ker} AP \subseteq \text{Ker}((AP)^* T_\gamma^{-1} AP)$ . This implies  $\hat{e} \in \overline{\text{span}}\{e_i\}_{i \in \mathbb{N}_{AP}}$ , which is a contradiction.

Now as shown above the eigenvalues of the operator  $(AP)^* T_\gamma^{-1} AP$  are given by  $\sigma_i^2 \langle y_i, f_i \rangle$  for all  $i \in \mathbb{N}_{AP}$ , but in terms of still unknown  $y_i$ . The rest of proof shows how to obtain  $y_i$  for all  $i \in \mathbb{N}_{AP}$ , hence the eigenvalues of the operator  $(AP)^* T_\gamma^{-1} AP$ .

Since the set  $\{v_{in}\}_{n \in \mathbb{N}_{P_i A}}$  forms an orthonormal basis of the space  $\mathcal{V}_i$  and  $f_i, y_i \in \mathcal{V}_i$ , we can expand  $f_i$  and  $y_i$  as

$$f_i = \sum_n c_{in} v_{in}, \quad y_i = \sum_n b_{in} v_{in}$$

where  $c_{in} := \langle f_i, v_{in} \rangle$  and  $b_{in} := \langle y_i, v_{in} \rangle$ . Note that  $c_{in} \neq 0$  by the definition of  $\mathcal{V}_i$ . Note that  $c_n \neq 0$  as  $\langle v_{in}, f_i \rangle \neq 0$  (see Lemma 4.10.3).

Now,

$$\begin{aligned} f_i &= T_\gamma y_i = (I - \gamma^{-2} A_N A_N^*) P_i y_i \\ &= \left( I - \gamma^{-2} A A^* P_i + \gamma^{-2} A P (A P)^* P_i \right) y_i \\ &= y_i - \gamma^{-2} \sum_n \alpha_{in}^2 \langle y_i, v_{in} \rangle v_{in} + \gamma^{-2} \sigma_i^2 \langle y_i, f_i \rangle f_i \end{aligned} \quad (4.62)$$

It will now be shown that  $y_i = \beta_i \sum_k c_{ik} \zeta_{ik} v_{ik}$  satisfy (4.62) where  $\zeta_{ik} = (1 - \gamma^{-2} \alpha_{ik}^2)^{-1}$  and  $\beta_i = (1 + \gamma^{-2} \sigma_i^2 \sum_n |c_{in}|^2 \zeta_{in})^{-1}$ . From (4.62), we have

$$\begin{aligned} T_\gamma y_i &= \left( y_i - \gamma^{-2} \sum_n \alpha_{in}^2 \langle y_i, v_{in} \rangle v_{in} \right) + \gamma^{-2} \sigma_i^2 \langle y_i, f_i \rangle f_i \\ &= \left( \beta_i \sum_n c_{in} \zeta_{in} v_{in} - \gamma^{-2} \sum_n \alpha_{in}^2 \beta_i c_{in} \zeta_{in} v_{in} \right) + \gamma^{-2} \sigma_i^2 \langle y_i, f_i \rangle f_i \\ &= \left( \beta_i \sum_n c_{in} \zeta_{in} (1 - \gamma^{-2} \alpha_{in}^2) v_{in} \right) + \gamma^{-2} \sigma_i^2 \langle y_i, f_i \rangle f_i \end{aligned}$$



$$\begin{aligned}
&= \beta_i \sum_n c_{in} v_{in} + \beta_i \gamma^{-2} \sigma_i^2 \left( \sum_k |c_{ik}|^2 \zeta_{ik} \right) f_i \\
&= \beta_i f_i + \beta_i \gamma^{-2} \sigma_i^2 \left( \sum_k |c_{ik}|^2 \zeta_{ik} \right) f_i \\
&= \beta_i \left( 1 + \gamma^{-2} \sigma_i^2 \sum_k |c_{ik}|^2 \zeta_{ik} \right) f_i = f_i
\end{aligned}$$

This shows  $y_i = \beta_i \sum_k c_{ik} \zeta_{ik} v_{ik}$  satisfy (4.62). Hence  $\langle y_i, f_i \rangle = \beta_i \sum_n |c_{in}|^2 \zeta_{in}$ .

Therefore  $\eta_i^2(\gamma) = \sigma_i^2 \langle y_i, f_i \rangle = \sigma_i^2 \beta_i \sum_n |c_{in}|^2 \zeta_{in}$ . Since for  $\gamma > \|A_N\|$ ,  $T_\gamma$  is a positive operator and invertible,  $(AP)^* T_\gamma^{-1} AP$  is also a positive operator, this implies eigenvalues of  $(AP)^* T_\gamma^{-1} AP$  belongs to  $[0, \infty)$ .  $\square$

*Proof of Lemma 4.10.6.* First we calculate  $\gamma_F := \|A - FB\|$  for all  $F \in \mathfrak{F}$ . We define  $i_F := \arg \max_i \{\alpha_{i0}\}_{i \notin \mathfrak{C}}$ . The first eigen vector of  $P_{i_F} A A^* P_{i_F}$  is given by  $v_{i_F 0}$ . Since  $i_F \notin \mathfrak{C}$ , by Assumption  $\mathcal{A}_1$ ,  $v_{i_F 0} \perp \bigoplus_{i \in \mathfrak{C}} \mathcal{V}_i$  and  $\{f_i\}_{i \in \mathfrak{C}} \in \bigoplus_{i \in \mathfrak{C}} \mathcal{V}_i$ , therefore

$$\begin{aligned}
&\langle f_i, v_{i_F 0} \rangle = 0 \quad \forall i \in \mathfrak{C} \\
&\Rightarrow (FB)^* v_{i_F 0} = \sum_{i \in \mathfrak{C}} \sigma_i \langle v_{i_F 0}, f_i \rangle e_i = 0 \\
&\Rightarrow \|(A - FB)^* v_{i_F 0}\| = \|A^* P_{i_F} v_{i_F 0}\| = \alpha_{i_F 0} \\
&\Rightarrow \|(A - FB)^*\| \geq \alpha_{i_F 0} \quad \Rightarrow \quad \|(A - FB)\| \geq \alpha_{i_F 0}
\end{aligned}$$

This means  $\gamma_F \geq \alpha_{i_F 0}$ .

If  $\alpha_{i_F 0} > \|A_N\|$ , then the singular values of the operator  $T_{\alpha_{i_F 0}^{-\frac{1}{2}}}(AP - FB)$  are given by the set  $\{\eta_i(\alpha_{i_F 0})\}_{i \in \mathfrak{C}}$  as  $AP - FB = \sum_{i \notin \mathfrak{C}} \sigma_i \langle x, e_i \rangle f_i$ . For a given  $i \notin \mathfrak{C}$ , we have  $\alpha_{i_F 0} \geq \{\alpha_{i0}\} \geq \{\alpha_{ik}\}_{k \in \mathbb{N}_{P_i A}}$ , hence

$$\begin{aligned}
&\frac{1}{\sigma_i^2 \sum_k \frac{|(f_i, v_{ik})|^2}{\alpha_{i_F 0}^2 - \alpha_{ik}^2}} \geq 0, \quad \forall i \notin \mathfrak{C} \& k \in \mathbb{N}_{P_i A} \\
&\Rightarrow \left( 1 + \frac{1}{\sigma_i^2 \sum_k \frac{|(f_i, v_{ik})|^2}{\alpha_{i_F 0}^2 - \alpha_{ik}^2}} \right)^{-1} \leq 1, \quad \forall i \notin \mathfrak{C} \& k \in \mathbb{N}_{P_i A} \\
&\Rightarrow \left( \alpha_{i_F 0}^{-2} + \frac{1}{\sigma_i^2 \sum_k \frac{|(f_i, v_{ik})|^2}{1 - \alpha_{i_F 0}^{-2} \alpha_{ik}^2}} \right)^{-1} \leq \alpha_{i_F 0}^2, \quad \forall i \notin \mathfrak{C} \& k \in \mathbb{N}_{P_i A} \\
&\Rightarrow \eta_i(\alpha_{i_F 0}) \leq \alpha_{i_F 0}, \quad \forall i \notin \mathfrak{C}
\end{aligned}$$

Here, we used the assumption that  $\langle f_i, v_{i_F k} \rangle \neq 0$  which follows from the construction of  $\mathcal{V}_i$  in (4.32). The above inequality implies all singular value of  $T_{\alpha_{i_F 0}}^{-\frac{1}{2}}(AP - FB)$  are less than or equal to  $\alpha_{i_F 0}$ , hence  $\|T_{\alpha_{i_F 0}}^{-\frac{1}{2}}(AP - FB)\| \leq \alpha_{i_F 0}$  if  $\alpha_{i_F 0} > \|A_N\|$ . Now, if  $\alpha_{i_F 0} > \|A_N\|$ , then by Lemma 4.10.2

$$\|T_{\alpha_{i_F 0}}^{-\frac{1}{2}}(AP - FB)\| \leq \alpha_{i_F 0} \Rightarrow \|A - FB\| \leq \alpha_{i_F 0}$$

Therefore we can say that if  $\alpha_{i_F 0} > \|A_N\|$ , then  $\|A - FB\| = \alpha_{i_F 0}$ .

If  $\alpha_{i_F 0} \leq \|A_N\|$ , choose any  $\gamma_\epsilon := \|A_N\| + \epsilon$ , where  $\epsilon > 0$ . For a given  $i \notin \mathfrak{C}$ , we have  $\gamma_\epsilon > \alpha_{i_F 0} \geq \{\alpha_{i_0}\} \geq \{\alpha_{ik}\}_{k \in \mathbb{N}_{P_i A}}$ . Similar to  $\alpha_{i_F 0} > \|A_N\|$  case we can show that,

$$\eta_i(\gamma_\epsilon) \leq \gamma_\epsilon \quad \forall i \notin \mathfrak{C}$$

Therefore, by Lemma 4.10.2

$$\|T_{\gamma_\epsilon}^{-\frac{1}{2}}(AP - FB)\| \leq \gamma_\epsilon \Rightarrow \|A - FB\| \leq \gamma_\epsilon$$

Since  $\|A - FB\| \geq \|A_N\|$ , letting  $\epsilon \rightarrow 0$ , we can say that  $\|A - FB\| = \|A_N\|$ .

Combining all of the above, we can say that

$$\gamma_F = \|A - FB\| = \max(\alpha_{i_F 0}, \|A_N\|)$$

Note that if  $F \neq F_0$ , then  $i_F \in D_r(\mathfrak{S})$ . But if  $F = F_0$ , then  $i_F = i_0 \notin D_r(\mathfrak{S})$  where  $i_0 := \arg \max_i \{\alpha_{i_0}\}_{i \notin D_r(\mathfrak{S})}$ . Now it follows from definition of  $D_r(\mathfrak{S})$  that

$$\max(\alpha_{i_0}, \|A_N\|) \leq \max(\alpha_{i_F 0}, \|A_N\|) \quad \forall F \in \mathfrak{F}$$

Hence,

$$\|A - F_0 B\| \leq \|A - FB\| \quad \forall F \in \mathfrak{F}$$

□

*Proof of Theorem 4.10.7.* Lets define  $\gamma_F := \|A - FB\|$ ,  $\gamma_{\text{opt}} := \min_{F \in \mathbb{F}_r} \|A - FB\|$  and  $F_{\text{opt}} := \arg \min_{F \in \mathbb{F}_r} \|A - FB\|$  where  $\mathbb{F}_r$  denote the space of bounded operators of rank  $r$ .

Notice that if  $\overline{\text{span}}\{e_i\}_{i \in D_r(\mathfrak{S})} \subseteq \text{Im } B^* \oplus (\text{Im } B^*)^\perp$ ,  $F_0$  is bounded (See Theorem 4.9.2 for the proof), so  $F_0 \in \mathbb{F}_r$  also.

First we show that the rank- $r$  approximation of  $T_\gamma^{-\frac{1}{2}}AP$  for any  $\gamma > \|A_N\|$  belongs to the set  $\mathfrak{F}$  (defined in Lemma 4.10.6). By Lemma 4.10.5, if  $\gamma > \|A_N\|$  then the SVD of  $T_\gamma^{-\frac{1}{2}}AP$  exist with singular values  $\eta_i(\gamma)$  and right singular vectors  $e_i$  for all  $i \in \mathbb{N}_{AP}$ . Therefore, if  $\gamma > \|A_N\|$ , then  $T_\gamma^{-\frac{1}{2}}AP$  has a SVD of the form

$$T_\gamma^{-\frac{1}{2}}APx = \sum_{i \in \mathbb{N}_{AP}} \eta_{m_i}(\gamma) \langle x, e_{m_i} \rangle z_i \quad x \in \mathcal{H}, z_i \in \mathcal{H}_0$$

where  $m_i$  gives the re-ordering of in the summation such that the set  $\{\eta_{m_i}(\gamma)\}_{i \in \mathbb{Z}}$  is in descending order. This implies

$$T_\gamma^{-\frac{1}{2}} A P e_{m_i} = T_\gamma^{-\frac{1}{2}} \sigma_{m_i} f_{m_i} = \eta_{m_i}(\gamma) z_i$$

Now let us define an  $F \in \mathfrak{F}$  with  $\mathfrak{C} = \{m_i\}_{i=(0, \dots, r-1)}$ . Then,

$$\begin{aligned} T_\gamma^{-\frac{1}{2}} F B x &= \sum_{k \in \mathfrak{C}} \sigma_k \langle B^+ B x, e_k \rangle T_\gamma^{-\frac{1}{2}} f_k & x \in \mathcal{H} \\ &= \sum_{i=0}^{r-1} \langle x, e_{m_i} \rangle \eta_{m_i}(\gamma) z_i, \end{aligned}$$

which is indeed optimal rank- $r$  approximation of  $T_\gamma^{-\frac{1}{2}} A P$  for any  $\gamma > \|A_N\|$ . This shows that rank- $r$  optimal approximation of  $T_\gamma^{-\frac{1}{2}} A P$  for any  $\gamma > \|A_N\|$  must belong to the set  $\mathfrak{F}$ .

Now let us consider the case  $\gamma_{\text{opt}} = \|A_N\|$ . We will show that in this case  $\|A - F_0 B\| = \|A_N\|$ , hence we can take  $F_{\text{opt}} = F_0$ . Assume that  $\gamma_0 := \|A - F_0 B\| > \|A_N\|$ . By Lemma 4.10.2, for any  $\gamma \in (\|A_N\|, \gamma_0)$  we have

$$\gamma_{\text{opt}} = \|A - F_{\text{opt}} B\| \leq \gamma \Rightarrow \|T_\gamma^{-\frac{1}{2}} (A P - F_{\text{opt}} B)\| \leq \gamma.$$

This implies, for some  $F \in \mathfrak{F}$ ,  $\|T_\gamma^{-\frac{1}{2}} (A P - F B)\| \leq \|T_\gamma^{-\frac{1}{2}} (A P - F_{\text{opt}} B)\| \leq \gamma$  (due to an SVD of  $T_\gamma^{-\frac{1}{2}} A P$ ). This means for same  $F \in \mathfrak{F}$ ,

$$\begin{aligned} \|T_\gamma^{-\frac{1}{2}} (A P - F B)\| \leq \gamma &\Rightarrow \|A - F B\| \leq \gamma \\ &\Rightarrow \|A - F B\| < \gamma_0. \end{aligned}$$

But since  $\gamma_F \geq \gamma_0$  for any  $F \in \mathfrak{F}$  (see Lemma 4.10.6), this leads to a contradiction. This implies  $\gamma_0 \not> \|A_N\|$  i.e.  $\gamma_0 = \|A_N\|$ . Since  $F_0$  attains  $\|A_N\|$  (the lower bound for  $\|A - F B\|$  for all  $F \in \mathbb{F}_r$ ), therefore  $\gamma_{\text{opt}} = \gamma_0$  and we can select  $F_{\text{opt}} = F_0$ .

Now let us consider the case when  $\gamma_{\text{opt}} > \|A_N\|$ . By Lemma 4.10.2, we have

$$\|A - F_{\text{opt}} B\| \leq \gamma_{\text{opt}} \Rightarrow \|T_{\gamma_{\text{opt}}}^{-\frac{1}{2}} (A P - F_{\text{opt}} B)\| \leq \gamma_{\text{opt}}.$$

This implies, for some  $F \in \mathfrak{F}$ ,  $\|T_{\gamma_{\text{opt}}}^{-\frac{1}{2}} (A P - F B)\| \leq \|T_{\gamma_{\text{opt}}}^{-\frac{1}{2}} (A P - F_{\text{opt}} B)\| \leq \gamma_{\text{opt}}$  (due to an SVD of  $T_{\gamma_{\text{opt}}}^{-\frac{1}{2}} A P$ ). This means for same  $F \in \mathfrak{F}$ ,

$$\|T_{\gamma_{\text{opt}}}^{-\frac{1}{2}} (A P - F B)\| \leq \gamma_{\text{opt}} \Rightarrow \|A - F B\| \leq \gamma_{\text{opt}}.$$

This means  $\gamma_F \leq \gamma_{\text{opt}}$  but by definition  $\gamma_{\text{opt}} \leq \gamma_F$  for same  $F \in \mathfrak{F}$ . This implies  $\gamma_F = \gamma_{\text{opt}} > \|A_N\|$  for same  $F \in \mathfrak{F}$ . Since we know that  $\gamma_0 \leq \gamma_F$  for all  $F \in \mathfrak{F}$ , therefore  $\gamma_0 = \gamma_{\text{opt}}$  and we can select  $F_{\text{opt}} = F_0$ .  $\square$

## 4.D Proofs of the results in Section 4.11

*Proof of Lemma 4.11.1.* The discrete lifted transform of signal  $\bar{w}_c$  for  $m \in \mathbb{M}$ , is given by

$$\begin{aligned}
 \bar{w}_c(e^{j\theta}; m) &= \sum_{k \in \mathbb{Z}} \bar{w}_c[Mk + m] e^{-j\theta k} \\
 &= \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \bar{g}_n[Mk + m - l] \bar{w}_n[l] e^{-j\theta k} \\
 &= \sum_{k \in \mathbb{Z}} \sum_{i=0}^{M-1} \sum_{l \in \mathbb{Z}} \bar{g}_n[Mk + m - Ml - i] \bar{w}_n[Ml + i] e^{-j\theta k} \\
 &= \sum_{k \in \mathbb{Z}} \sum_{i=0}^{M-1} \sum_{l \in \mathbb{Z}} \bar{g}_n[M(k-l) + m - i] \bar{w}_n[Ml + i] e^{-j\theta k} \\
 &= \sum_{i=0}^{M-1} \sum_{k \in \mathbb{Z}} \bar{g}_n[Mk + m - i] e^{-j\theta k} \sum_l \bar{w}_n[Ml + i] e^{-j\theta l} \\
 &= \sum_{i=0}^{M-1} \bar{g}_n(e^{j\theta}, m - i) \bar{w}_n(e^{j\theta}, i)
 \end{aligned}$$

Using discrete key lifting formula defined in Corollary 4.5.5, at each  $\theta \in [-\pi, \pi]$  we have

$$\begin{aligned}
 \bar{w}_c(e^{j\theta}; m) &= \frac{1}{M} \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} \bar{G}_n(e^{j\omega_k h'}) e^{j\omega_k h' (m-i)} \bar{w}_n(e^{j\theta}, i) \\
 &= \frac{1}{M} \sum_{k=0}^{M-1} \bar{G}_n(e^{j\omega_k h'}) e^{j\omega_k h' m} \sum_{i=0}^{M-1} \bar{w}_n(e^{j\theta}, i) e^{j\omega_k h' (-i)} \\
 &= \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} \bar{G}_n(e^{j\omega_k h'}) e^{j\omega_k h' m} \langle \bar{w}_n, \bar{e}_k \rangle \\
 &= \sum_{k=0}^{M-1} \bar{G}_n(e^{j\omega_k h'}) \langle \bar{w}_n, \bar{e}_k \rangle \bar{e}_k(e^{j\theta}; m)
 \end{aligned}$$

Note that  $\bar{G}_n \in L^\infty$  implies the boundedness of all singular values of  $\check{G}_n$  at almost each  $\theta \in [-\pi, \pi]$ .  $\square$

*Proof of Lemma 4.11.3.*

$$\begin{aligned}
 \check{G}_B \check{w}_s &:= [\check{S}_y \quad \check{G}_n] \check{w}_s \\
 &= \sqrt{\frac{M}{h}} \sum_{k \in \mathbb{M}} \alpha_k \langle \check{w}, \check{p}_k \rangle \bar{e}_k + \sum_{i \in \mathbb{M}} \bar{G}_n(e^{j\omega_i h'}) \langle \bar{w}_n, \bar{e}_i \rangle \bar{e}_i
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{k \in \mathbb{M}} \left( \left\langle \check{w}, \sqrt{\frac{M}{h}} \alpha_k \check{p}_k \right\rangle + \left\langle \check{w}_n, \tilde{G}_n(e^{j\omega_k h'})^* \check{e}_k \right\rangle \right) \check{e}_k \\
&= \sum_{k \in \mathbb{M}} \left\langle \begin{bmatrix} \check{w} \\ \check{w}_n \end{bmatrix}, \begin{bmatrix} \sqrt{\frac{M}{h}} \alpha_k \check{p}_k \\ \tilde{G}_n(e^{j\omega_k h'})^* \check{e}_k \end{bmatrix} \right\rangle \check{e}_k \\
&= \sum_{k \in \mathbb{M}, \mu_k \neq 0} \mu_k \left\langle \begin{bmatrix} \check{w} \\ \check{w}_n \end{bmatrix}, \frac{1}{\mu_k} \begin{bmatrix} \sqrt{\frac{M}{h}} \alpha_k \check{p}_k \\ \tilde{G}_n(e^{j\omega_k h'})^* \check{e}_k \end{bmatrix} \right\rangle \check{e}_k
\end{aligned}$$

Note that  $\check{p}_{n,k}$  are mutually orthonormal. □

*Proof of Lemma 4.11.4.* Using (4.47) and

$$\left\langle \begin{bmatrix} \check{e}_{l+Mi} \\ 0 \end{bmatrix}, \check{p}_{n,k} \right\rangle = \begin{cases} \sqrt{\frac{M}{h}} \frac{\psi_y^*(j\omega_{l+Mi})}{\mu_l} & k = l \\ 0 & k \neq l \end{cases},$$

for any  $\check{w}_s \in L^2[0, h) \times \mathbb{C}^M$  we have

$$\begin{aligned}
\check{G}_A \check{P}_B \check{w}_s &= \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \left\langle \check{P}_B \check{w}_s, \begin{bmatrix} \check{e}_i \\ 0 \end{bmatrix} \right\rangle \check{e}_i \\
&= \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \left\langle \check{w}_s, \check{P}_B \begin{bmatrix} \check{e}_i \\ 0 \end{bmatrix} \right\rangle \check{e}_i \\
&= \sum_{i \in \mathbb{Z}} G_v(j\omega_i) \left\langle \check{w}_s, \sum_{k \in \mathbb{M}, \mu_k \neq 0} \left\langle \begin{bmatrix} \check{e}_i \\ 0 \end{bmatrix}, \check{p}_{n,k} \right\rangle \check{p}_{n,k} \right\rangle \check{e}_i \\
&= \sum_{i \in \mathbb{Z}} \sum_{l \in \mathbb{M}} G_v(j\omega_{l+Mi}) \left\langle \check{w}_s, \sum_{k \in \mathbb{M}, \mu_k \neq 0} \left\langle \begin{bmatrix} \check{e}_{l+Mi} \\ 0 \end{bmatrix}, \check{p}_{n,k} \right\rangle \check{p}_{n,k} \right\rangle \check{e}_{l+Mi} \\
&= \sum_{i \in \mathbb{Z}} \sum_{l \in \mathbb{M}, \mu_l \neq 0} \sqrt{\frac{M}{h}} \frac{G_v(j\omega_{l+Mi}) \psi_y^*(j\omega_{l+Mi})}{\mu_l} \langle \check{w}_s, \check{p}_{n,l} \rangle \check{e}_{l+Mi} \\
&= \sum_{l \in \mathbb{M}, \mu_l \neq 0} \langle \check{w}_s, \check{p}_{n,l} \rangle \sum_{i \in \mathbb{Z}} \sqrt{\frac{M}{h}} \frac{G_v(j\omega_{l+Mi}) \psi_y^*(j\omega_{l+Mi})}{\mu_l} \check{e}_{l+Mi} \\
&= \sum_{l \in \mathbb{M}, \mu_l \neq 0} \rho_l \langle \check{w}_s, \check{p}_{n,l} \rangle \sum_{i \in \mathbb{Z}} \sqrt{\frac{M}{h}} \frac{G_v(j\omega_{l+Mi}) \psi_y^*(j\omega_{l+Mi})}{\rho_l \mu_l} \check{e}_{l+Mi} \\
&= \sum_{l \in \mathbb{M}, \mu_l \neq 0} \rho_l \langle \check{w}_s, \check{p}_{n,l} \rangle \check{q}_l
\end{aligned}$$

□

*Proof of Lemma 4.11.12.* Since,

$$\| \check{G}_A - \check{F}_{n,\text{opt}} \check{G}_B \| = \| [\check{G}_v - \check{F}_{n,\text{opt}} \check{S}_y \quad -\check{F}_{n,\text{opt}} \check{G}_n] \|$$

$$\geq \|\check{G}_v - \check{F}_{n,\text{opt}}\check{S}_y\| \geq \|\check{G}_v - \check{F}_{\text{opt}}\check{S}_y\|$$

where last inequality follows from the fact that  $\check{F}_{\text{opt}}$  is optimal over all  $\check{F}$  of rank  $r$ . This implies  $\|\mathcal{G}_A - \mathcal{F}_{n,\text{opt}}\mathcal{G}_B\|_{L^\infty} \geq \|\mathcal{G}_v - \mathcal{F}_{\text{opt}}\mathcal{S}_y\|_{L^\infty}$ .

To proof second part of the theorem, lets assume  $\mathcal{F}_{\text{opt}}$  and  $\mathcal{F}_{n,\text{opt}}$  are of full rank  $M$ . This implies  $\check{G}_v\check{P} - \check{F}_{\text{opt}}\check{S}_y = 0$  and  $\check{G}_A\check{P}_B - \check{F}_{n,\text{opt}}\check{G}_B = 0$ . Now

$$\begin{aligned} \|\check{G}_A - \check{F}_{n,\text{opt}}\check{G}_B\| &= \left\| \begin{bmatrix} \check{G}_v\check{P}_B - \check{F}_{n,\text{opt}}\check{S}_y & \check{G}_A(I - \check{P}_B) \end{bmatrix} \right\| \\ &= \left\| \begin{bmatrix} 0 & \check{G}_A(I - \check{P}_B) \end{bmatrix} \right\| = \|\check{G}_A(I - \check{P}_B)\| \end{aligned}$$

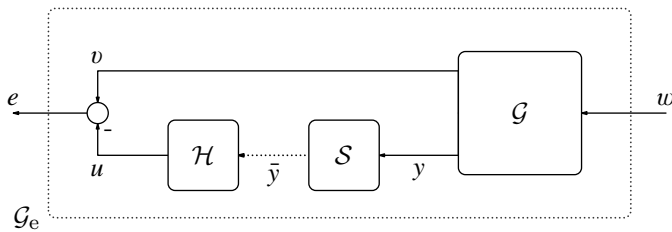
Similarly, for full rank  $\mathcal{F}_{\text{opt}}$ ,  $\|\check{G}_v - \check{F}_{\text{opt}}\check{S}_y\| = \|\check{G}_v(I - \check{P})\|$ . Since, for any rank  $\|\check{G}_A - \check{F}_{n,\text{opt}}\check{G}_B\| \geq \|\check{G}_v - \check{F}_{\text{opt}}\check{S}_y\|$ , then it must be true for full rank, this implies  $\|\check{G}_A(I - \check{P}_B)\| \geq \|\check{G}_v(I - \check{P})\|$ .  $\square$



# Chapter 5

## Relaxed causal sampling

### 5.1 Introduction



**Figure 5.1:** Sampled-data setup

In this chapter, we consider the problem of designing an optimal sampler  $\mathcal{S}$  given a hold  $\mathcal{H}$  and a signal generator  $\mathcal{G} := [\mathcal{G}_v \quad \mathcal{G}_y]^T$  (see Figure 5.1). A non-causal sampler can be obtained by using the solution provided in [31]. Ignoring causality is not a realistic scenario because most of the systems in practice are causal or relaxed causal in nature. Relaxed causal systems loosely speaking are systems whose present output depends not only on the present and the past inputs but also on a limited set of future of the input (see Section 2.5 for precise definition of relaxed causality). For example, assume that the input signal is stored in a compact disk. In this case, we can design a relaxed causal system for signal processing because some future inputs are available for processing. The maximum number of future inputs available depends upon the data storage capability of the compact disk. Live transmission of a marathon is another example. There we can allow a delay of few milliseconds in transmission i.e. what we see really on the television has happened few milliseconds in the past. We can design a relaxed causal system for signal processing because the data in those milliseconds can be considered as available future input. In summary, limited future inputs are sometimes available



for signal processing. For this reason in this chapter we concentrate on the design of relaxed-causal samplers instead of non-causal samplers, given causal hold  $\mathcal{H}$  and signal generator  $\mathcal{G}$ .

Throughout this chapter, we assume that the signal generator  $\mathcal{G}$  is known to us. The dual problem of designing an optimal relaxed causal hold  $\mathcal{H}$  given a sampler is studied and solved by [40] and [29]. The design of an optimal causal zero order hold  $\mathcal{H}$  given an ideal sampler  $\mathcal{S}$  (or vice-versa) is also well known (see [37, 38, 43]). All of these papers use the lifting technique to achieve the goal (see Chapter 2 for a review of lifting and lifting transforms). In this chapter we study the problem of designing optimal stable (see Section 2.6) relaxed causal samplers  $\mathcal{S}$  given causal  $\mathcal{H}$  and  $\mathcal{G}$ . This problem is similar to a two-sided model matching for LCTI systems (see [26]) but with the fundamental difference that systems involved here are required to be linear  $h$ -time shift invariant. Lifting also helps in obtaining a solution for this problem. We provide a frequency domain solution as well as a ready to use state space solution using the machinery given in [35]. This chapter is based on the papers [53] and [51].

The rest of the chapter is organized in three sections. In Section 5.2 we state our problem more precisely and provide a (lifted) frequency domain abstract solution. In Section 5.3, we review the fundamentals of the state-space for linear  $h$ -time shift invariant systems. In Section 5.4 a state space solution is provided to the problem.

## 5.2 Problem formulation and solution

In this section we formulate the problem of designing relaxed causal samplers given a hold and a model  $\mathcal{G} := [\mathcal{G}_v \quad \mathcal{G}_y]^T$ . We also provide a (lifted) frequency domain solution of this problem.

Recall that an analog signal  $y$  is *causal* if  $y(t) = 0 \forall t < 0$ , and for a given  $l \in \mathbb{N}$ , a sampler  $\mathcal{S}$  is (*relaxed*)  $l$ -*causal* if its output  $\bar{y}$  to any causal signal is  $l$ -causal (i.e.  $\bar{y}[n] = 0 \forall n < -l$ ).

Now, we state our problem more precisely:

**Problem  $\mathcal{P}_4$ :** *Given causal  $\mathcal{G}_v$  and  $\mathcal{G}_y$ , causal and stable hold  $\mathcal{H}$ , and  $l \in \mathbb{N}$ , find an  $l$ -causal and stable sampler  $\mathcal{S}$  such that  $\mathcal{G}_e := \mathcal{G}_v - \mathcal{H}\mathcal{S}\mathcal{G}_y$  is stable and  $\|\mathcal{G}_e\|_{L^2}$  is minimized.*

Intuitively, all the instabilities of  $\mathcal{G}_v$  must be contained in  $\mathcal{H}\mathcal{S}\mathcal{G}_y$  in order for  $\mathcal{G}_e$  to be stable. As  $\mathcal{H}$  and  $\mathcal{S}$  are stable, this implies that  $\mathcal{G}_v$  and  $\mathcal{G}_y$  must have the same type of instabilities. Moreover,  $\mathcal{H}$  must pass these instabilities. Therefore, the presence of a hold  $\mathcal{H}$  complicates the question of existence of a solution of  $\mathcal{P}_4$ . The complexity of  $\mathcal{P}_4$  is further increased as it is not immediately clear how the  $l$ -causality constraint can be imposed on the sampler. Similar to [40], lifting (and the lifted transform) can be used here to reduce some of these complexities. It is shown in Section 2.6 that if a system is stable then it belongs to  $L^\infty$  and if a system is  $l$ -causal and stable then it belongs to  $z^l H^\infty$ . Also, nothing can be said about the

space of causal  $\check{G}_v$  and  $\check{G}_y$  as they can be unstable. Therefore, problem  $\mathcal{P}_4$  in lifted  $z$ -domain is:

**Problem  $\mathcal{P}_5$**  : Given  $\check{G}_v$  and  $\check{G}_y$  are causal,  $\check{H} \in H^\infty$  and  $l \in \mathbb{N}$ , find  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e := \check{G}_v - \check{H}\check{S}\check{G}_y \in L^\infty \cap L^2$  and  $\|\check{G}_e\|_{L^2}$  is minimized.

**Remark 5.2.1.** If there exists a solution of Problem  $\mathcal{P}_5$  then  $\check{G}_e$  belongs to  $z^l H^\infty$ . This because  $\check{G}_v, \check{G}_y, \check{H}$  are causal and  $\check{S}$  is  $l$ -causal.

In order to solve  $\mathcal{P}_5$ , we break it into two parts according to the norm of  $\check{G}_e$ :

1. *Stabilization problem*: find all  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e := \check{G}_v - \check{H}\check{S}\check{G}_y \in L^\infty$ .
2. *Optimization problem*: find an  $\check{S} \in z^l H^\infty$  such that it solves the Stabilization problem and  $\|\check{G}_e\|_{L^2}$  is finite and minimized.

First, we consider the stabilization problem and after parameterizing all its solutions, we consider the optimization problem. For existence and parameterization of all the solutions of the *stabilization problem*, we need the following assumptions:

**Assumption  $\mathcal{A}_2$** :  $\check{G}_y$  is rational, proper and causal.

**Assumption  $\mathcal{A}_3$** : There exists a factorization of  $\check{H} = \check{H}_1 \check{H}_0$  with inner  $\check{H}_1 \in H^\infty$  (i.e.  $\check{H}_1^* \check{H}_1 = I$ ), and bistable and bicausal  $\check{H}_0 \in H^\infty$ .

The factorization in Assumption  $\mathcal{A}_3$  is an example of *inner-outer factorization* of hold  $\check{H}$  (see [62, §6.3] for details).

Assumption  $\mathcal{A}_2$  guarantees the existence of a coprime factorization of  $\check{G}_y$  over  $H^\infty$  (follows from [62, theorem 4.2.4], see also [4, chap. 2,8]).  $\check{N}_y$  and  $\check{M}_y$  are said to be left coprime factors in  $H^\infty$  of  $\check{G}_y$  if  $\check{N}_y$  and  $\check{M}_y$  are in  $H^\infty$ ,  $\check{G}_y = \check{M}_y^{-1} \check{N}_y$ , and there exist Bezout factors  $\check{X}_l \in H^\infty$  and  $\check{Y}_l \in H^\infty$  such that

$$\check{M}_y \check{X}_l + \check{N}_y \check{Y}_l = I$$

To have nice mathematical properties, the holds considered in this chapter are left invertible in  $L^\infty$ . Assumption  $\mathcal{A}_3$  implies left invertibility and stability of the hold. Assumption  $\mathcal{A}_3$  also helps in obtaining and parameterizing all the solutions of the stabilization problem as we will see later in this section.

Both of assumptions  $\mathcal{A}_2$  and  $\mathcal{A}_3$  are used in the following proposition which states the condition of existence of solutions of the stabilization problem.

**Proposition 5.2.2.** Given  $\check{H} \in H^\infty$ , causal  $\check{G}_v$  and  $\check{G}_y$ , and  $l \in \mathbb{N}$ . If assumptions  $\mathcal{A}_2$  and  $\mathcal{A}_3$  are satisfied, then there exists a sampler  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e \in L^\infty$  iff the following three conditions hold

1.  $\Pi_{\check{H}} \check{G}_v \in L^\infty$  where<sup>1</sup>  $\Pi_{\check{H}} := 1 - \check{H}_1 \check{H}_1^*$ ,

<sup>1</sup>This condition is with the constraint that  $\check{G}_v$  is causal.

2. there exists a coprime factorization over  $H^\infty$  of  $\check{G} := [\check{G}_v \quad \check{G}_y]^T$  of the form

$$\check{G} = \begin{bmatrix} I & \check{M}_v \\ 0 & \check{M}_y \end{bmatrix}^{-1} \begin{bmatrix} \check{N}_v \\ \check{N}_y \end{bmatrix} \quad (5.1)$$

with  $\check{M}_v, \check{N}_y$  left coprime.

3. there exists a  $\check{V} \in L^\infty$  such that  $\check{M}_h := \check{H}_i \check{M}_v - \check{V} \check{M}_y \in z^l H^\infty$ .

*Proof.* See Appendix 5.A (page 167).  $\square$

The above result can also be obtained by transforming the results of Kristalny [26] to the sampled-data setting. Condition 1 in Proposition 5.2.2 says that if an instability of  $\check{G}_v$  does not "belong" to the space  $\text{Im } \check{H}$  then we cannot cancel them by choice of  $\check{S}$ . Existence of a factorization of the form (5.1) in Condition 2 roughly speaking says that instabilities of  $\check{G}_v$  must be contained in  $\check{G}_y$ . These two conditions are sufficient and necessary to obtain a stable sampler  $\check{S}$  (i.e.  $\check{S} \in L^\infty$ ) such that  $\check{G}_e \in L^\infty$ . To obtain  $l$ -causal and stable sampler  $\check{S}$  (i.e.  $\check{S} \in z^l H^\infty$ ) we need an extra condition that there exists a  $\check{V} \in L^\infty$  such that  $\check{M}_h := \check{H}_i \check{M}_v - \check{V} \check{M}_y$  is in  $z^l H^\infty$  (Condition 3 in Proposition 5.2.2). There may exist several such  $\check{V}$ 's, so let us define the subspace  $\mathfrak{V} := \{\check{V} \in L^\infty : \check{M}_h \in z^l H^\infty\}$ . Now, we show that for any two  $\check{V}_1, \check{V}_2 \in \mathfrak{V} \subseteq L^\infty$ ,  $\text{proj}_{L^2 \setminus z^l H^2}(\check{V}_1 - \check{V}_2) = 0$ . This is used later in Proposition 5.2.4 to obtain a parameterization of all solutions of the stabilization problem in a single parameter. Note that  $\check{V}$  is a sampler, therefore, if it is in  $L^\infty$  then it is in  $L^2$  by Lemma 2.4.6. Hence, it makes sense to use the projection of a  $\check{V} \in \mathfrak{V}$ .

**Lemma 5.2.3.** *If  $\check{V}_1, \check{V}_2 \in L^\infty$  are such that  $\check{H}_i \check{M}_v - \check{V}_i \check{M}_y \in z^l H^\infty$  ( $i = 1, 2$ ) then,*

$$\text{proj}_{L^2 \setminus z^l H^2}(\check{V}_1 - \check{V}_2) = 0.$$

where  $\check{H}_i, \check{M}_v$  and  $\check{M}_y$  are defined in Proposition 5.2.2.

*Proof.* See Appendix 5.A (page 168).  $\square$

Lemma 5.2.3 is utilized in the following result. The proof is similar to the proof of [39, lemma 1].

**Proposition 5.2.4.** *If all the conditions of Proposition 5.2.2 are satisfied, then all samplers  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e := \check{G}_v - \check{H} \check{S} \check{G}_y \in L^\infty$  can be parameterized in parameter  $\check{S}_\alpha \in z^l H^\infty$  as*

$$\check{S} = \check{H}_0^{-1}(\check{S}_\alpha \check{M}_y - \check{M}_h) \quad (5.2)$$

where  $\check{M}_h := \check{H}_i \check{M}_v - \check{V} \check{M}_y$ . In this case

$$\check{G}_e = \check{G}_v + \check{H}_i \check{M}_h \check{G}_y - \check{H}_i \check{S}_\alpha \check{N}_y \quad (5.3)$$

*Proof.* See Appendix 5.A (page 168).  $\square$

After solving the *stabilization problem* in Proposition 5.2.4, we can now concentrate on the *optimization problem*. For this, we need the following assumption:

**Assumption  $\mathcal{A}_4$ :**  $\check{N}_y(e^{j\theta})\check{N}_y(e^{j\theta})^* > 0$  for all  $\theta \in [-\pi, \pi]$

Assumption  $\mathcal{A}_4$  along with Assumption  $\mathcal{A}_2$  is essential to make  $\check{N}_y$  co-inner (i.e.  $\check{N}_y\check{N}_y^\sim = I$ ) in (5.1). Now, we provide a solution to the Problem  $\mathcal{P}_5$  in the following lemma:

**Proposition 5.2.5.** *Let assumptions  $\mathcal{A}_2$ - $\mathcal{A}_4$  be satisfied. If the stabilization problem has a solution, then*

1. *there exist a coprime factorization over  $H^\infty$  of  $\check{G} := [\check{G}_v \quad \check{G}_y]^T$  of the form (5.1) with co-inner  $\check{N}_y$ .*
2.  *$\check{G}_e \in L^2$  iff  $\Pi_{\check{H}}\check{G}_v \in L^2$  where  $\Pi_{\check{H}} := 1 - \check{H}_i\check{H}_i^\sim$ .*

*In that case, there is a unique sampler that solves*

$$\check{S}_{\text{opt}} := \arg \inf_{\check{S} \in z^l H^\infty} \|\check{G}_v - \check{H}\check{S}\check{G}_y\|_{L^2} = \check{H}_0^{-1}(\check{S}_{\alpha,\text{opt}}\check{M}_y - \check{M}_h) \quad (5.4)$$

where

$$\check{S}_{\alpha,\text{opt}} = \text{proj}_{z^l H^2}(\check{H}_i^\sim \check{N}_v \check{N}_y^\sim - \check{V}). \quad (5.5)$$

Moreover,

$$\|\check{G}_{e,\text{opt}}\|_{L^2}^2 := \|\check{G}_v - \check{H}\check{S}_{\text{opt}}\check{G}_y\|_{L^2}^2 = \|\check{G}_v + \check{H}_i\check{M}_h\check{G}_y\|_{L^2}^2 - \|\check{S}_{\alpha,\text{opt}}\|_{L^2}^2 \quad (5.6)$$

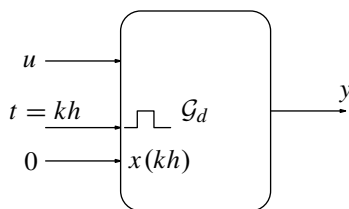
*Proof.* See Appendix 5.A (page 169).  $\square$

Note that  $\check{H}$  is a hold therefore we can never take  $\check{H} = I$ , in other words we will never have  $\Pi_{\check{H}} = 0$ .

Our aim in the rest of this chapter is to apply the results of Section 5.2 to a sampled-data setup where the signal generator and hold are given in state-space.

## 5.3 State-space of linear $h$ -time shift invariant systems

In this section, we define state-space representation of linear  $h$ -time shift invariant systems and list some of the properties that are useful in obtaining the optimal relaxed causal sampler. State-space representation of LCTI systems is well-known [71]. In order to understand the meaning of state-space for linear  $h$ -time shift invariant systems consider an example of reset system.



**Figure 5.2:** Reset system  $\mathcal{G}_d$  of Example 5.3.1. Here at time  $t = kh \forall k \in \mathbb{Z}$ , the output  $y$  and internal states are reset to zero irrespective of input  $u$ .

**Example 5.3.1.** Consider a linear  $h$ -time shift invariant reset system  $\mathcal{G}_d$  shown in Figure 5.2 whose output  $y$  is given by

$$y(t) = C \int_{kh}^t e^{A(t-s-kh)} B u(s) ds \quad kh \leq t < kh + h$$

where  $A$ ,  $B$  and  $C$  are complex matrices and  $k \in \mathbb{Z}$ . At time  $t = kh$ , the output  $y$  and internal states are reset to zero even if input  $u$  is non-zero.

Changing the variable  $t = kh + \tau$ , we have

$$y(kh + \tau) = C \int_0^\tau e^{A(\tau-\tau_1)} B u(kh + \tau_1) d\tau_1 \quad \tau \in [0, h). \quad (5.7)$$

We aim first to write a differential equation for the above. Using our state-space knowledge, we can express  $y(kh + \tau)$  in (5.7) as the solution of

$$\begin{aligned} \dot{x}(kh + \tau) &= Ax(kh + \tau) + Bu(kh + \tau), & x(kh) &= 0 \\ y(kh + \tau) &= Cx(kh + \tau). \end{aligned}$$

The above differential-equations resemble state-space with the difference that here states are allowed to jump at time instants  $kh$ . The condition  $x(kh) = 0$  says that whatever the value of  $x(kh^-)$  is,  $x(kh)$  is always zero. Taking lifted  $z$ -transform of  $x, u$  and  $y$ , we have

$$\dot{\check{x}}(z; \tau) = A\check{x}(z; \tau) + B\check{u}(z; \tau), \quad \tau \in [0, h) \quad (5.8a)$$

$$\check{y}(z; \tau) = C\check{x}(z; \tau) \quad (5.8b)$$

with (boundary) condition

$$\check{x}(z; 0) = 0. \quad (5.9)$$

The state-space equations (5.8) along with boundary condition (5.9) is an example of *state-space representation with two point boundary condition (STPBC)*

<sup>2</sup>This condition is with the constraint that  $\check{G}_v$  is causal.

[36, 25, 35]. Although boundary condition (5.9) consists of one point  $\check{x}(z; 0)$ , in general it is a relation between  $\check{x}(z; 0)$  and  $\check{x}(z; h^-)$ . Hence the name two point boundary condition. This point will be more clear if we take another example of a rational LCTI system.

**Example 5.3.2.** Consider a rational LCTI system  $\mathcal{G} : L^2(\mathbb{R}, \mathbb{R}^n) \rightarrow L^2(\mathbb{R}, \mathbb{R}^n)$  given in the state-space

$$\dot{x}(t) = Ax(t) + Bu(t), \quad t \in \mathbb{R} \quad (5.10a)$$

$$y(t) = Cx(t) + Du(t) \quad (5.10b)$$

where  $A, B, C$  and  $D$  are real matrices. For a given  $k \in \mathbb{N}$ , the lifted states  $\check{x}[k](\tau)$  satisfy the linear differential equation

$$\dot{\check{x}}[k](\tau) = A\check{x}[k](\tau) + B\check{u}[k](\tau) \quad \tau \in (0, h).$$

To obtain the initial condition  $\check{x}[k](0)$  for the (lifted) differential equation above we have to use an extra condition which tell us about the jump at  $kh$  i.e. a relationship between  $\check{x}[k](0)$  and  $\check{x}[k-1](h^-)$ . In case of the system  $\mathcal{G}$  is given by state-space (5.10), then  $x(t)$  is continuous for inputs  $u \in L^2$  [57, chapter 9]. Hence, we have the condition  $\check{x}[k](0) = \check{x}[k-1](h^-)$ . Therefore, the system  $\mathcal{G}$  in the lifted domain is given by

$$\begin{aligned} \dot{\check{x}}[k](\tau) &= A\check{x}[k](\tau) + B\check{u}[k](\tau) \\ \check{y}[k](\tau) &= C\check{x}[k](\tau) + D\check{u}[k](\tau) \end{aligned} \quad \tau \in (0, h)$$

with condition

$$\check{x}[k](0) = \check{x}[k-1](h^-).$$

Taking  $z$ -transform of the lifted  $\check{x}$ , we have

$$\begin{aligned} \dot{\check{x}}(z; \tau) &= A\check{x}(z; \tau) + B\check{u}(z; \tau) \\ \check{y}(z; \tau) &= C\check{x}(z; \tau) + D\check{u}(z; \tau) \end{aligned}$$

with (boundary) condition

$$\check{x}(z; 0) = \frac{1}{z}\check{x}(z; h^-)$$

Example 5.3.2 shows that a rational LCTI system  $\mathcal{G}$  is also an example of state-space with two point boundary condition (STPBC).

As described in examples 5.3.1 and 5.3.2, we allow states to jump at time instants  $kh, k \in \mathbb{Z}$ . These jumps allow us to write not only LCTI systems but also a much more bigger class of system. Later on we show that STPBC can be written for holds and samplers also.

### 5.3.1 State-space with two point boundary condition (STPBC)

Mirkin introduced STPBC for lifted systems in [36] using the earlier work of Krener [25], and Gohberg and Kaashoek [18]. In this section, we define STPBC and summarize some of its properties. This section is for reference purpose only.

We consider in this section a system  $\mathcal{G}$  mapping  $u \in L^2[0, h)$  to  $y \in L^2[0, h)$  defined by the linear differential equations as

$$\dot{x}(\tau) = Ax(\tau) + B(\tau)u(\tau) \quad (5.11a)$$

$$y(\tau) = C(\tau)x(\tau) + Du(\tau) \quad (5.11b)$$

with boundary condition

$$\Omega x(0) + \Upsilon x(h^-) = 0$$

where  $\tau \in [0, h)$ . Here for integers  $k, m$  and  $n$ ,  $A, \Omega, \Upsilon \in \mathbb{C}^{n \times n}$ ,  $D \in \mathbb{C}^{k \times m}$ ,  $B \in L^2([0, h), \mathbb{C}^{n \times m})$  and  $C \in L^2([0, h), \mathbb{C}^{k \times n})$ . The above representation of systems is known as *state-space with two point boundary condition (STPBC)*.  $D$  is known as *direct feed-through term* of the STPBC representation of the system  $\mathcal{G}$  (or in short *direct feed-through term* of the system  $\mathcal{G}$ ).

The system  $\mathcal{G}$  given by (5.11) is represented by the following notation in this thesis

$$y = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \begin{bmatrix} \Omega & \Upsilon \end{bmatrix} u \quad (5.12)$$

The usefulness of the STPBC representation is already established in the [36, 35, 40, 29].

These linear differential equations are defined *well-posed* if the output  $y$  is uniquely determined by the input  $u$  [25, 18]. It is shown in [25, 18] that well-posedness is equivalent to invertibility of the matrix

$$\Xi_{\mathcal{G}} := \Omega + \Upsilon e^{Ah}.$$

A condition for invertibility of  $\Xi_{\mathcal{G}}$  is stated in the following corollary.

**Corollary 5.3.3.** *A necessary (but not sufficient) condition for invertibility of matrix  $\Xi_{\mathcal{G}} := \Omega + \Upsilon e^{Ah}$  is that  $\begin{bmatrix} \Omega & \Upsilon \end{bmatrix}$  has full row rank.*

If (5.11) is well posed then the output  $y$  is given by

$$y(\tau) = Du(\tau) + \int_0^h K_{\mathcal{G}}(\tau, \sigma)u(\sigma)d\sigma \quad (5.13)$$

where

$$K_{\mathcal{G}}(\tau, \sigma) = \begin{cases} C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Omega e^{-A\sigma}B(\sigma) & \text{if } 0 \leq \sigma < \tau \leq h \\ -C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Upsilon e^{A(h-\sigma)}B(\sigma) & \text{if } 0 \leq \tau < \sigma \leq h \end{cases} \quad (5.14)$$

Alternatively, the output  $y(\tau)$  can be written as

$$\begin{aligned} y(\tau) &= Du(\tau) - C(\tau) \int_{\tau}^h e^{A(\tau-\sigma)} B(\sigma) u(\sigma) d\sigma \\ &\quad + C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} \Omega \int_0^h e^{-A\sigma} B(\sigma) u(\sigma) d\sigma \end{aligned} \quad (5.15a)$$

$$\begin{aligned} &= Du(\tau) + C(\tau) \int_0^{\tau} e^{A(\tau-\sigma)} B(\sigma) u(\sigma) d\sigma \\ &\quad - C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} \Upsilon \int_0^h e^{A(h-\sigma)} B(\sigma) u(\sigma) d\sigma. \end{aligned} \quad (5.15b)$$

The  $y(\tau)$  given in (5.15) is sometimes more useful than  $y(\tau)$  given in (5.13). The proof of (5.15) is given in Appendix 5.B (page 170).

We say that two STPBCs  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are *equivalent* if for the same input, the output of both systems are equal in  $L^2$  sense. In other words, the systems  $\mathcal{G}_1$  and  $\mathcal{G}_2$  given by

$$\mathcal{G}_i = \left[ \begin{array}{c|c} A_i & B_i \\ \hline C_i & D_i \end{array} \right] \begin{bmatrix} \Omega_i & \Upsilon_i \end{bmatrix} \quad (5.16)$$

where  $i = \{1, 2\}$ , are equivalent iff  $D_1 = D_2$  and  $K_{\mathcal{G}_1}(\tau, \sigma) = K_{\mathcal{G}_2}(\tau, \sigma)$ . The following corollary is immediate from the (5.13).

**Corollary 5.3.4.** *Suppose that the system  $\mathcal{G}$  given by STPBC (5.12) is well posed. Define  $\mathcal{G}_1$  with STPBC*

$$\mathcal{G}_1 = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \begin{bmatrix} S\Omega & S\Upsilon \end{bmatrix}$$

where  $S$  is a real matrix. If  $S$  is an invertible matrix then

$$\mathcal{G} \equiv \mathcal{G}_1$$

*Proof.* As feed-through matrix  $D$  is same,  $\mathcal{G} \equiv \mathcal{G}_1$  iff  $K_{\mathcal{G}}(\tau, \sigma) = K_{\mathcal{G}_1}(\tau, \sigma)$ . This follows from the fact that

$$\begin{aligned} C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} \Omega e^{-A\sigma} B(\sigma) &= C(\tau) e^{A\tau} (S \Xi_{\mathcal{G}})^{-1} S \Omega e^{-A\sigma} B(\sigma) \\ -C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} \Upsilon e^{A(h-\sigma)} B &= -C(\tau) e^{A\tau} (S \Xi_{\mathcal{G}})^{-1} S \Upsilon e^{A(h-\sigma)} B(\sigma) \end{aligned}$$

□

Now we list some basic operations in terms of STPBC [36]:

**Lemma 5.3.5.** *Let  $\mathcal{G}$  and  $\mathcal{G}_i$  for  $i = \{1, 2\}$  be given by STPBC (5.12) and (5.16) respectively. If  $\mathcal{G}$  and  $\mathcal{G}_i$  are well-posed, then*



## 1. Similarity transformation

$$\left[ \begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array} \right]_{\left[ \begin{array}{cc} S\Omega T^{-1} & S\Upsilon T^{-1} \end{array} \right]} \equiv \mathcal{G}$$

## 2. Parallel interconnection

$$\mathcal{G}_1 + \mathcal{G}_2 = \left[ \begin{array}{cc|c} A_1 & 0 & B_1 \\ 0 & A_2 & B_2 \\ \hline C_1 & C_2 & D_1 + D_2 \end{array} \right] \left[ \begin{array}{cc} \left[ \begin{array}{cc} \Omega_1 & 0 \\ 0 & \Omega_2 \end{array} \right] & \left[ \begin{array}{cc} \Upsilon_1 & 0 \\ 0 & \Upsilon_2 \end{array} \right] \end{array} \right]$$

## 3. Series interconnection

$$\mathcal{G}_1 \mathcal{G}_2 = \left[ \begin{array}{cc|c} A_1 & B_1 C_2 & B_1 D_2 \\ 0 & A_2 & B_2 \\ \hline C_1 & D_1 C_2 & D_1 D_2 \end{array} \right] \left[ \begin{array}{cc} \left[ \begin{array}{cc} \Omega_1 & 0 \\ 0 & \Omega_2 \end{array} \right] & \left[ \begin{array}{cc} \Upsilon_1 & 0 \\ 0 & \Upsilon_2 \end{array} \right] \end{array} \right]$$

All of these system are well-posed as well.

*Proof.* See [36]. □

**Lemma 5.3.6.** Suppose that the system  $\mathcal{G} : L^2[0, h) \rightarrow L^2[0, h)$  given by STPBC (5.12) is well posed. The adjoint system of  $\mathcal{G}$  is given by

$$\mathcal{G}^* = \left[ \begin{array}{c|c} -A^* & C^* \\ \hline -B^* & D^* \end{array} \right]_{\left[ \begin{array}{cc} \Upsilon_d^* & \Omega_d^* \end{array} \right]}$$

where  $\Upsilon_d$  and  $\Omega_d$  are such that  $\Omega \Upsilon_d = \Upsilon \Omega_d$  and  $\left[ \begin{array}{c} \Omega_d \\ \Upsilon_d \end{array} \right]$  has maximal column rank.

*Proof.* See Appendix 5.B (page 171). □

### 5.3.2 Systems in STPBC

In this section, we present STPBCs of different linear  $h$ -time shift invariant systems including LCTI systems, holds and samplers. Just like the state-space does not represent all LCTI systems, STPBCs do not represent all linear  $h$ -time shift invariant systems. However, it represents a fairly large class of systems including rational LCTI systems. With some modification, STPBCs can be used to represent holds and samplers also. A generic linear  $h$ -time shift invariant system  $y = \mathcal{G}u$  mapping  $L^2(\mathbb{R})$  to  $L^2(\mathbb{R})$  that can be represented by STPBC is given by

$$\check{y}(z) = \check{G}(z)\check{u}(z) : \quad \check{y}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]_{\left[ \begin{array}{cc} \Omega(z) & \Upsilon(z) \end{array} \right]} \check{u}(z), \quad (5.17)$$

in lifted  $z$ -domain. Here for integers  $k, m$  and  $n$ ,  $A \in \mathbb{C}^{n \times n}$ ,  $B \in L^2([0, h), \mathbb{C}^{n \times m})$ ,  $C \in L^2([0, h), \mathbb{C}^{k \times n})$ ,  $D \in \mathbb{C}^{k \times m}$ , and  $\Omega(z)$  and  $\Upsilon(z)$  are square discrete transfer matrices.  $D$  is known as *direct feed-through term* of the STPBC representation of system  $\check{G}$  (or in short, direct feed-through term of the system  $\check{G}$ ). The STPBC of  $\check{G}$  is well posed [35] if

$$\det(\Omega(z) + \Upsilon(z)e^{Ah}) \neq 0.$$

Those values of  $z \in \mathbb{C}$  for which the above does not hold are called *poles of the representation*. Using (5.15), poles of  $\check{G}(z)$  depend upon the factor  $(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Upsilon(z)$  or  $(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Omega(z)$ . Therefore, the common roots of  $\Omega(z)$  and  $\Upsilon(z)$  are the poles of the representation but not the poles of system  $\check{G}$ . Therefore, in STPBC representation, common roots of  $\Omega(z)$  and  $\Upsilon(z)$  can be removed without any loss of generality. Region of convergence do play a role in case of STPBC. For example, if our system  $\check{G}$  is causal and stable then all poles of the STPBC  $\check{G}(z)$  must lie within the unit disc  $\mathbb{D}$ .

### Rational LCTI system in STPBC

We start with an STPBC of a rational LCTI system.

**Lemma 5.3.7.** *Assume an LCTI system  $\mathcal{G} : L^2 \rightarrow L^2$  is given in state-space*

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t). \end{aligned}$$

where  $A, B, C$  and  $D$  are constant matrices. Then transfer function  $\check{G}(z)$  of the lifted system  $\check{G}$  can be written as

$$\check{G}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]_{\begin{bmatrix} zI & -I \end{bmatrix}} \quad (5.18)$$

*Proof.* See Example 5.3.2. □

### STPBC for Holds

A hold is a system which converts a discrete signal  $\bar{u} : \mathbb{Z} \rightarrow \mathbb{C}^{n_{\bar{u}}}$  back to an analog signal  $u : \mathbb{R} \rightarrow \mathbb{C}^{n_u}$ . Here  $n_{\bar{u}}$  and  $n_u$  are positive integers. If the hold is linear and  $h$ -time shift invariant then it is given by

$$u = \mathcal{H}\bar{u} : \quad u(t) = \sum_{n \in \mathbb{Z}} \phi(t - nh - \sigma) \bar{u}[n], \quad t \in \mathbb{R}$$

where  $\sigma \in (0, h)$  and  $\phi(t - \sigma)$  is the hold function. The hold given above can be viewed as a cascade of a modulated impulse train and an  $h$ -time shift invariant system  $\mathcal{G}$  with kernel  $g(t, s)$  such that  $g(t, nh + \sigma) = \phi(t - nh - \sigma)$  because

$$u(t) = \int_{-\infty}^{\infty} g(t, s) \sum_{n \in \mathbb{Z}} \delta(s - nh - \sigma) \bar{u}[n] ds = \sum_{n \in \mathbb{Z}} g(t, nh + \sigma) \bar{u}[n].$$

We assume that linear  $h$ -time shift invariant system  $\mathcal{G}$  has a well-posed STPBC representation in lifted  $z$ -domain as

$$\check{G}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} \Omega(z) & \Upsilon(z) \end{bmatrix} \quad (5.19)$$

Here for integers  $k, m$  and  $n$ ,  $A \in \mathbb{C}^{n \times n}$ ,  $D \in \mathbb{C}^{k \times m}$ ,  $B \in C^2([0, h], \mathbb{C}^{n \times m})$  (space of continuous functions in  $L^2([0, h], \mathbb{C}^{n \times m})$ ),  $C \in L^2([0, h], \mathbb{C}^{k \times n})$ , and  $\Omega(z)$  and  $\Upsilon(z)$  are square discrete transfer matrices. *Note that we assumed that  $B$  is continuous.* The feed through term  $D$  is assumed 0 in (5.19) so that we have a stable hold [57, chapter 9].

Using  $\check{G}(z)$  given in (5.19), the hold  $\mathcal{H}$  in the lifted  $z$ -domain is given by (see [35, 36] also)

$$\check{H}(z) = \check{G}(z)\mathcal{J}_\sigma \quad (5.20)$$

where the *impulse operator*  $\mathcal{J}_\sigma$  is defined as

$$\mathcal{J}_\sigma \eta := \delta(\tau - \sigma) \eta \quad \eta \in \mathbb{C}^n, \tau, \sigma \in (0, h) \quad (5.21)$$

The function  $\delta(\tau)$  has a meaning only in the sense of the integral

$$g(\sigma) = \int_0^h g(\tau) \delta(\tau - \sigma) d\tau, \quad \sigma \in (0, h).$$

Here we always assume that  $g(\tau)$  is a continuous function on  $(0, h)$  to avoid mathematical subtleties. This assumption is satisfied in case of hold  $\check{H}$  given in (5.20) because here  $e^{A\sigma} B(\sigma)$  (see (5.15)) is assumed continuous.

For any  $g \in C^2(0, h)$  (space of continuous functions in  $L^2((0, h), \mathbb{C}^n)$ ), we have

$$\begin{aligned} \langle \mathcal{J}_\sigma \eta, g \rangle_{C^2(0, h)} &= \int_0^h g(\tau) \delta(\tau - \sigma) \eta d\tau \\ &= g(\sigma)^* \eta \\ &= \langle \eta, \mathcal{J}_\sigma^* g \rangle_{\mathbb{C}^n} \end{aligned}$$

where *sampling operator*  $\mathcal{J}_\sigma^*$  is defined as

$$\mathcal{J}_\sigma^* g := g(\sigma) \quad g \in C^2(0, h). \quad (5.22)$$

As long as  $g$  is continuous we can treat  $\mathcal{J}_\sigma^*$  as the adjoint of  $\mathcal{J}_\sigma$  and vice-versa. For a detailed discussion on the operators  $\mathcal{J}_\sigma^*$  see [36]. The operator  $\mathcal{J}_\sigma^*$  helps in writing an STPBC for samplers.

Using (5.15), the output  $\check{y}(z) = \check{H}(z)\bar{u}(z)$  is given by

$$\check{y}(z; \tau) = C(\tau) e^{A\tau} \left( \mathbb{1}_{[0, h]}(\tau - \sigma) - (\Omega(z) + \Upsilon(z) e^{Ah})^{-1} \Upsilon(z) \right) B(\sigma) \bar{u}(z) \quad (5.23)$$

where  $B(\sigma) \in \mathbb{C}^{n \times m}$  is the value of  $B$  evaluated at the given  $\sigma$ .

Let us consider few examples now.

**Example 5.3.8.** Suppose the hold  $\mathcal{H}$  is a cascade of a modulated impulse train and a causal rational LCTI system  $\mathcal{G}$  given in state-space as

$$\mathcal{G} = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]$$

where  $A$ ,  $B$  and  $C$  are constant complex matrices. Now, the STPBC of  $\check{\mathcal{G}}(z)$  can be obtained by (5.18). Therefore,  $\mathcal{H}$  in lifted  $z$ -domain is given by

$$\dot{\mathcal{H}}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]_{[zI \quad -I]} \mathcal{J}_{0^+}$$

Now, we consider the generalized zero order hold discussed in [40].

**Example 5.3.9.** A generalized zero order hold  $\mathcal{H}_z : \mathbb{C}^n \rightarrow L^2$  is a hold whose hold function has support on  $[0, h)$ . Here we consider an example of generalized zero order hold  $\mathcal{H}_z$  given by (see [57, chapter 10])

$$\dot{x}(kh + \tau) = Ax(kh + \tau) + B\bar{u}[k], \quad k \in \mathbb{Z}, \tau \in [0, h) \quad (5.24a)$$

$$y(kh + \tau) = Cx(kh + \tau) \quad (5.24b)$$

with boundary condition

$$0 = x(kh) \quad (5.25)$$

where  $A$ ,  $B$ , and  $C$  are constant complex matrices.

Solving (5.24) with boundary condition (5.25), we have

$$y(kh + \tau) = Ce^{A\tau} B\bar{u}[k], \quad k \in \mathbb{Z}$$

Taking  $z$ -transform we have,

$$\check{y}(z; \tau) = Ce^{A\tau} B\check{u}(z)$$

with boundary condition

$$x(z; 0) = 0$$

The STPBC of the generalized zero order hold  $\mathcal{H}_z : \mathbb{C}^n \rightarrow L^2$  given by (5.24) with boundary condition (5.25) in lifted  $z$ -domain is given by

$$\dot{\mathcal{H}}_z(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]_{[I \quad 0]} \mathcal{J}_{0^+}$$

The ideal zero order hold is a special case of generalized zero order hold discussed in Example 5.3.9.

**Example 5.3.10.** A ideal zero order hold  $\mathcal{H}_{iz} : \mathbb{C}^n \rightarrow L^2$  is a hold whose hold function is the rectangular pulse  $\mathbb{1}_{[0,h)}(t)$ . The ideal zero order hold  $\mathcal{H}_{iz}$  is a special case of the generalized hold given by (5.24) i.e

$$y = \mathcal{H}_{iz}u \quad : \quad y(kh + \tau) = \bar{u}[k]$$

with boundary condition

$$0 = x(kh)$$

Therefore, the ideal zero order hold  $\mathcal{H}_{iz}$  in the lifted  $z$ -domain is given by

$$\dot{\mathcal{H}}_{iz}(z) = \left[ \begin{array}{c|c} 0 & I \\ \hline I & 0 \end{array} \right] \begin{bmatrix} \mathcal{J}_{0^+} \\ I & 0 \end{bmatrix}$$

### STPBC for samplers

A sampler  $\mathcal{S}$  is a system that maps an analog signal  $y : \mathbb{R} \rightarrow \mathbb{C}^{n_y}$  to a discrete signal  $\bar{y} : \mathbb{Z} \rightarrow \mathbb{C}^{n_{\bar{y}}}$ . Here  $n_y$  and  $n_{\bar{y}}$  are positive integers. If the sampler is linear and  $h$ -time shift invariant then it is given by

$$\bar{y} = \mathcal{S}y : \quad \bar{y}[n] = \int_{-\infty}^{\infty} \psi(nh + \sigma - s)y(s) ds$$

where  $\sigma \in (0, h)$  and  $\psi(t - \sigma)$  is the sampling function. The sampler given above can be viewed as a cascade of an  $h$ -time shift invariant system  $\mathcal{G}$  with kernel  $g(t, s)$  such that  $g(nh + \sigma, s) = \psi(nh + \sigma - s)$  and an ideal sampler sampling at every  $nh + \sigma$  time instants because

$$\bar{y}[n] = \int_{-\infty}^{\infty} g(nh + \sigma, s)y(s)ds.$$

We assume that the linear  $h$ -time shift invariant system  $\mathcal{G}$  has a well-posed STPBC representation in lifted  $z$ -domain as

$$\check{\mathcal{G}}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} \Omega(z) & \Upsilon(z) \end{bmatrix} \quad (5.26)$$

Here for integers  $k, m$  and  $n$ ,  $A \in \mathbb{C}^{n \times n}$ ,  $D \in \mathbb{C}^{k \times m}$ ,  $B \in L^2([0, h), \mathbb{C}^{n \times m})$ ,  $C \in C^2([0, h), \mathbb{C}^{k \times n})$ , and  $\Omega(z)$  and  $\Upsilon(z)$  are square discrete transfer matrices. Note that we assumed that  $C$  is continuous.

Using  $\check{\mathcal{G}}(z)$  given in (5.26), the sampler  $\mathcal{S}$  in the lifted  $z$ -domain is given by (see [35, 36] also)

$$\hat{\mathcal{S}}(z) = \mathcal{J}_{\sigma}^* \check{\mathcal{G}}(z), \quad \sigma \in (0, h)$$

where  $\mathcal{J}_{\sigma}^*$  is defined in (5.22).

We know that  $\mathcal{J}_\sigma^* g(\tau)$  has a meaning only if  $g(\tau)$  is continuous in  $(0, h)$ . This condition is satisfied in case of sampler  $\acute{S}$  because here  $C(\tau)e^{A\tau}$  (see (5.15)) is assumed continuous.

It is not necessary that the feed-through term (i.e.  $D$ ) for sampler in (5.26) is zero. However, if  $D \neq 0$  then the domain of sampler must be restricted to have a proper meaning of the sampling operator  $\mathcal{J}_\sigma^*$ . Then, for a given  $z$  and  $\sigma \in (0, h)$ ,

$$\acute{S}(z)\check{u}(z) = \mathcal{J}_\sigma^* \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]_{[\Omega \quad \Upsilon]} \check{u}(z), \quad \check{u}(z) \in L^2(0, h) \text{ \& continuous}$$

is well-defined. Here domain is restricted to continuous inputs only.

Let us consider few examples now. First, we take an example of ideal non-causal sampling.

**Example 5.3.11.** *The ideal non-causal sampler  $\mathcal{S}_{\text{idl}}$  is defined as*

$$\bar{y} = \mathcal{S}_{\text{idl}} u \quad : \quad \bar{y}[k] = u(kh^+)$$

where  $u(kh + \tau)$  is continuous in  $(0, h)$  for each  $k \in \mathbb{Z}$ . Taking lifted  $z$ -transform of  $\bar{y}$  and  $u$ , we have

$$\bar{y}(z) = \check{u}(z; 0^+) = \mathcal{J}_{0^+}^* \check{u}(z; \tau)$$

Hence, the ideal non-causal sampler in the lifted  $z$ -domain is given by

$$\acute{S}_{\text{idl}}(z) = \mathcal{J}_{0^+}^*$$

Now, we consider a cascade of an LCTI system with ideal non-causal sampler.

**Example 5.3.12.** *Given a sampler  $\mathcal{S}$  which is a cascade of a causal rational LCTI system  $\mathcal{G}$  given by state-space*

$$\mathcal{G} = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]$$

and the ideal non-causal sampler sampling with sampling period  $h$ . Here  $A$ ,  $B$  and  $C$  are constant complex matrices. Now, the STPBC of  $\acute{G}(z)$  can be obtained by (5.18). Therefore, it can be shown that  $\mathcal{S}$  in lifted  $z$ -domain is given by

$$\acute{S}(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]_{[zI \quad -I]}.$$

Now, we take an example of ideal causal sampling.

**Example 5.3.13.** *The ideal causal sampler  $\mathcal{S}_{\text{idl}}$  is defined as*

$$\bar{y} = \mathcal{S}_{\text{idl}} u \quad : \quad \bar{y}[k] = u(kh^-)$$

assuming  $u(kh + \tau)$  is continuous in  $(0, h)$  for each  $k \in \mathbb{Z}$ . Taking lifted  $z$ -transform of  $\bar{y}$  and  $u$ , we have

$$\bar{y}(z) = \check{u}(z; 0^-) = \frac{1}{z} \check{u}(z; h^-) = \frac{1}{z} \mathcal{J}_h^* \check{u}(z; \tau)$$

Hence, the ideal causal sampler in the lifted  $z$ -domain is given by

$$\check{S}_{\text{idl}}(z) = \frac{1}{z} \mathcal{J}_h^*$$

### Conjugate of the system given in STPBC

The conjugate of the system  $\check{G}(z)$  is defined as  $\check{G}^\sim(z) := [G(z^{-*})]^*$  [35]. By using Lemma 5.3.6, it can be shown that conjugate  $\check{G}^\sim(z)$  of the system  $\check{G}(z)$  given in (5.17) has STPBC (see [35] for details)

$$\check{G}^\sim(z) = \left[ \begin{array}{c|c} -A^* & C^* \\ \hline -B^* & D^* \end{array} \right] \begin{bmatrix} \Upsilon_d^\sim(z) & \Omega_d^\sim(z) \end{bmatrix} \quad (5.27)$$

where  $\Omega_d(z)$  and  $\Upsilon_d(z)$  are any square discrete transfer matrices satisfying

$$\Omega(z)\Upsilon_d(z) = \Upsilon(z)\Omega_d(z)$$

and such that  $\begin{bmatrix} \Omega_d(z) \\ \Upsilon_d(z) \end{bmatrix}$  has full normal rank.

It is clear from (5.15) that as long as a system  $\mathcal{G}$  has zero feed through term  $D$ , its output is continuous for a  $L^2$  input. This fact is used in the following Lemma which is useful in obtaining conjugates of samplers and holds.

**Lemma 5.3.14.** *Assume that for integers  $k, m$  and  $n$ ,  $A, \Omega, \Upsilon \in \mathbb{C}^{n \times n}$ ,  $D \in \mathbb{C}^{k \times m}$ ,  $B \in L^2([0, h], \mathbb{C}^{n \times m})$  and  $C \in L^2([0, h], \mathbb{C}^{k \times n})$ . Let system  $\mathcal{G}$  be given by STPBC*

$$\mathcal{G} = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} \Omega & \Upsilon \end{bmatrix}. \quad (5.28)$$

Now, for a given  $\sigma \in (0, h)$ ,

1. If  $B$  is continuous then  $(\mathcal{G}\mathcal{J}_\sigma)^* = \mathcal{J}_\sigma^* \mathcal{G}^*$ .
2. If  $C$  is continuous then  $(\mathcal{J}_\sigma^* \mathcal{G})^* = \mathcal{G}^* \mathcal{J}_\sigma$ .

*Proof.* See [36]. □

Using above Lemma 5.3.14, we obtain the conjugates of samplers and holds.

**Corollary 5.3.15.** *Given a  $\sigma \in (0, h)$ .*

1. Let system  $\check{G}$  is given by STPBC (5.19) with  $B$  continuous and hold  $\check{H} = \check{G}\mathcal{J}_\sigma$ . Then, the conjugate  $\check{H}^\sim$  of hold  $\check{H}$  is given by  $\check{H}^\sim = \mathcal{J}_\sigma^*\check{G}^\sim$ .
2. Let system  $\check{G}$  is given by STPBC (5.26) with  $C$  continuous and sampler  $\check{S} = \mathcal{J}_\sigma^*\check{G}$ . Then, the conjugate  $\check{S}^\sim$  of sampler  $\check{S}$  is given by  $\check{S}^\sim = \check{G}^\sim\mathcal{J}_\sigma$ .

*Proof.* Follows from Lemma 5.3.14. □

### 5.3.3 Stability and causality of systems given in STPBC

In this section, we describe the stability and causality condition of a system given in STPBC.

It is explained in Section 2.6 that if a system is stable and causal then the system is in  $H^\infty$ . We start with a simple case. Consider a linear  $h$ -time shift invariant analog system  $\mathcal{T}$  whose lifted impulse response system  $\check{T}[k]$  is zero at all  $k \in \mathbb{Z}$  except at  $k = 0$ . The kernel of  $\check{T}[0]$  can be a complicated function of time and may render a non- $L^2(\mathbb{R})$  output of the system  $\mathcal{T}$  to a  $L^2(\mathbb{R})$  input (i.e. the system may not be in  $L^\infty$ ). Therefore, the stability of such a system is not trivial and depends upon the kernel of  $\check{T}[0]$ . The following result helps in identifying the stability of all analog systems  $\mathcal{T}$  whose lifted impulse response system  $\check{T}[k]$  is zero at all  $k \in \mathbb{Z}$  except at  $k = 0$ .

**Lemma 5.3.16.** *Let  $g \in L^2[0, h) \times L^2[0, h)$  and  $0 \leq C < \infty$ . Let a linear  $h$ -time shift invariant system  $\mathcal{T}$  defined in lifted  $z$ -domain by*

$$\check{f} = \check{T}\check{u} : \quad \check{f}(z; \tau) = \int_0^h g(\tau, \sigma)\check{u}(z; \sigma)d\sigma, \quad \tau \in [0, h)$$

where

$$\begin{aligned} \text{ess sup}_{\tau \in [0, h)} \int_0^h |g(\tau, \sigma)|d\sigma &\leq C \\ \text{ess sup}_{\sigma \in [0, h)} \int_0^h |g(\tau, \sigma)|d\tau &\leq C. \end{aligned}$$

Now, we have that  $\mathcal{T}$  belongs to  $H^\infty$ .

*Proof.* See Appendix 5.B (page 173). □

The condition for stability and causality of an analog system given in STPBC is described in the following lemma.

**Lemma 5.3.17.** *For integers  $k, m$  and  $n$ , let  $A \in \mathbb{C}^{n \times n}$ ,  $D \in \mathbb{C}^{k \times m}$ ,  $B \in C^2([0, h), \mathbb{C}^{n \times m})$  and  $C \in C^2([0, h), \mathbb{C}^{k \times n})$ . Then, there exists matrices  $\bar{B}$  and  $\bar{C}$  that satisfy*

$$\bar{B}\bar{B}^* := \int_0^h e^{A(h-\sigma)}B(\sigma)B(\sigma)^*e^{A^*(h-\sigma)}d\sigma, \quad (5.29a)$$



$$\bar{C}^* \bar{C} := \int_0^h e^{A^* \tau} C^*(\tau) C(\tau) e^{A \tau} d\tau. \quad (5.29b)$$

Let  $\check{G}$  be a system with STPBC

$$\check{G}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \begin{bmatrix} \Omega(z) & \Upsilon(z) \end{bmatrix} \quad (5.30)$$

where  $\Omega(z)$  and  $\Upsilon(z)$  are square discrete transfer matrices. Then,

$$\check{G} \in H^\infty \iff \bar{Y} \in H^\infty$$

where  $\bar{Y}$  is a discrete system with

$$\bar{Y}(z) = \bar{C}(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Upsilon(z)\bar{B} \quad (5.31)$$

*Proof.* See Appendix 5.B (page 174).  $\square$

Similar to Lemma 5.3.17, the condition for stability and causality of holds and samplers can be stated.

**Lemma 5.3.18.** *Let  $A$ ,  $D$ ,  $B$ ,  $C$ ,  $\Omega(z)$  and  $\Upsilon(z)$  be as in Lemma 5.3.17. Let matrices  $\bar{B}$  and  $\bar{C}$  be such that they satisfy (5.29a) and (5.29b) respectively. Now,*

1. *if a hold  $\dot{H}$  is given by STPBC*

$$\dot{H}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} \Omega(z) & \Upsilon(z) \end{bmatrix} \mathcal{J}_r \quad (5.32)$$

*for some fixed  $r \in (0, h)$ , then*

$$\dot{H} \in H^\infty \iff \bar{Y}_H \in H^\infty$$

*where  $\bar{Y}_H$  is a discrete system with*

$$\bar{Y}_H(z) = \bar{C}(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Omega(z)e^{-Ar}B(r) \quad (5.33)$$

2. *if a sampler  $\acute{S}$  is given by STPBC*

$$\acute{S}(z) = \mathcal{J}_r^* \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} \Omega(z) & \Upsilon(z) \end{bmatrix} \quad (5.34)$$

*for some fixed  $r \in (0, h)$ , then*

$$\acute{S} \in H^\infty \iff \bar{Y}_S \in H^\infty$$

*where  $\bar{Y}_S$  is a discrete system with*

$$\bar{Y}_S(z) = C(r)e^{Ar}(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Upsilon(z)\bar{B} \quad (5.35)$$

*Proof.* Similar to the proof Lemma 5.3.17.  $\square$

**Remark 5.3.19.** Lemmas 5.3.16, 5.3.17 and 5.3.18 hold even if  $H^\infty$  is replaced by  $L^\infty$ . This will help in the cases where we are only interested in stability.

We can write a state space representation of the discrete system  $\bar{Y}$  in (5.31) more easily if  $\Omega(z) = zI$  and  $\Upsilon(z) = -\Upsilon \in \mathbb{C}^{n \times n}$ . In this case, we can apply our knowledge about state-space representation. The same can be said for discrete systems  $\bar{Y}_H$  in (5.33) and  $\bar{Y}_S$  in (5.35).

**Corollary 5.3.20.** If  $\Omega(z) = zI$  and  $\Upsilon(z) = -\Upsilon \in \mathbb{C}^{n \times n}$  in Lemma 5.3.17 and Lemma 5.3.18 then the system  $\check{G}$  given in (5.30), the hold  $\check{H}$  given by (5.32) and the sampler  $\check{S}$  (5.34) are in  $H^\infty$  if  $\Upsilon e^{Ah}$  is Schur (i.e. having eigenvalues in  $\mathbb{D}$ ).

*Proof.* The proof follows from Lemma 5.3.17 and Lemma 5.3.18.  $\square$

### 5.3.4 $H^2$ norm of systems given in STPBC

This section is devoted to the  $H^2$  norm of systems that are represented as STPBC. Similar to Section 5.3.3, we start with static systems.

**Lemma 5.3.21.** Let  $g \in L^2[0, h) \times L^2[0, h)$ . If an operator  $\mathcal{T} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  is given in lifted  $z$ -domain by

$$\check{f} = \check{T}\check{u} : \check{f}(z; \tau) = \int_0^h g(\tau, \sigma)\check{u}(z; \sigma)d\sigma, \quad \tau \in [0, h)$$

then  $\mathcal{T} \in H^2$ .

*Proof.* Since  $\mathcal{T}$  belongs to  $L^2$  (see [67, section 7.3] for a proof) and it is causal (see Lemma 2.5.8), we have that  $\mathcal{T} \in H^2$ .  $\square$

It is well known that if the direct feed through term of an analog system given in state-space is not zero, then that system does not have a finite  $H^2$  norm. The same can be said about systems represented by STPBC.

**Lemma 5.3.22.** Let  $\check{G}$  is a system with STPBC given by (5.30). Now, if  $D \neq 0$  then  $\check{G} \notin L^2$ .

*Proof.* The proof follows from the fact that constant multiplicative operators mapping  $L^2[0, h)$  to  $L^2[0, h)$  are not compact (hence not Hilbert-Schmidt).  $\square$

A result similar to Lemma 5.3.17 and Lemma 5.3.18 can be stated for the  $H^2$  norm of the systems. However, most of the systems given in this thesis are in  $H^\infty$  and for such systems we have the following simple result.

**Lemma 5.3.23.** *Let  $A, B, C, \Omega(z)$  and  $\Upsilon(z)$  as in Lemma 5.3.17. Let a system  $\check{G}$  in STPBC is given by*

$$\check{G}(z) := \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} \Omega(z) & -\Upsilon(z) \end{bmatrix}$$

Then  $\mathcal{G} \in H^\infty$  implies  $\mathcal{G} \in H^2$ .

*Proof.* Using (5.15), we have that

$$\check{G} = \check{X} - \check{Y}$$

where

$$\begin{aligned} \check{X}(z)\check{u}(z) &:= C(\tau) \int_0^\tau e^{A(\tau-\sigma)} B(\sigma)\check{u}(\sigma) d\sigma, \quad \tau \in [0, h) \\ \check{Y}(z)\check{u}(z) &:= C(\tau)e^{A\tau}(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Upsilon(z) \int_0^h e^{A(h-\sigma)} B(\sigma)\check{u}(\sigma) d\sigma. \end{aligned}$$

Since  $e^{A(\tau-\sigma)}\mathbf{1}(\tau-\sigma) \in L^2[0, h) \times L^2[0, h)$  and bounded, it follows that  $\check{X} \in H^\infty \cap H^2$  by lemmas 5.3.16 and 5.3.21. Since  $\check{Y}(z)$  is a hybrid signal processor (i.e. a cascade of sampler, discrete system and a hold), we have that  $\text{rank } \check{Y}(e^{j\theta})$  is uniformly bounded for all  $\theta \in [-\pi, \pi]$ . Therefore if  $\check{Y}$  in  $H^\infty$ , it is in  $H^2$  (see Lemma 2.4.6).  $\square$

**Remark 5.3.24.** *Lemma 5.3.23 holds even if  $H^\infty$  and  $H^2$  are replaced by  $L^\infty$  and  $L^2$  respectively.*

As shown in the above lemma, we check that the systems are in  $H^2$  or not without worrying much about  $\Omega(z)$  and  $\Upsilon(z)$ . However the calculation of the  $H^2$  norm depends upon the actual value of  $\Omega(z)$  and  $\Upsilon(z)$ . Therefore, in this section we calculate the  $H^2$  norm of systems that have  $\Omega(z) = zI$  and  $\Upsilon(z) = -\Upsilon \in \mathbb{C}^{n \times n}$ . Most of the systems discussed later in this chapter have these boundary conditions. These systems are such that their  $H^2$  norm is equal to the  $H^2$  norm of a discrete system.

**Lemma 5.3.25.** *For integers  $k, m$  and  $n$ , let  $A, \Upsilon \in \mathbb{C}^{n \times n}$ ,  $B \in C^2([0, h), \mathbb{C}^{n \times m})$  and  $C \in C^2([0, h), \mathbb{C}^{k \times n})$ . Let matrices  $\check{B}$  and  $\check{C}$  be such that they satisfy (5.29a) and (5.29b) respectively. Let  $\check{G}$  be a causal system with STPBC given by*

$$\check{G}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] \begin{bmatrix} zI & -\Upsilon \end{bmatrix}. \quad (5.36)$$

If  $\check{G} \in H^2$  then the squared  $H^2$  norm of  $\check{G}$  equals

$$\|\check{G}\|_{H^2}^2 = \frac{1}{h} \|\check{D}\|_{HS}^2 + \|\bar{Y}\|_{H^2}^2 \quad (5.37)$$

where  $\check{D} : L^2[0, h] \rightarrow L^2[0, h]$  is given by

$$\check{y} = \check{D}\check{u} : \check{y}(\tau) = \int_0^h C(\tau)e^{A(\tau-\sigma)}B(\sigma)\mathbb{1}(\tau-\sigma)\check{u}(\sigma)d\sigma \quad (5.38)$$

and  $\bar{Y}$  is a discrete system with

$$\bar{Y}(z) = \bar{C}(zI - \Upsilon e^{Ah})^{-1}\Upsilon\bar{B}$$

The squared Hilbert-Schmidt norm of  $\check{D}$  is given by

$$\|\check{D}\|_{HS}^2 = \text{tr} \int_0^h \int_0^h C(\tau)e^{A(\tau-\sigma)}B(\sigma)B(\sigma)^*e^{A^*(\tau-\sigma)}C^*(\tau)\mathbb{1}(\tau-\sigma)d\sigma d\tau \quad (5.39)$$

*Proof.* See Appendix 5.B (page 175). □

In a similar way, we can state the following result about  $H^2$  norm of holds and samplers.

**Lemma 5.3.26.** *Let  $A$ ,  $B$ ,  $C$  and  $\Upsilon$  as in Lemma 5.3.25. Let matrices  $\bar{B}$  and  $\bar{C}$  are such that they satisfy (5.29a) and (5.29b) respectively. Now,*

1. *if a causal hold  $\check{H}$  is given by STPBC*

$$\check{H}(z) = \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]_{[zI \quad -\Upsilon]} \mathcal{J}_{0^+}$$

*then*

$$\|\check{H}\|_{H^2} = \|\bar{Y}_H\|_{H^2}$$

*where  $\bar{Y}_H$  is a discrete system with*

$$\bar{Y}_H(z) = \bar{C}(zI - \Upsilon e^{Ah})^{-1}B(0^+)$$

2. *if a causal sampler  $\acute{S}$  is given by STPBC*

$$\acute{S}(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right]_{[zI \quad -\Upsilon]}$$

*then*

$$\|\acute{S}\|_{H^2} = \|\bar{Y}_S\|_{H^2}$$

*where  $\bar{Y}_S$  is a discrete system with*

$$\bar{Y}_S(z) = C(0)(zI - \Upsilon e^{Ah})^{-1}\Upsilon\bar{B}$$

*Proof.* The proof is similar to the proof Lemma 5.3.25. We also used

$$\|\bar{C}(zI - \Upsilon e^{Ah})^{-1}B(0^+)\|_{\mathbb{H}^2} = \|\bar{C}(zI - \Upsilon e^{Ah})^{-1}(zI)B(0^+)\|_{\mathbb{H}^2}$$

in the proof. □

**Remark 5.3.27.** *The adjoint of an anti-causal system is causal, therefore  $L^2$  norm of the anti-causal system can be calculated by using lemmas 5.3.25 and 5.3.26.*

To calculate the  $H^2$  norm of a discrete system the following standard result is very useful.

**Lemma 5.3.28.** *Let a causal discrete system  $\bar{G}$  be given by state space*

$$\bar{G} = \left( \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right)$$

where  $A$ ,  $B$ ,  $C$  and  $D$  are constant matrices. If  $A$  is Schur then

$$\|\bar{G}\|_{\mathbb{H}^2}^2 = \frac{1}{h}(\text{tr}(DD^*) + \text{tr}(C^*CW_c)) = \frac{1}{h}(\text{tr}(DD^*) + \text{tr}(W_bBB^*))$$

where

$$W_c = AW_cA^* + BB^*, \quad W_b = A^*W_bA + C^*C$$

*Proof.* Standard. □

Note that the standard  $H^2$  norm of a discrete system given in [71] is a scaled version of  $H^2$  norm defined in Section 2.4.3.

### 5.3.5 Computations

Integrals given in (5.29) and (5.39) seems very tedious to evaluate. However if  $B$  and  $C$  are constant then these integral can be calculated using matrix exponentials [28, 3, 36].

**Lemma 5.3.29.** *If  $A$  is a square constant matrix, and  $B$  and  $C$  are constant matrices of appropriate dimensions then*

$$\begin{aligned} \int_0^h e^{As}BB^*e^{A^*s}ds &= \Gamma_{33}^*(A, B)\Gamma_{23}(A, B) \\ \int_0^h \int_0^t e^{As}BB^*e^{A^*s}dsdt &= \Gamma_{33}^*(A, B)\Gamma_{13}(A, B) \\ \int_0^h e^{A^*s}C^*Ce^{As}ds &= \Lambda_{22}^*(A, C)\Lambda_{12}(A, C) \end{aligned}$$

where

$$\begin{aligned}\Gamma(A, B) &= \begin{bmatrix} \Gamma_{11}(A, B) & \Gamma_{12}(A, B) & \Gamma_{13}(A, B) \\ 0 & \Gamma_{22}(A, B) & \Gamma_{23}(A, B) \\ 0 & 0 & \Gamma_{33}(A, B) \end{bmatrix} \\ &= \exp \left\{ \begin{bmatrix} -A & I & 0 \\ 0 & -A & BB^* \\ 0 & 0 & A^* \end{bmatrix} h \right\}\end{aligned}$$

and

$$\Lambda(A, C) = \begin{bmatrix} \Lambda_{11}(A, C) & \Lambda_{12}(A, C) \\ 0 & \Lambda_{22}(A, C) \end{bmatrix} = \exp \left\{ \begin{bmatrix} -A^* & C^*C \\ 0 & A \end{bmatrix} h \right\}$$

*Proof.* The proof is given in [3]. □

### Other STPBC results

This section contains some lemmas which are useful in this chapter. Most of the results are from [36, 35] given here for reference purpose.

The following lemma explains the different methods of writing discrete systems mapping from  $\ell^2$  to  $\ell^2$  in STPBC.

**Lemma 5.3.30.** *Given a well-posed system  $\mathcal{G}$  with STPBC (5.28). Assume that  $B$  and  $C$  are continuous in (5.28). Then*

1.  $\mathcal{J}_h^* \mathcal{G} \mathcal{J}_{0^+} = C(h^-)e^{Ah}(\Omega + \Upsilon e^{Ah})^{-1}\Omega B(0^+)$
2.  $\mathcal{J}_{0^+}^* \mathcal{G} \mathcal{J}_h = -C(0^+)(\Omega + \Upsilon e^{Ah})^{-1}\Upsilon B(h^-)$

*In addition, if  $C(0^+)B(0^+) = 0$ , then*

3.  $\mathcal{J}_{0^+}^* \mathcal{G} \mathcal{J}_{0^+} = C(0^+)(\Omega + \Upsilon e^{Ah})^{-1}\Omega B(0^+)$

*and if  $C(h^-)B(h^-) = 0$ , then*

4.  $\mathcal{J}_h^* \mathcal{G} \mathcal{J}_h = -C(h^-)e^{Ah}(\Omega + \Upsilon e^{Ah})^{-1}\Upsilon B(h^-)$

*Proof.* Follows from (5.15). See [36] for detail. □

The following lemma explains how to write the impulse input system as a system without impulse input.

**Lemma 5.3.31.** *The well-posed system mapping  $\begin{bmatrix} u \\ \bar{\eta} \end{bmatrix} \in L^2[0, h) \times \mathbb{R}^n$  to  $y_1 \in L^2[0, h)$  given by STPBC*

$$\begin{aligned}\dot{x}_1 &= Ax_1 + Bu + B_\eta \mathcal{J}_{0^+} \bar{\eta}, & \Omega x_1(0) + \Upsilon x_1(h^-) &= 0 \\ y_1 &= Cx_1\end{aligned}$$

is equivalent (i.e. for same inputs  $u$  and  $\bar{\eta}$ ,  $y_1 = y_2$  in Lebesgue sense) to system

$$\begin{aligned}\dot{x}_2 &= Ax_2 + Bu, \quad \Omega(x_2(0) - B_\eta \bar{\eta}) + \Upsilon x_2(h^-) = 0 \\ y_2 &= Cx_2.\end{aligned}$$

*Proof.* See [35, Proposition A.2]. □

We can also represent a hybrid signal processor (i.e. a cascade of a sampler and a hold) in STPBC. The following lemma essentially from [36] helps here.

**Lemma 5.3.32.** *Given*

$$\mathcal{G}_1 := \left[ \begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & 0 \end{array} \right]_{[\Omega_1 \quad \Upsilon_1]} \mathcal{J}_{\lambda_1}, \quad \mathcal{G}_2 := \mathcal{J}_{\lambda_2}^* \left[ \begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & 0 \end{array} \right]_{[\Omega_2 \quad \Upsilon_2]}$$

where  $\lambda_i, i = 1, 2$  be either  $0^+$  or  $h^-$ ,  $A_i, \Omega_i, \Upsilon_i$  are square matrices, and  $B_i$  and  $C_i$  are matrix valued functions in  $C^2([0, h))$  of appropriate dimensions then the STPBC of system  $\mathcal{G}_f := \mathcal{G}_1 \mathcal{J}_{\lambda_1} \mathcal{J}_{\lambda_2}^* \mathcal{G}_2$  is given by

$$\mathcal{G}_f := \left[ \begin{array}{cc|c} A_1 & 0 & 0 \\ 0 & A_2 & B_2 \\ \hline C_1 & 0 & 0 \end{array} \right] \left[ \begin{array}{cc} [\Omega_1 & (1 - \lambda_2)M_1] \\ 0 & \Omega_2 \end{array} \right] \left[ \begin{array}{cc} [\Upsilon_1 & \lambda_2 M_1] \\ 0 & \Upsilon_2 \end{array} \right]$$

where  $M_1 := (\lambda_1 \Upsilon_1 - (1 - \lambda_1)\Omega_1)B_1C_2$ .

*Proof.* See [36, lemma 3]. □

## 5.4 STPBC solution

In this section we apply the results of Section 5.2 to a sampled-data setup where the signal generator and hold are given in state-space with two-point boundary condition (STPBC). We assume that signal generator  $\mathcal{G}$  is causal and in Laplace domain is given by,

$$G(s) = \begin{bmatrix} G_v(s) \\ G_y(s) \end{bmatrix} = D + C(sI - A)^{-1}B \quad (5.40)$$

where  $C := \begin{bmatrix} C_v \\ C_y \end{bmatrix}$  and  $D := \begin{bmatrix} 0 \\ D_y \end{bmatrix}$ . This is a LCTI system therefore the STPBC of  $\mathcal{G}$  in lifted  $z$ -domain is given by (see Lemma 5.3.7):

$$\check{G}(z) = \begin{bmatrix} \check{G}_v(z) \\ \check{G}_y(z) \end{bmatrix} = \left[ \begin{array}{c|c} A & B \\ \hline C_v & 0 \\ C_y & D_y \end{array} \right]_{[zI \quad -I]} \quad (5.41)$$

The reason of zero direct feed-through term of  $\check{G}_v$  is explained in Remark 5.4.1. Without loss of generality, we assume that  $A = \text{diag}\{A_s, A_u\}$ , where  $A_s$  has all its eigenvalues in  $\mathbb{C}^- := \{z \in \mathbb{C} : \text{real}(z) < 0\}$  and  $A_u$  has all its eigenvalues in  $\mathbb{C} \setminus \mathbb{C}^-$ . Also let  $C_v := \begin{bmatrix} C_{vs} & C_{vu} \end{bmatrix}$  be the partition of  $C_v$  according to  $A_s$  and  $A_u$ . Therefore

$$\check{G}(z) = \left[ \begin{array}{cc|c} A_s & 0 & B \\ 0 & A_u & \\ \hline C_{vs} & C_{vu} & 0 \\ & C_y & D_y \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix}. \quad (5.42)$$

Note that systems  $\mathcal{G}_v$  and  $\mathcal{G}_y$  are causal (but not necessarily stable) by assumption. However, to obtain an optimal sampler described in Proposition 5.2.5, in addition to causality, we need that  $\mathcal{G}_y$  must be rational and proper (Assumption  $\mathcal{A}_2$ ). Since  $\mathcal{G}_y$  is represented in state-space, it is rational and proper. Also, we need that there exists a left coprime-factorization of  $\check{G}$  of the form (5.1). To this end, we need the following assumption.

**Assumption  $\mathcal{A}_5$ :**  $(C_y, A)$  is observable and  $(A, B)$  is controllable.

Later it will be explained in Section 5.4.1 that  $\mathcal{A}_5$  allows the existence of a coprime factorization, and assumptions  $\mathcal{A}_4$  and  $\mathcal{A}_5$  allow the existence of a coprime factorization  $\check{G}_y = \check{M}_y^{-1} \check{N}_y$  with  $\check{N}_y$  co-inner.

Also, we consider hold  $\check{H}$  with STPBC given by [35]:

$$\check{H}(z) := \left[ \begin{array}{c|c} A_H & B_H \\ \hline C_H & 0 \end{array} \right] \begin{bmatrix} zI & -E \end{bmatrix} \mathcal{J}_{0^+} \quad (5.43)$$

where impulse operator  $\mathcal{J}_{0^+}$  defined in (5.21) is needed to perform the discrete to analog domain conversion. The holds given by STPBC (5.43) can represent a large class of stable holds with infinite or finite impulse response. For example the ideal zero order hold  $\check{H}_{iz}$  can be obtained by setting  $I = C_H = B_H$  and  $0 = A_H = E$ . We also assume the following about  $\check{H}$ :

**Assumption  $\mathcal{A}_6$ :**  $Ee^{A_H h}$  is a Schur matrix,

**Assumption  $\mathcal{A}_7$ :**  $B_H$  has full column rank.

**Assumption  $\mathcal{A}_8$ :**  $(C_H, A_H)$  is observable.

$Ee^{A_H h}$  is a Schur matrix is just a restatement of the fact that  $\check{H} \in H^\infty$  (see Corollary 5.3.20). Assumption  $\mathcal{A}_6$  allow us to obtain a right coprime factorization of  $\check{H}$  and assumptions  $\mathcal{A}_6$ - $\mathcal{A}_8$  allow us to obtain an inner-outer factorization of  $\check{H}$ . This is explained in Section 5.4.3.

**Remark 5.4.1.** If the direct feed-through term of  $\check{G}_v$  is not zero in (5.41) then STPBC of  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v$  will have a non-zero direct feed-through term. This implies  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \notin \mathcal{L}^2$  (see Lemma 5.3.22). Therefore the solution of Problem  $\mathcal{P}_5$  does not exist (see Proposition 5.2.5). Hence, the direct feed-through term of  $\check{G}_v$  is chosen zero in (5.41).



### 5.4.1 Left coprime-factorization of $\check{G}$

The following lemma describes the left coprime-factorization over  $H^\infty$  of  $\check{G}$  which is required in Proposition 5.2.5.

**Lemma 5.4.2.** *Let  $\begin{bmatrix} \check{G}_v \\ \check{G}_y \end{bmatrix}$  have STPBC given in (5.41). If Assumption  $\mathcal{A}_5$  is satisfied then there exists an  $L$  such that  $A + LC_y$  is Hurwitz. In that case*

$$\begin{bmatrix} \check{G}_v \\ \check{G}_y \end{bmatrix} = \begin{bmatrix} I & \check{M}_v \\ 0 & \check{M}_y \end{bmatrix}^{-1} \begin{bmatrix} \check{N}_v \\ \check{N}_y \end{bmatrix} \quad (5.44)$$

for

$$\begin{bmatrix} \check{M}_v & \check{N}_v \\ \check{M}_y & \check{N}_y \end{bmatrix} := \left[ \begin{array}{cc|cc} A + LC_y & L & B + LD_y & \\ \hline C_v & 0 & 0 & \\ Z_y C_y & Z_y & Z_y D_y & \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix} \quad (5.45)$$

where  $Z_y$  is any invertible complex matrix. In this case,  $\check{M}_y$  and  $\check{N}_y$  are left coprime, and  $\check{M}_v, \check{M}_y, \check{N}_y$  and  $\check{N}_v$  belong to  $H^\infty$ .

*Proof.* See Appendix 5.C (page 176). □

We also need that  $\check{N}_y$  to be co-inner in the Proposition 5.2.5. Here, Assumption  $\mathcal{A}_4$  helps. We start with the following standard result to check Assumption  $\mathcal{A}_4$  in the state space.

**Lemma 5.4.3.** *Let  $\check{N}_y$  be as in (5.45). If  $(C_y, A)$  is observable (see  $\mathcal{A}_5$ ) then Assumption  $\mathcal{A}_4$  ( $\check{N}_y(e^{j\theta})\check{N}_y^*(e^{j\theta}) > 0 \forall \theta \in [-\pi, \pi]$ ) is satisfied iff  $D_y$  has full row rank and*

$$\begin{bmatrix} A - j\omega I & B \\ C_y & D_y \end{bmatrix}$$

has full row rank for all  $\omega \in \mathbb{R}$ .

*Proof.* See Appendix 5.C (page 177). □

The following result explains how to do the left coprime factorization  $\check{G}_y = \check{M}_y^{-1}\check{N}_y$  with  $\check{N}_y$  co-inner.

**Lemma 5.4.4.** *If assumptions  $\mathcal{A}_4$  and  $\mathcal{A}_5$  are satisfied then by Lemma 5.4.2 there exists a coprime factorization of  $\check{G}$  given in (5.41) of the form (5.44),  $R := D_y D_y^*$  is invertible and there exists a unique stabilizing solution  $X$  (i.e. such that matrix  $A + -(XC_y^* + BD_y^*)R^{-1}C_y$  is Hurwitz) of the Riccati equation*

$$AX + XA^* - (XC_y^* + BD_y^*)R^{-1}(C_y X + D_y B^*) + BB^* = 0.$$

If we choose

$$\begin{aligned} Z_y &= R^{-\frac{1}{2}} \\ L &= -(XC_y^* + BD_y^*)R^{-1} \end{aligned}$$

then  $\check{N}_y$  defined in (5.45) is co-inner.

*Proof.* Follows from Lemma 5.4.3 and [71, Theorem 13.35].  $\square$

The following lemma is useful later in obtaining the optimal sampler.

**Lemma 5.4.5.** *Let  $L$ ,  $Z_y$  and  $X$  be as in Lemma 5.4.4, and  $\check{N}_v$  and  $\check{N}_y$  as in Lemma 5.4.2. Now,*

$$\check{N}_v \check{N}_y \sim \left[ \begin{array}{c|c} -(A + LC_y) & (Z_y C_y)^* \\ \hline -C_v X & 0 \end{array} \right]$$

*Proof.* See Appendix 5.C (page 177).  $\square$

## 5.4.2 Simplification of Assumption $\mathcal{A}_3$

It is desirable to have a simple criterion which tells us that Assumption  $\mathcal{A}_3$  (i.e. assumption of existence of an inner-outer factorization of  $\hat{H} = \hat{H}_i \bar{H}_o$  with inner  $\hat{H}_i \in H^\infty$ , and bistable and bicausal  $\bar{H}_o \in H^\infty$ ) is satisfied or not. To this end, using (5.23) we write the hold  $\hat{H}(z)$ , defined in (5.43), as

$$\begin{aligned} \hat{H}(z) &= \left[ \begin{array}{c|c} A_H & B_H \\ \hline C_H & 0 \end{array} \right] \begin{bmatrix} zI & -E \end{bmatrix} \mathcal{J}_{0^+} \\ &= C_H e^{A_H \tau} \left( I + (zI - E e^{A_H h})^{-1} E e^{A_H h} \right) B_H \\ &= C_H e^{A_H \tau} \bar{H}_s(z) \end{aligned} \quad (5.46)$$

where  $\bar{H}_s$  is a discrete system which is rational in  $z$  and it is given by

$$\bar{H}_s(z) := \left( \begin{array}{c|c} E e^{A_H h} & E e^{A_H h} B_H \\ \hline I & B_H \end{array} \right) \quad (5.47)$$

Hence,

$$\hat{H} \sim (z) \hat{H}(z) = \bar{H}_s(z) \sim \bar{C}_H^* \bar{C}_H \bar{H}_s(z)$$

where  $\bar{C}_H$  is a matrix which satisfies

$$\bar{C}_H^* \bar{C}_H = \int_0^h e^{A_H^* \tau} C_H^* C_H e^{A_H \tau} d\tau. \quad (5.48)$$

The following proposition explains the relationship of Assumption  $\mathcal{A}_3$  with other criteria.

**Proposition 5.4.6.** *Let  $\bar{C}_H$  and  $\bar{H}_s(z)$  as in (5.48) and (5.47) respectively. If the hold  $\dot{H} \in \mathbb{H}^\infty$  given by STPBC (5.43) satisfies Assumption  $\mathcal{A}_6$  (i.e.  $Ee^{A_H h}$  is Schur) then the following are equivalent:*

1. *The inner outer factorization of hold  $\dot{H} = \dot{H}_1 \bar{H}_0$  exists with inner  $\dot{H}_1 \in \mathbb{H}^\infty$ , and bistable and bicausal  $\bar{H}_0 \in \mathbb{H}^\infty$  (Assumption  $\mathcal{A}_3$ ).*
2. *The spectral factorization of  $\dot{H} \sim \dot{H}$  exists i.e. there exists a bistable and bicausal spectral factor  $\bar{W}$  such that  $\dot{H} \sim \dot{H} = \bar{W} \sim \bar{W}$ .*
3.  *$\dot{H} \sim (e^{j\theta}) \dot{H} (e^{j\theta}) > 0$  i.e. the matrix*

$$\begin{bmatrix} Ee^{A_H h} - e^{j\theta} I & B_H \\ \bar{C}_H & 0 \end{bmatrix}$$

*has full column rank for every  $\theta \in [-\pi, \pi]$ .*

4. *The discrete algebraic Riccati equation*

$$Q_0 = e^{A_H^* h} E^* (Q_0 - Q_0 B_H (B_H^* Q_0 B_H)^{-1} B_H^* Q_0) E e^{A_H h} + \bar{C}_H^* \bar{C}_H$$

*has a unique solution  $Q_0$  for which  $(E - B_H (B_H^* Q_0 B_H)^{-1} B_H^* Q_0 E) e^{A_H h}$  is Schur-stable.*

5. *There exists an  $\epsilon > 0$  such  $\|\mathcal{H}\bar{u}\|_2 \geq \epsilon \|\bar{u}\|_2$  for all  $\bar{u} \in \ell^2$ . Here  $\mathcal{H} \in \mathbb{H}^\infty$  is the hold in the time domain.*

*Proof.* See Appendix 5.C (page 178). □

Thus, using Proposition 5.4.6, we can easily check that Assumption  $\mathcal{A}_3$  is satisfied or not. We also use Proposition 5.4.6 later in obtaining an inner-outer factorization of the hold.

Now, we show that Assumption  $\mathcal{A}_3$  is satisfied if assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied.

**Lemma 5.4.7.** *Let STPBC of hold  $\dot{H}$  be given by (5.43). If assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied then there exists a factorization of  $\dot{H} = \dot{H}_1 \bar{H}_0$  with inner  $\dot{H}_1 \in \mathbb{H}^\infty$ , and bistable and bicausal  $\bar{H}_0 \in \mathbb{H}^\infty$  (i.e. Assumption  $\mathcal{A}_3$  is satisfied).*

*Proof.* Since  $(C_H, A_H)$  is observable, we have  $\bar{C}_H^* \bar{C}_H > 0$ . This implies  $\bar{C}_H$  have full column rank. As  $B_H$  and  $\bar{C}_H$  has full column rank, condition 3 of Proposition 5.4.6 is satisfied. Now, the results follows from Proposition 5.4.6. □

### 5.4.3 Inner-outer factorization of hold

There are many ways of obtaining an inner-outer factorization for  $\dot{H}$  given in (5.43). We adopted the method used by [40] i.e we first obtain a right coprime factorization (RCF) over  $\mathbb{H}^\infty$  of  $\dot{H} = \dot{N}_H \bar{M}_H^{-1}$  and then we make  $\dot{N}_H$  inner (i.e.

$\tilde{N}_H \tilde{N}_H = I$ ).  $\tilde{N}_H$  and  $\tilde{M}_H$  are said to be coprime in  $H^\infty$  if  $\tilde{N}_H$  and  $\tilde{M}_H$  are in  $H^\infty$  and there exist Bezout factors  $\tilde{X}_r \in H^\infty$  and  $\tilde{Y}_r \in H^\infty$  such that

$$\tilde{X}_r \tilde{M}_H + \tilde{Y}_r \tilde{N}_H = I.$$

We start with a right coprime factorization (RCF) of  $\dot{H}$ .

**Lemma 5.4.8.** *Consider the hold  $\dot{H}$  given by (5.43). If Assumption  $\mathcal{A}_6$  is satisfied, then there exists a matrix  $F$  such that  $(E + B_H F)e^{A_H h}$  is Schur. Now, the Hold  $\dot{H} = \tilde{N}_H(\tilde{M}_H)^{-1}$ , where  $\tilde{N}_H \in H^\infty$  and  $\tilde{M}_H \in H^\infty$  are right coprime and given by*

$$\begin{bmatrix} \tilde{M}_H(z) \\ \tilde{N}_H(z) \end{bmatrix} = \left[ \begin{array}{c|c} A_H & B_H \mathcal{J}_0^+ \\ \hline \frac{1}{z} \mathcal{J}_h^* F & I \\ C_H & 0 \end{array} \right] \begin{bmatrix} zI & -(E + B_H F) \end{bmatrix} \quad (5.49)$$

*Proof.* See Appendix 5.C (page 179). □

We give an important result about the Sylvester differential equation which will be useful later.

**Lemma 5.4.9.** *Consider the Sylvester differential equation*

$$\dot{X}(t) = AX(t) + X(t)B + C$$

*with initial condition  $X(0) = X_0$ , where  $A \in \mathbb{C}^{m \times m}$ ,  $B \in \mathbb{C}^{n \times n}$ ,  $C \in \mathbb{C}^{m \times n}$ . The solution of the Sylvester differential equation is given by*

$$X(t) = e^{At} X_0 e^{Bt} + \int_0^t e^{As} C e^{Bs} ds$$

*Proof.* See [24, chapter 8]. □

Now, we concentrate on exploiting the factor  $F$  in (5.49) to make  $\tilde{N}_H$  inner in Lemma 5.4.8. This is because if  $\tilde{N}_H$  is inner then the inner factor  $\dot{H}_i$  of  $\dot{H}$  is  $\tilde{N}_H$  and the bicausal and bistable factor  $\dot{H}_o$  of  $\dot{H}$  is  $\tilde{M}_H^{-1}$ . Note that  $\tilde{M}_H^{-1}$  is stable and causal if the hold  $\dot{H}$  is stable and causal.

Since  $\dot{H}$  is assumed to be in  $H^\infty$ ,  $F = 0$  renders a trivial RCF i.e  $\tilde{N}_H = \dot{H}$  and  $\tilde{M}_H = I$ . However, in general  $\dot{H}$  is not inner. Therefore, we need a non-trivial  $F$  to make  $\tilde{N}_H$  inner.

**Lemma 5.4.10** (Inner-outer factorization of the Hold). *Consider the Hold  $\dot{H}$  given by (5.43) and suppose that assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied. Then, there exists a unique stabilizing solution  $Q_0 > 0$  of the Riccati equation*

$$Q_0 = e^{A_H^* h} E^* (Q_0 - Q_0 B_H (B_H^* Q_0 B_H)^{-1} B_H^* Q_0) E e^{A_H h} + \bar{C}_H^* \bar{C}_H$$

where  $\bar{C}_H$  is a matrix which satisfies (5.48) and  $B_H^* Q_0 B_H$  is invertible. If we define

$$\begin{aligned} Z &:= (B_H^* Q_0 B_H)^{-\frac{1}{2}} \\ F &:= -(B_H^* Q_0 B_H)^{-1} B_H^* Q_0 E \\ \dot{H}_i &:= \dot{N}_H Z \end{aligned} \quad (5.50)$$

$$\bar{H}_0^{-1} := \bar{M}_H Z \quad (5.51)$$

where  $\bar{M}_H$  and  $\dot{N}_H$  are given by (5.49), then  $\dot{H}_i \bar{H}_0$  forms an inner-outer factorization of  $\dot{H}$  with inner  $\dot{H}_i \in H^\infty$ , and bistable and bicausal  $\bar{H}_0 \in H^\infty$ .

*Proof.* See Appendix 5.C (page 180).  $\square$

#### 5.4.4 The condition $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$

The first thing we need to check is the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  (see Proposition 5.2.2) for the existence of a solution of Problem  $\mathcal{P}_5$ . Note that we have to check  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  with the constraint that  $\check{G}_v$  is causal. A state space formulation of the condition can be done by using the STPBC of  $\dot{H}_i$  (the inner factor of hold  $\dot{H}$ ). However, the construction of  $\dot{H}_i$  requires a Riccati equation (see Section 5.4.3). The main aim of this section is to check the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  with a computationally efficient method which does not require Riccati equations. Advantage is that if the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  is not satisfied, then there is no need of doing the inner-outer factorization of hold.

We now state some results that will later help in simplifying the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$ . We start with a factorization of hold  $\dot{H}$  into a zero order hold and a discrete system.

**Lemma 5.4.11.** *Assume  $\bar{C}_H$  is a matrix which satisfy (5.48) and let the discrete system  $\bar{H}_s$  be as in (5.47). Let  $\dot{V}_H$  be a zero order hold defined as*

$$\check{y}(z) = \dot{V}_H(z) \bar{x}(z) : \quad \check{y}(z; \tau) = \int_0^h C_H e^{A_H \tau} \bar{C}_H^+ \bar{x}(z) d\sigma, \quad \tau \in [0, h)$$

where  $\bar{C}_H^+$  is the pseudo-inverse of the matrix  $\bar{C}_H$ . Then,

1. The discrete system  $\dot{V}_H^\sim \dot{V}_H$  is an orthogonal projection onto  $\text{Im } \bar{O}_{\bar{C}_H}$  where  $\bar{O}_{\bar{C}_H} := \bar{C}_H$  is a static discrete system in  $H^\infty$ .
2.  $\dot{V}_H \dot{V}_H^\sim \bar{C}_H = \bar{C}_H$  where  $\bar{C}_H \in H^\infty$  is a lifted zero order hold whose hold function in lifted  $z$ -domain is given by  $C_H e^{A_H \tau}$ .
3. Hold  $\dot{H}$  given by STPBC (5.43) can be factorized into a zero order hold and a discrete system as  $\dot{H} = \dot{V}_H \bar{H}_{cs}$  where  $\bar{H}_{cs}$  is a discrete system defined as

$$\bar{H}_{cs} := \dot{V}_H^\sim \dot{H} = \bar{C}_H \bar{H}_s = \left( \begin{array}{c|c} E e^{A_H h} & E e^{A_H h} B_H \\ \hline \bar{C}_H & \bar{C}_H B_H \end{array} \right). \quad (5.52)$$

*Proof.* See Appendix 5.C (page 182).  $\square$

**Remark 5.4.12.** *The above lemma is generic in the sense that we did not assume the observability of  $(C_H, A_H)$  (Assumption  $\mathcal{A}_8$ ). However, if  $(C_H, A_H)$  is observable then  $\bar{C}_H$  is left invertible. Hence,  $\bar{C}_H^+ \bar{C}_H = I$ . Now Lemma 5.4.11.(3) follows from  $\dot{H} = \dot{C}_H \bar{C}_H^+ \bar{C}_H \bar{H}_s = \dot{V}_H \bar{H}_{cs}$ . Note that  $\dot{V}_H \neq 0$  in this case.*

Similarly, we factorize the inner factor  $\dot{H}_i$  of hold  $\dot{H}$  into a zero order hold and a discrete system.

**Lemma 5.4.13.** *Let the zero order hold  $\dot{V}_H$  and matrix  $\bar{C}_H$  be as in Lemma 5.4.11. If assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied then*

1.  $\dot{H}_i$  is factorized into a hold and a discrete system as

$$\dot{H}_i = \dot{V}_H \bar{H}_{ics}$$

where  $\bar{H}_{ics}$  is a discrete system defined as

$$\bar{H}_{ics} := \left( \begin{array}{c|c} (E + B_H F)e^{A_H h} & (E + B_H F)B_H \\ \hline \bar{C}_H & \bar{C}_H B_H \end{array} \right) (B_H^* Q_0 B_H)^{-\frac{1}{2}}. \quad (5.53)$$

Here  $F$  and  $Q_0$  are as defined in Lemma 5.4.10.

2.  $\bar{H}_{ics}$  defined in (5.53) is inner i.e.  $\bar{H}_{ics}^{\sim} \bar{H}_{ics} = I$ .

*Proof.* If assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied then  $\dot{H}_i \in H^\infty$  exists by Lemma 5.4.7. Using Lemma 5.4.10, we have

$$\dot{H}_i(z) = \left[ \begin{array}{c|c} A_H & B_H \\ \hline C_H & 0 \end{array} \right] \begin{bmatrix} zI & -(E + B_H F) \end{bmatrix} Z \mathcal{J}_0^*$$

where  $Z := (B_H^* Q_0 B_H)^{-\frac{1}{2}}$ . Using (5.23), we have

$$\dot{H}_i(z) = C_H e^{A_H \tau} \left( I + (zI - (E + B_H F)e^{A_H h})^{-1} (E + B_H F) \right) B_H Z$$

The rest of the proof of the part 1 is similar to the proof of Lemma 5.4.11.

Since  $\dot{V}_H^{\sim} \dot{V}_H$  is an orthogonal projection onto the space  $(\text{Ker } \dot{V}_H^{\sim} \dot{V}_H)^\perp = \text{Im } \bar{C}_H \neq 0$  (see Lemma 5.4.11), we have  $\dot{V}_H^{\sim} \dot{V}_H \bar{H}_{ics} = \bar{H}_{ics}$ . Now,

$$\bar{H}_{ics}^{\sim} \bar{H}_{ics} = \bar{H}_{ics}^{\sim} \dot{V}_H^{\sim} \dot{V}_H \bar{H}_{ics} = \dot{H}_i^{\sim} \dot{H}_i = I$$

as  $\dot{V}_H \neq 0$  (see Remark 5.4.12).  $\square$

Note that in the above result, assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are required for existence of inner  $\hat{H}_i \in H^\infty$  and non-zero  $\hat{V}_H$ .

Since the factorization of  $\hat{H}$  and  $\hat{H}_i$  contains the same hold  $\hat{V}_H$  (see lemmas 5.4.11 and 5.4.13), we can expect some relationship between the discrete systems  $\bar{H}_{cs}$  and  $\bar{H}_{ics}$ . Indeed.

**Lemma 5.4.14.** *Let assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  be satisfied. Also, let  $\hat{V}_H$  and  $\bar{H}_{cs}$  be as in Lemma 5.4.11 and  $\bar{H}_{ics}$  be as in Lemma 5.4.13. Then,*

$$\bar{H}_{cs} = \bar{H}_{ics} \bar{H}_0$$

*Proof.* See Appendix 5.C (page 183). □

We define

$$\hat{H}^+ := \bar{H}_0^{-1} \hat{H}_i^{\sim} \quad (5.54a)$$

$$\bar{H}_{cs}^+ := \bar{H}_0^{-1} \bar{H}_{ics}^{\sim} \quad (5.54b)$$

Note that by definition  $\hat{H}^+$  and  $\bar{H}_{cs}^+$  exist (in  $L^\infty$ ) iff inner  $\hat{H}_i$  and (bistable and bicausal)  $\bar{H}_0$  in  $H^\infty$  exist (see also Lemma 5.4.13). Also, the existence of  $\hat{H}_i$  and  $\bar{H}_0$  in  $H^\infty$  is guaranteed by assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  (see Lemma 5.4.7). Now, using Lemma 5.4.13, we have

$$\begin{aligned} \hat{H}^+ &= \bar{H}_0^{-1} \hat{H}_i^{\sim} = \bar{H}_0^{-1} \bar{H}_{ics}^{\sim} \hat{V}_H^{\sim} \\ &= \bar{H}_{cs}^+ \hat{V}_H^{\sim} \end{aligned}$$

Now, using  $\hat{H} = \hat{V}_H \bar{H}_{cs} = \hat{H}_i \bar{H}_0$ , we have the following result which states the equivalence of  $(I - \hat{H}_i \hat{H}_i^{\sim})$ . These equivalences are useful later in simplifying the condition  $(I - \hat{H}_i \hat{H}_i^{\sim}) \hat{G}_v \in L^\infty$ .

**Corollary 5.4.15.** *Let assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  be satisfied. Also, let  $\hat{V}_H$  and  $\bar{H}_{cs}$  be as in Lemma 5.4.11 and  $\bar{H}_{ics}$  be as in Lemma 5.4.13. Now,*

$$I - \hat{H}_i \hat{H}_i^{\sim} = I - \hat{H} \hat{H}^+ = I - \hat{V}_H \bar{H}_{cs} \bar{H}_{cs}^+ \hat{V}_H^{\sim} = I - \hat{V}_H \bar{H}_{ics} \bar{H}_{ics}^+ \hat{V}_H^{\sim} \quad (5.55)$$

where  $\hat{H}^+$  and  $\bar{H}_{cs}^+$  are as defined in (5.54).

*Proof.* Since  $\hat{H} = \hat{V}_H \bar{H}_{cs} = \hat{H}_i \bar{H}_0 = \hat{V}_H \bar{H}_{ics} \bar{H}_0$ , we have

$$\begin{aligned} I - \hat{H} \hat{H}^+ &= I - \hat{V}_H \bar{H}_{cs} \bar{H}_{cs}^+ \hat{V}_H^{\sim} \\ &= I - \hat{V}_H \bar{H}_{ics} \bar{H}_0 \bar{H}_0^{-1} \bar{H}_{ics}^+ \hat{V}_H^{\sim} = I - \hat{V}_H \bar{H}_{ics} \bar{H}_{ics}^+ \hat{V}_H^{\sim} \\ &= I - \hat{H}_i \hat{H}_i^{\sim} \end{aligned}$$

□

Our main aim is to check the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  linearly. For this purpose, we use the relationship  $I - \dot{H}_i \dot{H}_i^\sim = I - \dot{V}_H \bar{H}_{cs} \bar{H}_{cs}^+ \dot{V}_H^\sim$  to obtain a simplification of the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  in the following lemma.

**Lemma 5.4.16.** *Let assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  be satisfied. If  $\check{G}_v$  is causal then the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  is satisfied iff  $(I - \bar{H}_{cs} \bar{H}_{cs}^+) \dot{V}_H^\sim \check{G}_v \in L^\infty$  and  $(I - \dot{V}_H \dot{V}_H^\sim) \check{G}_v \in H^\infty$ .*

*Proof.* See Appendix 5.C (page 184). □

Although Lemma 5.4.16 simplified the condition  $(I - \dot{H}_i \dot{H}_i^\sim) \check{G}_v \in L^\infty$  a little, still the construction of  $\bar{H}_{cs}^+$  depends upon  $\dot{H}_i$  (see (5.54)). Since the construction of  $\dot{H}_i$  requires a Riccati equation, our purpose is not yet fulfilled. To this end, we define a perpendicular  $\bar{H}_{cs}^\perp$  and a left-inverse  $\bar{H}_{cs}^\perp$  of  $\bar{H}_{cs}$ .

**Definition 5.4.17.** *A perpendicular  $\bar{H}_{cs}^\perp$  and a left inverse  $\bar{H}_{cs}^L$  of a discrete system  $\bar{H}_{cs}$  given by (5.52), are the discrete systems which satisfy*

$$\begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix} \bar{H}_{cs} = \begin{bmatrix} I \\ 0 \end{bmatrix}$$

Perpendicular and left inverse are not necessarily unique. For example, we can take  $\bar{H}_{cs}^\perp = 0$ .  $I - \bar{H}_{cs} \bar{H}_{cs}^+$  is also a perpendicular and  $\bar{H}_{cs}^+$  is a left inverse of  $\bar{H}_{cs}$  given by (5.52). This is because

$$(I - \bar{H}_{cs} \bar{H}_{cs}^+) \bar{H}_{cs} = \bar{H}_{cs} - \bar{H}_{cs} \bar{H}_0^{-1} \bar{H}_{ics}^\sim \bar{H}_{ics} \bar{H}_0 = 0. \quad (5.56)$$

where we used Lemma 5.4.13 and (5.54). Similarly,

$$\bar{H}_{cs}^+ \bar{H}_{cs} = \bar{H}_0^{-1} \bar{H}_{ics}^\sim \bar{H}_{ics} \bar{H}_0 = I.$$

As mentioned earlier the construction of  $I - \bar{H}_{cs} \bar{H}_{cs}^+$  and  $\bar{H}_{cs}^+$  needs a Riccati equation. However, we need to construct a perpendicular and a left inverse of discrete system  $\bar{H}_{cs}$  without a Riccati equation. There is a standard method to explicitly write a left inverse  $\bar{H}_{cs}^L$  and perpendicular  $\bar{H}_{cs}^\perp$  of  $\bar{H}_{cs}$  which does not requires a Riccati equation. For that method, we need the condition that the feed-through term of  $\bar{H}_{cs}$  i.e.  $\bar{C}_H B_H$  be left invertible. The following lemma shows that it is indeed true under our assumptions.

**Lemma 5.4.18.** *If assumption  $\mathcal{A}_7$  and  $\mathcal{A}_8$  are satisfied then  $\bar{C}_H B_H$  is left invertible.*

*Proof.* Since  $(A_H, C_H)$  is observable (Assumption  $\mathcal{A}_8$ ), using (5.48) we have

$$\bar{C}_H^* \bar{C}_H = \int_0^h e^{A_H^* \tau} C_H^* C_H e^{A_H \tau} d\tau > 0.$$

Therefore,  $B_H^* \bar{C}_H^* \bar{C}_H B_H > 0$  as  $B_H$  is full column rank (Assumption  $\mathcal{A}_7$ ). This implies  $\bar{C}_H B_H$  has full column rank. □



Now, we write a left inverse  $\bar{H}_{cs}^L$  and a perpendicular  $\bar{H}_{cs}^\perp$  of  $\bar{H}_{cs}$  without any Riccati equation.

**Lemma 5.4.19.** *If assumption  $\mathcal{A}_7$  and  $\mathcal{A}_8$  are satisfied then  $\bar{C}_H B_H$  is left invertible and there exists an invertible matrix  $\bar{D}^\S := \begin{bmatrix} \bar{D}^+ \\ \bar{D}^\perp \end{bmatrix}$  such that  $\bar{D}^\S \bar{C}_H B_H = \begin{bmatrix} I \\ 0 \end{bmatrix}$ .*

Now,

1. A left inverse  $\bar{H}_{cs}^L$  and a perpendicular  $\bar{H}_{cs}^\perp$  of  $\bar{H}_{cs}$  given by (5.52) exist and they are given by

$$\begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix} = \left( \begin{array}{c|c} A_l & Ee^{A_H h} B_H \bar{D}^+ + L_{cs} \bar{D}^\perp \\ \hline -\bar{D}^+ \bar{C}_H & \bar{D}^+ \\ -\bar{D}^\perp \bar{C}_H & \bar{D}^\perp \end{array} \right)$$

where  $A_l := Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H - L_{cs} \bar{D}^\perp \bar{C}_H$  and  $L_{cs}$  is a matrix chosen in such a way that all modes of  $Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H$  which are detectable from  $\bar{D}^\perp \bar{C}_H$  are stabilized.

2.  $\bar{H}_{cs}^\perp$  defined above is in  $H^\infty$ .
3. the eigenvalues of  $Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H$  which are unobservable from  $\bar{D}^\perp \bar{C}_H$  are the invariant zeros of the realization of  $\bar{H}_{cs}$ .

*Proof.* Note that  $\bar{C}_H B_H$  is left invertible by Lemma 5.4.18. Therefore there exists an invertible matrix  $\bar{D}^\S$  such that  $\bar{D}^\S \bar{C}_H B_H = \begin{bmatrix} I \\ 0 \end{bmatrix}$ . The proof of 1 and 2 is standard (see for example [26, §A.8.3]).

The invariant zeros of a left invertible system  $\bar{H}_{cs}$  given by (5.52) are those values of  $z \in \mathbb{C}$  where the matrix

$$\begin{bmatrix} Ee^{A_H h} - zI & Ee^{A_H h} B_H \\ \bar{C}_H & \bar{C}_H B_H \end{bmatrix}$$

loses its normal rank. The rest of the proof of 3 is similar to the proof of [26, claim A.34].

□

In Lemma 5.4.19(1),  $L_{cs}$  is chosen in such a way that  $\bar{D}^\perp \bar{C}_H (zI - A_l)^{-1}$  is a stable discrete system. Note that matrix  $A_l$  is not necessarily Schur. The importance of Lemma 5.4.19(3) will be more clear in Lemma 5.4.27.

**Remark 5.4.20.** *In Lemma 5.4.19,  $\bar{H}_{cs}^\perp$  is in  $H^\infty$ . However, it is not necessary that the left inverse  $\bar{H}_{cs}^L$  is in  $H^\infty$ . This is because  $Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H - L_{cs} \bar{D}^\perp \bar{C}_H$  may have some unstable eigenvalues (i.e. eigenvalues having modulus greater than one) that are observable from  $\bar{D}^+ \bar{C}_H$ . However, under some assumptions  $\bar{H}_{cs}^L$  can be constructed stable. In the sequel we do not need stability of  $\bar{H}_{cs}^L$ , therefore we do not delve into this interesting topic (for details see [26, §A.8.3]). Note that by construction of  $L_{cs}$ , we did not face this problem for  $\bar{H}_{cs}^\perp$ .*

Although the above constructed  $\bar{H}_{cs}^\perp$  and  $(I - \bar{H}_{cs}\bar{H}_{cs}^+)$  are such that  $\bar{H}_{cs}^\perp\bar{H}_{cs} = 0$  and  $(I - \bar{H}_{cs}\bar{H}_{cs}^+)\bar{H}_{cs} = 0$ , it is not necessary that  $\bar{H}_{cs}^\perp = I - \bar{H}_{cs}\bar{H}_{cs}^+$  in general. However, there exists a relation between  $I - \bar{H}_{cs}\bar{H}_{cs}^+$  and  $\bar{H}_{cs}^\perp$  given by the following lemma.

**Lemma 5.4.21.** *If assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied then there exists a left invertible system  $\bar{W}_{cs} \in L^\infty$  such that*

$$\bar{W}_{cs}\bar{H}_{cs}^\perp = I - \bar{H}_{cs}\bar{H}_{cs}^+$$

*Proof.* See Appendix 5.C (page 184). □

**Remark 5.4.22.**  $\bar{H}_{cs}^\perp\hat{V}_H^\sim$  maps the image of  $\hat{H}$  to zero as

$$\bar{H}_{cs}^\perp\hat{V}_H^\sim\hat{H} = \bar{H}_{cs}^\perp\bar{H}_{cs} = 0.$$

However,  $\bar{H}_{cs}^\perp\hat{V}_H^\sim$  does not represent all perpendicular systems of hold  $\hat{H}$ . Indeed one such perpendicular is  $I - \hat{H}_1\hat{H}_1^\sim$ . Using (5.55), we have

$$\begin{aligned} I - \hat{H}_1\hat{H}_1^\sim &= I - \hat{V}_H\bar{H}_{cs}\bar{H}_{cs}^+\hat{V}_H^\sim \\ &= I - \hat{V}_H(I - \bar{W}_{cs}\bar{H}_{cs}^\perp)\hat{V}_H^\sim \\ &= I - \hat{V}_H\hat{V}_H^\sim + \hat{V}_H\bar{W}_{cs}\bar{H}_{cs}^\perp\hat{V}_H^\sim \end{aligned} \quad (5.57)$$

In general,  $I - \hat{V}_H\hat{V}_H^\sim \neq 0$  (see Lemma 5.4.11). Hence,  $\bar{H}_{cs}^\perp\hat{V}_H^\sim$  does not represent all operators perpendicular to  $\hat{H}$ .

For example, let  $\hat{H} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ , then  $\bar{H}_{cs} = \sqrt{\frac{e^{2h}-1}{2}} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$  and  $\hat{V}_H = e^\tau \sqrt{\frac{2}{e^{2h}-1}}$ . Therefore  $\bar{D}^\perp = 0$ , hence  $\bar{H}_{cs}^\perp = 0$ , but clearly  $I - \hat{V}_H\hat{V}_H^\sim \neq 0$ .

Using Lemma 5.4.21, the condition  $(I - \hat{H}_1\hat{H}_1^\sim)\check{G}_v \in L^\infty$  can be further simplified as explained in the following theorem.

**Theorem 5.4.23.** *Let assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  be satisfied. Also let  $\bar{H}_{cs}^\perp$  be as in Lemma 5.4.19 and  $\hat{V}_H$  be as in Lemma 5.4.11. If  $\check{G}_v$  is causal, then the condition  $(I - \hat{H}_1\hat{H}_1^\sim)\check{G}_v \in L^\infty$  is satisfied iff  $\bar{H}_{cs}^\perp\hat{V}_H^\sim\check{G}_v \in H^\infty$  and  $(I - \hat{V}_H\hat{V}_H^\sim)\check{G}_v \in H^\infty$ .*

*Proof.* See Appendix 5.C (page 185). □

Still it is not clear that Theorem 5.4.23 offers any advantage or not. However, since  $\hat{V}_H$  is a zero order hold and  $\bar{H}_{cs}^\perp$  can be constructed in finite steps (as demonstrated in Lemma 5.4.19), the conditions  $\bar{H}_{cs}^\perp\hat{V}_H^\sim\check{G}_v \in H^\infty$  and  $(I - \hat{V}_H\hat{V}_H^\sim)\check{G}_v \in H^\infty$  requires no Riccati equation. Thus, we avoided the Riccati equation needed if we directly check  $(I - \hat{H}_1\hat{H}_1^\sim)\check{G}_v \in L^\infty$  using STPBC of  $\hat{H}_1$ .

**Simple equations to check the condition**  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$

In this section, we use Theorem 5.4.23 to obtain simple equations to check the condition  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$ . If  $\check{G}_v$  belongs to  $H^\infty$  then the condition  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$  (with constraint that  $\check{G}_v$  is causal) is trivially satisfied. Therefore, we can add or subtract any  $L^\infty$  system from  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v$  without any loss of generality. The following lemma explains this idea.

**Lemma 5.4.24.** *Assume that causal  $\check{G}_v$  has STPBC given in (5.42). Now,  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$  is equivalent to  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_u \in L^\infty$  where  $\check{G}_u$  is causal and it is given by STPBC*

$$\check{G}_u = \left[ \begin{array}{c|c} A_u & B_u \\ \hline C_{vu} & 0 \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix} \quad (5.58)$$

*Proof.* The proof follows from the fact that  $\check{G}_v = \check{G}_u + \check{G}_s$  and  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_s \in L^\infty$  where

$$\check{G}_s = \left[ \begin{array}{c|c} A_s & B_s \\ \hline C_{vs} & 0 \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix} \in H^\infty.$$

□

Also, it follows from Theorem 5.4.23 that if  $\check{G}_u$  is causal then the condition  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_u \in L^\infty$  is satisfied iff  $\check{H}_{cs}^\perp \check{V}_H^\sim \check{G}_u \in H^\infty$  and  $(I - \check{V}_H \check{V}_H^\sim) \check{G}_u \in H^\infty$ . We first write a simple equation to check the condition  $(I - \check{V}_H \check{V}_H^\sim) \check{G}_u \in H^\infty$ . For obtaining the equation, we need the following standard but important result (written here for reference purpose).

**Lemma 5.4.25.** *If  $(A, B)$  is controllable, then*

1.  $(A_u, B_u)$  is controllable.
2.  $(e^{A_u h}, \bar{B}_u)$  is controllable where  $\bar{B}_u$  is a matrix that satisfies

$$\bar{B}_u \bar{B}_u^* = \int_0^h e^{A_u \tau} B_u B_u^* e^{A_u^* \tau} d\tau = \int_0^h e^{A_u(h-\tau)} B_u B_u^* e^{A_u^*(h-\tau)} d\tau. \quad (5.59)$$

*Proof.* Suppose  $(A, B)$  is controllable. Hence  $[A - \lambda I \quad B]$  has full row rank for all  $\lambda \in \mathbb{C}$ . Then

$$\left[ \begin{array}{ccc|c} A_s - \lambda I & 0 & B_s & \\ \hline 0 & A_u - \lambda I & B_u & \end{array} \right] \text{ has full row rank for all } \lambda \in \mathbb{C}.$$

Clearly  $[A_u - \lambda I \quad B_u]$  has full row rank as well.

Now, if  $(A_u, B_u)$  is controllable then  $\int_0^h e^{A_u \tau} B_u B_u^* e^{A_u^* \tau} d\tau = \bar{B}_u \bar{B}_u^* > 0$  [71]. This implies that  $\bar{B}_u$  has full row rank. Hence,  $[e^{A_u h} - \lambda I \quad \bar{B}_u]$  has full row rank for all  $\lambda \in \mathbb{C}$ .

The two integrals in (5.59) are equivalent as  $B_u$  is a constant matrix. □

We start with the condition  $(I - \dot{V}_H \dot{V}_H^\sim) \check{G}_u \in H^\infty$  first.

**Lemma 5.4.26.** *Let  $\check{G}_u$  be as in (5.58). Let  $\bar{C}_H$ ,  $\bar{B}_u$  and  $\bar{C}_{vu}$  satisfy (5.48), (5.59) and*

$$\bar{C}_{vu}^* \bar{C}_{vu} = \int_0^h e^{A_u^* \tau} C_{vu}^* C_{vu} e^{A_u \tau} d\tau, \quad (5.60)$$

respectively. Now, if  $(A, B)$  is controllable then  $(I - \dot{V}_H \dot{V}_H^\sim) \check{G}_u \in H^\infty$  iff

$$\bar{C}_{vu}^* \bar{C}_{vu} - P_u^* (\bar{C}_H^* \bar{C}_H)^+ P_u = 0$$

where  $P_u$  is given by

$$P_u = \int_0^h e^{A_H^* \tau} C_H^* C_{vu} e^{A_u \tau} d\tau$$

and the zero order hold  $\dot{V}_H$  is defined in Lemma 5.4.11.

*Proof.* See Appendix 5.C (page 185). □

Now, we simplify the condition  $\bar{H}_{cs}^\perp \dot{V}_H^\sim \check{G}_v \in L^\infty$ . To this end, we need the following assumption.

**Assumption  $\mathcal{A}_9$ :** *No pole of  $\check{G}_v$  in the region  $|z| \geq 1$  is a zero of the discrete system  $\bar{H}_{cs}$ .*

The poles of  $\check{G}_v$  in the region  $|z| \geq 1$  are the eigenvalues of  $e^{A_u h}$ . Also note that the eigenvalues of  $E e^{A_H h} - E e^{A_H h} B_H \bar{D}^+ \bar{C}_H$  which are unobservable from  $\bar{D}^\perp \bar{C}_H$  are the invariant zeros of the  $\bar{H}_{cs}$  (see Lemma 5.4.19(3)). Therefore, by Assumption  $\mathcal{A}_9$  and the suitable choice of  $L_{cs}$  (see Lemma 5.4.19),  $A_l$  and  $e^{A_u h}$  have no common eigenvalues. This is required in the following lemma which helps in simplifying the condition  $\bar{H}_{cs}^\perp \dot{V}_H^\sim \check{G}_u \in H^\infty$ .

**Lemma 5.4.27.** *Let assumptions  $\mathcal{A}_7$ – $\mathcal{A}_8$  be satisfied. Let  $\bar{H}_{cs}^\perp$  be as in Lemma 5.4.19. If  $(A, B)$  is controllable then  $\bar{H}_{cs}^\perp \dot{V}_H^\sim \check{G}_u \in H^\infty$  iff there exist a matrix  $X_l$  that satisfies the following linear equations*

$$\begin{aligned} (E e^{A_H h} - E e^{A_H h} B_H \bar{D}^+ \bar{C}_H) X_l - X_l e^{A_u h} + E e^{A_H h} B_H \bar{D}^+ &= 0 \\ -\bar{D}^\perp \bar{C}_H X_l + \bar{D}^\perp (\bar{C}_H^*)^+ P_u &= 0 \end{aligned}$$

*Proof.* See Appendix 5.C (page 187). □

Combining Lemma 5.4.27 and 5.4.26, we can write the following theorem:

**Theorem 5.4.28.** *Let assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  be satisfied. Also let  $\bar{C}_{vu}$ ,  $P_u$ ,  $\bar{C}_H$  be as in Lemma 5.4.26 and  $D^+$ ,  $D^\perp$  as in Lemma 5.4.19. If  $(A, B)$  is controllable and  $\check{G}_v$  is causal then  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$  iff*

$$\bar{C}_{vu}^* \bar{C}_{vu} - P_u^* (\bar{C}_H^* \bar{C}_H)^+ P_u = 0 \quad (5.61)$$

and there exists an  $X_l$  that satisfies the following linear equations

$$\left( E e^{A_H h} - E e^{A_H h} B_H \bar{D}^+ \bar{C}_H \right) X_l - X_l e^{A_u h} + E e^{A_H h} B_H \bar{D}^+ = 0 \quad (5.62a)$$

$$-\bar{D}^\perp \bar{C}_H X_l + \bar{D}^\perp (\bar{C}_H^*)^+ P_u = 0 \quad (5.62b)$$

*Proof.* Follows from Theorem 5.4.23 and lemmas 5.4.26 and 5.4.27.  $\square$

**Remark 5.4.29.** *If Assumption  $\mathcal{A}_8$  is satisfied then the condition given in (5.61) can be further simplified as explained below. Note that*

$$\int_0^h e^{A_{uH} \tau} C_{uH}^* C_{uH} e^{A_{uH} \tau} d\tau = \begin{bmatrix} \bar{C}_{vu}^* \bar{C}_{vu} & P_u^* \\ P_u & \bar{C}_H^* \bar{C}_H \end{bmatrix}$$

where  $A_{uH} := \begin{bmatrix} A_u & 0 \\ 0 & A_H \end{bmatrix}$  and  $C_{uH} := [C_{vu} \quad C_H]$ . If Assumption  $\mathcal{A}_8$  is satisfied then  $\bar{C}_H^* \bar{C}_H > 0$ . Hence,  $(\bar{C}_H^* \bar{C}_H)^+ = (\bar{C}_H^* \bar{C}_H)^{-1}$ . Using (5.61) and invertibility of  $\bar{C}_H^* \bar{C}_H$ , we have (see [71, §2.3])

$$\begin{aligned} \text{rank} \begin{bmatrix} \bar{C}_{vu}^* \bar{C}_{vu} & P_u^* \\ P_u & \bar{C}_H^* \bar{C}_H \end{bmatrix} &= \text{rank} \begin{bmatrix} \bar{C}_{vu}^* \bar{C}_{vu} - P_u^* (\bar{C}_H^* \bar{C}_H)^{-1} P_u & 0 \\ 0 & \bar{C}_H^* \bar{C}_H \end{bmatrix} \\ &= \text{rank} \bar{C}_H^* \bar{C}_H = \dim A_H. \end{aligned}$$

Here  $\dim A_H$  means number of rows of the square matrix  $A_H$ . This implies that the condition given in (5.61) is satisfied iff

$$\text{rank} \int_0^h e^{A_{uH} \tau} C_{uH}^* C_{uH} e^{A_{uH} \tau} d\tau = \dim A_H.$$

Now,  $\text{rank} \int_0^h e^{A_{uH} \tau} C_{uH}^* C_{uH} e^{A_{uH} \tau} d\tau$  is equal to the rank of the observability matrix associated with pair  $(C_{uH}, A_{uH})$  (see [71, theorem 3.3, 3.8]). Therefore, the condition given in (5.61) can be verified just by showing that the rank of the observability matrix associated with pair  $(C_{uH}, A_{uH})$  is equal to  $\dim A_H$ .

Theorem 5.4.28 is important because here we check  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$  with just a few linear equations. We do not need to solve a Riccati equation here to obtain  $\check{H}_1$ . The eigenvalue split of  $\check{G}_v$  required in Theorem 5.4.28 is also required if we directly check  $(I - \check{H}_1 \check{H}_1^\sim) \check{G}_v \in L^\infty$  with STPBC of  $\check{H}_1$ . Therefore, certainly we gain something by using Theorem 5.4.28. Let us consider an example now.

**Example 5.4.30.** Consider LCTI systems  $\mathcal{G}_v$  and  $\mathcal{G}_y$  given in the Laplace domain by  $G_v(s) = \begin{bmatrix} \frac{1}{s} & 0 \end{bmatrix}$  and  $G_y(s) = \begin{bmatrix} \frac{1}{s} & \epsilon \end{bmatrix}$  where  $\epsilon > 0$ .

The STPBC of  $\mathcal{G}_v$  and  $\mathcal{G}_y$  is given by

$$\begin{bmatrix} \check{G}_v(z) \\ \check{G}_y(z) \end{bmatrix} = \left[ \begin{array}{c|cc} 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ 1 & 0 & \epsilon \end{array} \right] \begin{bmatrix} z & -1 \end{bmatrix}.$$

We also assume that the hold is the ideal zero order hold given by

$$\check{H}_{iz}(z) = \left[ \begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array} \right] \begin{bmatrix} 1 & 0 \end{bmatrix} \mathcal{J}_{0^+}$$

The first thing we need to check is the condition  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^\infty$  (see Proposition 5.2.2) for the existence of a solution of Problem  $\mathcal{P}_5$  where  $\check{H}_i$  is the inner factor of hold  $\check{H}_{iz}$ . In Theorem 5.4.28, we saw that this condition can be checked without constructing the inner factor  $\check{H}_i$ . In this example,  $\bar{C}_{vu} = \bar{C}_H = \sqrt{h}$ ,  $P_u = h$ ,  $E = 0$ ,  $\bar{D}^+ = \frac{1}{\sqrt{h}}$  and  $\bar{D}^- = 0$ . As required in Theorem 5.4.28,

$$\bar{C}_{vu}^* \bar{C}_{vu} - P_u^* (\bar{C}_H^* \bar{C}_H)^+ P_u = h - h = 0$$

and  $X_l = 0$  satisfies (5.62).

### 5.4.5 The condition $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^2$

We also need to check the condition  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^2$  (see Proposition 5.2.5) for the existence of a solution of Problem  $\mathcal{P}_5$ . To this end Lemma 5.3.23 and Remark 5.3.24 are very useful and applied in the following result.

**Corollary 5.4.31.** Let system  $\check{G}$  and  $\check{H}$  are given by STPBC (5.42) and (5.43) respectively. Then  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^\infty$  implies  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^2$ .

*Proof.* Since the direct feed through term of  $\check{G}_v$  is 0, the STPBC of  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v$  has zero direct feed-through term. Therefore, it follows from Lemma 5.3.23 and Remark 5.3.24 that if  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^\infty$  then  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^2$ .  $\square$

Causality of  $\check{G}_v$  does not play a role in Corollary 5.4.31. This is because once it proved that  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^\infty$  with any constraint (here  $\check{G}_v$  is causal) then  $(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^2$  with the same constraint.

### 5.4.6 Obtaining $\check{V}$

Now, we concentrate on showing the existence of a  $\check{V} \in L^\infty$  such that  $\check{M}_h := \check{H}_i^\sim \check{M}_v - \check{V} \check{M}_y \in H^\infty$  (see Proposition 5.2.2). In this section, we show that

$$\check{V}(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} \frac{-A_H^*}{-(B_H Z)^*} & \frac{P(\tau) L Z_v^{-1}}{0} \end{array} \right] \begin{bmatrix} z(E + B_H F)^* & -I \end{bmatrix} \quad (5.63)$$

is in  $L^\infty$  and it is such that  $\dot{M}_h \in H^\infty$ , where  $P(\tau)$  is the solution of Sylvester differential equation

$$\dot{P}(\tau) + A_H^* P(\tau) + P(\tau) A + C_H^* C_v = 0, \quad (5.64)$$

with boundary condition  $(E + B_H F)^* P(0) = -P(h)$ .

**Remark 5.4.32.** *This STPBC of  $\dot{V}$  is not brute force. It can be obtained by constructing a rational function of  $\dot{H}_i \check{M}_v \check{M}_y^{-1}$  (with time varying co-efficient). Even though  $\dot{H}_i \check{M}_v \check{M}_y^{-1}$  may have no convergence for all  $z \in \mathbb{Z}$  (this can happen if region of convergence (ROC) of  $\dot{H}_i$  and ROC of  $\check{M}_v \check{M}_y^{-1}$  have empty intersection), but still as a rational  $\dot{H}_i \check{M}_v \check{M}_y^{-1}$  has some meaning. Doing partial fraction*

$$\dot{H}_i \check{M}_v \check{M}_y^{-1} = P_{\dot{H}_i} + P_{\check{M}_v \check{M}_y^{-1}}$$

where  $P_{\dot{H}_i}$  contains the poles (outside unit circle) of  $\dot{H}_i$  and  $P_{\check{M}_v \check{M}_y^{-1}}$  contains the poles (inside unit circle) of  $\check{M}_v \check{M}_y^{-1}$ , the desired  $\dot{V}$  can be obtained by taking  $\dot{V} = P_{\dot{H}_i}$ . Since the poles of  $P_{\dot{H}_i}$  are outside unit circle,  $\dot{V}$  is stable if it is anti-causal. Also

$$\dot{M}_h := \dot{H}_i \check{M}_v - \dot{V} \check{M}_y = (\dot{H}_i \check{M}_v \check{M}_y^{-1} - \dot{V}) \check{M}_y = P_{\check{M}_v \check{M}_y^{-1}} \check{M}_y.$$

It can be shown that this  $\dot{M}_h$  is stable and causal.

First, we show that Sylvester differential equation (5.64) with boundary condition  $(E + B_H F)^* P(0) = -P(h)$  has a solution. To this end, following Lemma is useful.

**Lemma 5.4.33.** *If Assumption  $\mathcal{A}_9$  (given at page 153) is satisfied then no pole of  $\dot{H}_i$  in the region  $|z| \geq 1$  is a pole of  $\check{G}_v$ .*

*Proof.* Similar to  $\dot{H} = \dot{V}_H \bar{H}_{cs}$  in Lemma 5.4.11, we can write  $\dot{H}_i = \dot{V}_H \bar{H}_{ics}$  where

$$\bar{H}_{ics} := \left( \begin{array}{c|c} (E + B_H F)e^{A_H h} & (E + B_H F)e^{A_H h} B_H \\ \hline \bar{C}_H & \bar{C}_H B_H Z \end{array} \right)$$

Since  $\bar{H}_0$  is bicausal and bistable, it will not have zeros in the region  $|z| \geq 1$ . Therefore, a zero of  $\bar{H}_{cs}$  in the region  $|z| \geq 1$  is a zero of  $\bar{H}_{ics}$ . Now  $\dot{H}_i = \bar{H}_{ics} \dot{V}_H$ . As  $\dot{V}_H$  is a sampler with no poles, the poles of  $\dot{H}_i$  are the poles of  $\bar{H}_{ics}$ . Also, since  $\bar{H}_{ics} \bar{H}_{ics} = I$ , the poles of  $\bar{H}_{ics}$  are the zero of  $\bar{H}_{ics}$ . Therefore, Assumption  $\mathcal{A}_9$  can be restated as "no pole of  $\dot{H}_i$  in the region  $|z| \geq 1$  is a pole of  $\check{G}_v$  (or  $\check{G}_u$ )".  $\square$

We use the above Lemma to show that Sylvester differential equation (5.64) with boundary condition  $(E + B_H F)^* P(0) = -P(h)$  has a solution.

**Lemma 5.4.34.** *If Assumption  $\mathcal{A}_9$  is satisfied, a unique solution  $P(\tau)$  of the differential Sylvester equation (5.64) with the boundary condition  $(E + B_H F)^* P(0) = -P(h)$  exists and it is given by*

$$P(\tau) = e^{-A_H^* \tau} P_0 e^{-A \tau} - R_3(\tau),$$

where

$$R_3(\tau) := \int_0^\tau e^{-A_H^* \tau_1} C_H^* C_v e^{-A \tau_1} d\tau_1$$

and  $P_0$  is the unique solution of Sylvester equation

$$P_0 = e^{A_H^* h} \left( -(E + B_H F)^* P_0 + R_3(h) \right) e^{A h}. \quad (5.65)$$

*Proof.* From Lemma 5.4.9, the solution of (5.64) with boundary condition  $P(0) = P_0$  is given by

$$P(\tau) = R_1(\tau) P_0 R_2(\tau) - R_3(\tau)$$

where  $R_1(\tau) := e^{-A_H^* \tau}$  and  $R_2(\tau) := e^{-A \tau}$ . Using above and the boundary condition  $(E + B_H F)^* P(0) = -P(h)$ , we have

$$P(h) = R_1(h) P_0 R_2(h) - R_3(h) = -(E + B_H F)^* P_0$$

this implies

$$P_0 = -R_1(h)^{-1} (E + B_H F)^* P_0 R_2(h)^{-1} + R_1(h)^{-1} R_3(h) R_2(h)^{-1}$$

The unique solution of the above discrete Sylvester equation exists by Lemma 5.4.33.  $\square$

Now, we show that  $\hat{V}$  defined in (5.63) is in  $L^\infty$ .

**Lemma 5.4.35.**  *$\hat{V}$  defined in (5.63) is in  $L^\infty$ .*

*Proof.* Since the  $P(\tau)$  exists and continuous (hence, bounded), and  $(E + B_H F) e^{A h}$  is Schur, it follows from Remark 5.3.19 and Lemma 5.3.18 that  $\hat{V} \in L^\infty$ .  $\square$

Now, we show that this  $\hat{V} \in L^\infty$  is such that  $\hat{M}_h := \hat{H}_i \tilde{M}_v - \hat{V} \check{M}_y \in H^\infty$ .

**Lemma 5.4.36.** *If  $\hat{V}$  is as defined in (5.63) then  $\hat{M}_h := \hat{H}_i \tilde{M}_v - \hat{V} \check{M}_y \in H^\infty$  and*

$$\hat{M}_h(z) = \mathcal{J}_{0^*}^* \left[ \begin{array}{c|c} A + LC_y & L \\ \hline -(B_H Z)^* P_0 & 0 \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix} \quad (5.66)$$

*Proof.* See Appendix 5.C (page 189).  $\square$



### 5.4.7 Optimal relaxed causal sampler

In this section, we write a STPBC for optimal relaxed causal sampler described in Proposition 5.2.5. For the solution described in Proposition 5.2.5, we need  $\tilde{H}_1 \tilde{N}_v \tilde{N}_y - \dot{V}$ . Following corollary show how to obtain a compact STPBC for this.

**Corollary 5.4.37.** Define  $A_p := \begin{bmatrix} -A_H^* & -C_H^* C_v X \\ 0 & -A_L^* \end{bmatrix}$ ,  $B_p := \begin{bmatrix} -P(\tau) L Z_y^{-1} \\ (Z_y C_y)^* \end{bmatrix}$ ,  $\Omega_p := \begin{bmatrix} (E + B_H F)^* & 0 \\ 0 & I \end{bmatrix}$ , and  $C_p := \begin{bmatrix} -(B_H Z)^* & 0 \end{bmatrix}$ . Now

$$\tilde{H}_1 \tilde{N}_v \tilde{N}_y(z) - \dot{V}(z) = \mathcal{J}_{0^*} \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p & 0 \end{array} \right] \begin{bmatrix} z \Omega_p & -I \end{bmatrix}$$

*Proof.* See Appendix 5.C (page 190).  $\square$

Now, we have all component to write the optimal relaxed causal sampler described in Proposition 5.2.5.

**Theorem 5.4.38.** Let system  $\check{G}$  and  $\check{H}$  are given by STPBC (5.42) and (5.43) respectively. Let assumptions  $\mathcal{A}_4$ - $\mathcal{A}_9$  be satisfied. If the conditions given in Theorem 5.4.28 are satisfied then

$$\dot{S}_{\text{opt}} := \arg \inf_{\dot{S} \in z^l \mathbb{H}^\infty} \|\check{G}_v - \check{H} \dot{S} \check{G}_y\|_{\mathbb{L}^2} = \bar{H}_0^{-1} (\dot{S}_{\alpha, \text{opt}} \check{M}_y - \dot{M}_h)$$

where  $\dot{S}_{\alpha, \text{opt}} := \text{proj}_{z^l \mathbb{H}^2} (\tilde{H}_1 \tilde{N}_v \tilde{N}_y - \dot{V})$  has STPBC

$$\dot{S}_{\alpha, \text{opt}}(z) = \mathcal{J}_{0^*} \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p (I - (z e^{-A_p h} \Omega_p)^{l+1}) & 0 \end{array} \right] \begin{bmatrix} \Omega_p & -I \end{bmatrix}$$

where  $A_p$ ,  $B_p$ ,  $\Omega_p$ , and  $C_p$  are defined in Corollary 5.4.37.

*Proof.* See Appendix 5.C (page 191).  $\square$

Theorem 5.4.38 says that strictly causal sampler (i.e.  $l = -1$ ) is given by  $-\bar{H}_0^{-1} \dot{M}_h$  as  $\tilde{H}_1 \tilde{N}_v \tilde{N}_y$  and  $\dot{V}$  are anti-causal.

**Remark 5.4.39.** Note that the causality (strictly speaking lifted causality) of a sampler is not equivalent to input/output causality (see Section 2.5 for detail).

### 5.4.8 Optimal $\mathbb{H}^2$ norm

In this section we provide a simple algebraic expression to calculate the optimal error norm  $\|\check{G}_{e, \text{opt}}\|_{\mathbb{L}^2}$  for signal generator  $\check{G}$  given in (5.41) and hold  $\check{H}$  given in (5.43). It is given in (5.6) that

$$\|\check{G}_{e, \text{opt}}\|_{\mathbb{L}^2}^2 := \|\check{G}_v - \check{H} \dot{S}_{\text{opt}} \check{G}_y\|_{\mathbb{L}^2}^2 = \|\check{G}_v + \check{H}_1 \dot{M}_h \check{G}_y\|_{\mathbb{L}^2}^2 - \|\dot{S}_{\alpha, \text{opt}}\|_{\mathbb{L}^2}^2$$

where  $\hat{S}_{\text{opt}}$  is given in Theorem 5.4.38 and

$$\|\hat{S}_{\alpha, \text{opt}}\|_{\mathbb{L}^2}^2 = \|\text{proj}_{z^l \mathbb{H}^2}(\hat{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})\|_{\mathbb{L}^2}^2 \quad (5.67)$$

The squared  $\mathbb{L}^2$  norm of  $\hat{S}_{\alpha, \text{opt}}$  can also be written as

$$\|\hat{S}_{\alpha, \text{opt}}\|_{\mathbb{L}^2}^2 = \|\hat{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}\|_{\mathbb{L}^2}^2 - \|\text{proj}_{\mathbb{L}^2 \setminus z^l \mathbb{H}^2}(\hat{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})\|_{\mathbb{L}^2}^2 \quad (5.68)$$

It is shown later that it is easier to calculate  $\|\hat{S}_{\alpha, \text{opt}}\|_{\mathbb{L}^2}$  using (5.68) than (5.67).

Now, we obtain STPBC of all systems required to calculate  $\|\check{G}_{e, \text{opt}}\|_{\mathbb{L}^2}$ . Using Corollary 5.4.37, the STPBC of sampler  $\hat{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}$  is given by

$$\hat{H}_i \tilde{N}_v \tilde{N}_y - \hat{V} := \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p & 0 \end{array} \right]_{[z \Omega_p \quad -I]} \quad (5.69)$$

where  $A_p$ ,  $B_p$ ,  $\Omega_p$ , and  $C_p$  are defined in Corollary 5.4.37. Proceeding as in the proof Theorem 5.4.38, we have

$$\text{proj}_{\mathbb{L}^2 \setminus z^l \mathbb{H}^2}(\hat{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}) := \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p (ze^{-A_p h} \Omega_p)^{l+1} & 0 \end{array} \right]_{[z \Omega_p \quad -I]} \quad (5.70)$$

Note that  $\hat{H}_i$  is a sampler and  $\hat{M}_h$  is a sampler, therefore to obtain the STPBC of  $\check{G}_v + \hat{H}_i \hat{M}_h \check{G}_y$ , we need the STPBC of  $\hat{H}_i \hat{M}_h$ . Using Lemma 5.3.32, (5.50) and (5.66), the STPBC of  $\hat{H}_i \hat{M}_h$  is given by

$$\begin{aligned} \hat{H}_i(z) \hat{M}_h(z) &= \left[ \begin{array}{cc|c} A_H & 0 & 0 \\ 0 & A + LC_y & L \\ \hline C_H & 0 & 0 \end{array} \right] \left[ z \begin{bmatrix} I & M_1 \\ 0 & I \end{bmatrix} \quad - \begin{bmatrix} E + B_H F & 0 \\ 0 & I \end{bmatrix} \right] \\ &= \left[ \begin{array}{cc|c} A_H & 0 & 0 \\ 0 & A + LC_y & L \\ \hline C_H & 0 & 0 \end{array} \right] \left[ z \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \quad - \begin{bmatrix} E + B_H F & -M_1 \\ 0 & I \end{bmatrix} \right] \end{aligned} \quad (5.71)$$

where

$$M_1 := -B_H Z (B_H Z)^* P_0. \quad (5.72)$$

Now, the STPBC of  $\check{G}_v + \hat{H}_i \hat{M}_h \check{G}_y$  is obtained in the following lemma.

**Lemma 5.4.40.** *Let the STPBC of  $\check{G} = \begin{bmatrix} \check{G}_v \\ \check{G}_y \end{bmatrix}$ ,  $\hat{M}_h$  and  $\hat{H}_i$  are given by (5.41), (5.66) and (5.50) respectively. Then, the STPBC of the system  $\check{G}_v + \hat{H}_i \hat{M}_h \check{G}_y$  is*

given by

$$\check{G}_v(z) + \check{H}_i(z)\check{M}_h(z)\check{G}_y(z) = \left[ \begin{array}{c|c} A_e & B_e \\ \hline C_e & 0 \end{array} \right] \begin{bmatrix} zI & -\Upsilon_e \end{bmatrix} \quad (5.73)$$

where

$$A_e := \begin{bmatrix} A_H & 0 & 0 \\ 0 & A + LC_y & 0 \\ 0 & 0 & A \end{bmatrix}, \quad B_e := \begin{bmatrix} 0 \\ B + LD_y \\ B \end{bmatrix}$$

$$C_e := [C_H \quad 0 \quad C_v], \quad \Upsilon_e := \begin{bmatrix} E + B_H F & -M_1 & M_1 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}$$

*Proof.* Define  $\Upsilon_\lambda := \begin{bmatrix} (E + B_H F) & -M_1 \\ 0 & I \end{bmatrix}$ . Using the STPBC of  $\check{H}_i(z)\check{M}_h(z)$  given in (5.71),

$$\check{G}_v(z) + \check{H}_i(z)\check{M}_h(z)\check{G}_y(z) = [I \quad \check{H}_i(z)\check{M}_h(z)] \begin{bmatrix} \check{G}_v(z) \\ \check{G}_y(z) \end{bmatrix}$$

and

$$[I \quad \check{H}_i(z)\check{M}_h(z)] = \left[ \begin{array}{cc|cc} A_H & 0 & 0 & 0 \\ 0 & A + LC_y & 0 & L \\ \hline C_H & 0 & I & 0 \end{array} \right] \begin{bmatrix} zI & -\Upsilon_\lambda \end{bmatrix}$$

$$\begin{bmatrix} \check{G}_v(z) \\ \check{G}_y(z) \end{bmatrix} = \left[ \begin{array}{c|c} A & B \\ \hline C_v & 0 \\ C_y & D_y \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix},$$

the result follows from the product of STPBC, the state-transformation using

$$T := \begin{bmatrix} I & 0 & 0 \\ 0 & I & I \\ 0 & 0 & I \end{bmatrix},$$

and multiplying the boundary condition from left by  $T$ . □

Since  $\check{G}_v + \check{H}_i\check{M}_h\check{G}_y$  is causal and in  $L^\infty$ , it is in  $H^\infty$ . By Lemma 5.3.23, this further implies  $\check{G}_v + \check{H}_i\check{M}_h\check{G}_y \in H^2$ . However, in the following theorem we show that the STPBC of the system  $\check{G}_v + \check{H}_i\check{M}_h\check{G}_y$  given in (5.73) is not minimal. It contains unobservable or uncontrollable poles that lie outside (open) unit disc of the complex plane. Since  $\check{G}_v + \check{H}_i\check{M}_h\check{G}_y \in H^\infty$ , this implies these poles must be canceled somehow. This complicates the calculation of  $H^2$  norm of  $\check{G}_v + \check{H}_i\check{M}_h\check{G}_y$  as shown in the following lemma.

**Lemma 5.4.41.** Let  $\check{G}_v$  is given by STPBC (5.42). Let the matrix functions  $\Lambda$  and  $\Gamma$  be as in Lemma 5.3.29. Also let  $A_e$ ,  $B_e$ ,  $\Upsilon_e$  and  $C_e$  be as in (5.73). Let  $\bar{C}_e$  and  $\bar{B}_e$  be any matrices which satisfy

$$\bar{C}_e^* \bar{C}_e = \int_0^h e^{A_e^* s} C_e^* C_e e^{A_e s} ds = \Lambda_{22}^*(A_e, C_e) \Lambda_{12}(A_e, C_e) \quad (5.74a)$$

$$\bar{B}_e \bar{B}_e^* = \int_0^h e^{A_e s} B_e B_e^* e^{A_e^* s} ds = \Gamma_{33}^*(A_e, B_e) \Gamma_{23}(A_e, B_e) \quad (5.74b)$$

respectively. Define for  $\tau \in [0, h]$

$$\check{D}_e \check{u} := \int_0^h C_e e^{A_e(\tau-\sigma)} B_e \mathbb{1}(\tau - \sigma) \check{u}(\sigma) d\sigma \quad (5.75)$$

then

$$\begin{aligned} \|\check{D}_e\|_{HS}^2 &= \text{tr} C_e \int_0^h \int_0^t e^{A_e s} B_e B_e^* e^{A_e^* s} ds dt C_e^* \\ &= \text{tr} C_e \Gamma_{33}^*(A_e, B_e) \Gamma_{13}(A_e, B_e) C_e^*. \end{aligned} \quad (5.76)$$

Let us partition  $M_1$  defined in (5.72) as  $[M_{1s} \quad M_{1u}]$  according to  $A_s$  and  $A_u$ , and define

$$A_{ms} := \begin{bmatrix} (E + B_H F) e^{A_H h} & -M_1 e^{(A+LC_y)h} & M_{1s} e^{A_s h} \\ 0 & e^{(A+LC_y)h} & 0 \\ 0 & 0 & e^{A_s h} \end{bmatrix}.$$

Also let us partition

$$\Upsilon_e \bar{B}_e =: \begin{bmatrix} \bar{B}_{ms} \\ \bar{B}_{mu} \end{bmatrix}, \quad \bar{C}_e =: [\bar{C}_{ms} \quad \bar{C}_{mu}]$$

according to  $A_{ms}$  and  $e^{A_u h}$ . Then

$$\|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y\|_{H^2}^2 = \frac{1}{h} \|\check{D}_e\|_{HS}^2 + \frac{1}{h} \text{tr}(\bar{C}_{ms}^* \bar{C}_{ms} W_{ec}) \quad (5.77)$$

$$= \frac{1}{h} \|\check{D}_e\|_{HS}^2 + \frac{1}{h} \text{tr}(W_{eo}(\bar{B}_{ms} - X_m \bar{B}_{mu})(\bar{B}_{ms} - X_m \bar{B}_{mu})^*) \quad (5.78)$$

where  $W_{ec}$  and  $W_{eo}$  are matrices satisfying the Lyapunov equations

$$W_{ec} = A_{ms} W_{ec} A_{ms}^* + (\bar{B}_{ms} - X_m \bar{B}_{mu})(\bar{B}_{ms} - X_m \bar{B}_{mu})^*,$$

$$W_{eo} = A_{ms}^* W_{eo} A_{ms} + \bar{C}_{ms}^* \bar{C}_{ms}$$

and  $X_m$  satisfies the Sylvester equation

$$X_m = A_{ms} X_m e^{-A_u h} + \begin{bmatrix} M_{1u} \\ 0 \\ 0 \end{bmatrix}. \quad (5.79)$$

*Proof.* See Appendix 5.C (page 192).  $\square$

$\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}$  and  $\text{proj}_{L^2 \setminus z^l \mathbb{H}^2}(\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})$  are anti-causal systems in  $L^\infty$ , therefore their conjugates are in  $H^\infty$ . Hence, their  $H^2$  norm can be obtained by Lemma 5.3.26. Since  $\|z^l \tilde{G}\|_{L^2} = \|\tilde{G}\|_{L^2}$  for a system  $\tilde{G}$ ,  $L^2$  norm of the system  $\text{proj}_{L^2 \setminus z^l \mathbb{H}^2}(\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})$  can be calculated easily.

**Lemma 5.4.42.** *Let  $A_p, B_p, \Omega_p,$  and  $C_p$  be as in Corollary 5.4.37. Let the STP-BCs of the systems  $\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}$  and  $\text{proj}_{L^2 \setminus z^l \mathbb{H}^2}(\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})$  be given by (5.69) and (5.70) respectively. Then*

1. *the  $L^2$  norm of (anti-causal) sampler  $\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}$  is given by*

$$\|\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V}\|_{L^2}^2 = \frac{1}{h} \text{tr}(\bar{B}_p \bar{B}_p^* W_{sc}) = \frac{1}{h} \text{tr}(W_{so} C_p^* C_p)$$

where  $W_{sc}$  and  $W_{so}$  are matrices satisfying the Lyapunov equations

$$\begin{aligned} W_{sc} &= \Omega_p^* e^{-A_p^* h} W_{sc} e^{-A_p h} \Omega_p + \bar{C}_p^* \bar{C}_p, \\ W_{so} &= e^{-A_p h} \Omega_p W_{so} \Omega_p^* e^{-A_p^* h} + \bar{B}_p \bar{B}_p^* \end{aligned}$$

and  $\bar{B}_p$  is any matrix satisfying

$$\begin{aligned} \bar{B}_p \bar{B}_p^* &= \int_0^h e^{-A_p \tau} B_p(\tau) B_p(\tau)^* e^{-A_p^* \tau} d\tau \\ &= P_z \Lambda_{22}^*(A_z^*, B_z^*) \Lambda_{12}(A_z^*, B_z^*) P_z^* \end{aligned}$$

in which

$$\begin{aligned} A_z &:= \begin{bmatrix} A_H & C_H^* C_v X & C_H^* C_v \\ 0 & A + L C_y & 0 \\ 0 & 0 & -A \end{bmatrix} & B_z &:= \begin{bmatrix} 0 \\ (Z_y C_y)^* \\ L Z_y^{-1} \end{bmatrix} \\ P_z &:= \begin{bmatrix} I & 0 & -P_0 \\ 0 & I & 0 \end{bmatrix} \end{aligned}$$

2. *the  $L^2$  norm of (anti-causal) sampler  $\text{proj}_{L^2 \setminus z^l \mathbb{H}^2}(\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})$  is given by*

$$\|\text{proj}_{L^2 \setminus z^l \mathbb{H}^2}(\tilde{H}_i \tilde{N}_v \tilde{N}_y - \hat{V})\|_{L^2}^2 = \frac{1}{h} \text{tr}(\bar{B}_p \bar{B}_p^* W_{pc}) = \frac{1}{h} \text{tr}(W_{po} C_{pm}^* C_{pm})$$

where  $C_{pm} := C_p(e^{-A_p h} \Omega_p)^{l+1}$  and  $W_{pc}$  and  $W_{po}$  are matrices satisfying the Lyapunov equations

$$\begin{aligned} W_{pc} &= \Omega_p^* e^{-A_p^* h} W_{pc} e^{-A_p h} \Omega_p + \bar{C}_{pm}^* \bar{C}_{pm}, \\ W_{po} &= e^{-A_p h} \Omega_p W_{po} \Omega_p^* e^{-A_p^* h} + \bar{B}_p \bar{B}_p^* \end{aligned}$$

*Proof.* See Appendix 5.C (page 193).  $\square$

Now we summarize all of the results in this section to obtain the optimal error norm  $\|\check{G}_{e,\text{opt}}\|_{\mathbb{L}^2}$ .

**Theorem 5.4.43.** *Let system  $\check{G}$  and  $\check{H}$  are given by STPBC (5.42) and (5.43) respectively. Let assumptions  $\mathcal{A}_4$ - $\mathcal{A}_9$  be satisfied. If the conditions given in Theorem 5.4.28 are satisfied then the optimal error norm  $\|\check{G}_{e,\text{opt}}\|_{\mathbb{L}^2}$  is given by*

$$\begin{aligned} \|\check{G}_{e,\text{opt}}\|_{\mathbb{L}^2}^2 &= \|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y\|_{\mathbb{H}^2}^2 - \|\check{H}_i \check{N}_v \check{N}_y - \check{V}\|_{\mathbb{L}^2}^2 \\ &\quad + \|\text{proj}_{\mathbb{L}^2 \setminus \check{z}' \mathbb{H}^2}(\check{H}_i \check{N}_v \check{N}_y - \check{V})\|_{\mathbb{L}^2}^2 \end{aligned}$$

where  $\|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y\|_{\mathbb{H}^2}$  is obtained in Lemma 5.4.41, and  $\|\check{H}_i \check{N}_v \check{N}_y - \check{V}\|_{\mathbb{L}^2}$  and  $\|\text{proj}_{\mathbb{L}^2 \setminus \check{z}' \mathbb{H}^2}(\check{H}_i \check{N}_v \check{N}_y - \check{V})\|_{\mathbb{L}^2}$  are obtained in Lemma 5.4.42.

*Proof.* The proof follows from lemmas 5.4.4, 5.4.10, 5.4.41 and 5.4.42.  $\square$

## 5.4.9 Example

In this section we consider an example to explain the theory discussed till now.

**Example 5.4.44.** *Consider the systems  $\check{G}_v$ ,  $\check{G}_y$  and  $\check{H}_{iz}$  as in Example 5.4.30. In this example we obtain strictly causal, causal, 1-causal and non-causal optimal samplers given the systems  $\check{G}_v$ ,  $\check{G}_y$  and  $\check{H}_{iz}$ . It follows from Example 5.4.30 that  $(I - \check{H}_i \check{H}_i) \check{G}_v \in \mathbb{L}^\infty$ .*

Using Lemma 5.4.10, we have  $F = 0$ ,  $Z = \frac{1}{\sqrt{h}}$  and an inner-outer factorization of  $\check{H}_{iz} := \check{H}_i \check{H}_o$  is given by

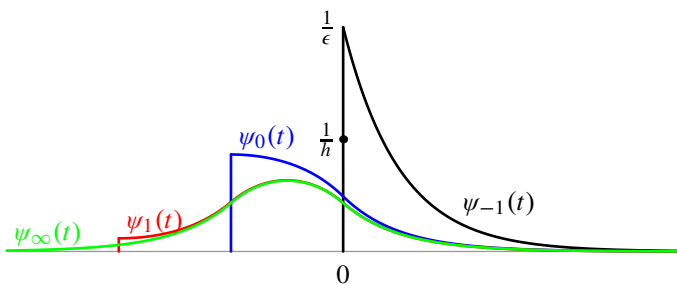
$$\check{H}_i(z) = \frac{1}{\sqrt{h}} \left[ \begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array} \right]_{\left[ \begin{array}{c} \mathcal{J}_{0^+} \\ 1 \quad 0 \end{array} \right]}, \quad \check{H}_o(z) = \sqrt{h}$$

And using Lemma 5.4.4, we have  $Z_y = \frac{1}{\epsilon}$ ,  $X = \epsilon$ , and  $L = -\frac{1}{\epsilon}$  and

$$\left[ \begin{array}{c} \check{M}_v(z) \\ \check{M}_y(z) \end{array} \right] = \left[ \begin{array}{c|c} -\frac{1}{\epsilon} & -\frac{1}{\epsilon} \\ \hline 1 & 0 \\ \frac{1}{\epsilon} & \frac{1}{\epsilon} \end{array} \right]_{\left[ \begin{array}{c} z \\ -1 \end{array} \right]}$$

Using Lemma 5.4.34,  $P(t) = h - t$ . From (5.63) and Lemma 5.4.36, we have

$$\begin{aligned} \check{V}(z) &= \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} 0 & -P(\tau) \\ \hline -\frac{1}{\sqrt{h}} & 0 \end{array} \right]_{\left[ \begin{array}{c} 0 \\ -I \end{array} \right]} \\ \check{M}_h(z) &= \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} -\frac{1}{\epsilon} & \frac{1}{\epsilon} \\ \hline -\sqrt{h} & 0 \end{array} \right]_{\left[ \begin{array}{c} z \\ -1 \end{array} \right]} \end{aligned}$$



**Figure 5.3:** Sampling functions of optimal  $-1, 0, 1, \infty$ -causal samplers in Example 5.4.44

Using Corollary 5.4.37, we have

$$\hat{H}_i^{\sim}(z)\check{N}_v(z)\check{N}_y^{\sim}(z) - \check{V}(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix}$$

where  $A_p := \begin{bmatrix} 0 & -\epsilon \\ 0 & \frac{1}{\epsilon} \end{bmatrix}$ ,  $B_p := \begin{bmatrix} P(\tau) \\ \frac{1}{\epsilon} \end{bmatrix}$ ,  $\Omega_p := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ , and  $C_p := \begin{bmatrix} -\frac{1}{\sqrt{h}} & 0 \end{bmatrix}$ .

Using Theorem 5.4.38,  $\hat{S}_{\alpha, \text{opt}} := \text{proj}_{z^l \mathbb{H}^2}(\hat{H}_i^{\sim}\check{N}_v\check{N}_y^{\sim} - \check{V})$  is given by STPBC

$$\hat{S}_{\alpha, \text{opt}}(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p (I - (ze^{-A_p h} \Omega_p)^l) & 0 \end{array} \right] \begin{bmatrix} \Omega_p & -I \end{bmatrix}$$

and the optimal relaxed causal sampler  $\hat{S}_{\text{opt}}$  is given by (5.4).

Now, we write strictly causal, causal, 1-causal and non-causal optimal samplers in a more tangible form.

1. Strictly causal optimization (i.e.  $l = -1$ )

Since  $\text{proj}_{z^{-1}\mathbb{H}^2}\{\hat{H}_i^{\sim}\check{N}_v\check{N}_y^{\sim} - \check{V}\} = 0$ , we have strictly causal sampler  $-\bar{H}_0^{-1}\bar{M}_h$  which is a cascade of the LCTI causal system  $\frac{1}{\epsilon}1/(s + \frac{1}{\epsilon})\mathbb{1}_{[0, \infty)}$  and the ideal sampler. Note that the strictly causal sampler does not depend upon  $h$  and tend to a causal impulse when  $\epsilon \rightarrow 0$ . Therefore, strictly causal  $\hat{S}_{\text{opt}}$  in time domain is a sampler given by

$$y[n] = \int_{-\infty}^{\infty} \psi_{-1}(nh - t)u(t)dt$$

where sampling function  $\psi_{-1} := \frac{1}{\epsilon}e^{-\frac{t}{\epsilon}}\mathbb{1}_{[0, \infty)}$ . See Figure 5.3.

2. Causal optimization (i.e.  $l = 0$ )

Since  $\hat{H}_i^{\sim} \in \mathbb{H}^\infty$ , therefore  $\check{V} = 0$  satisfies  $\hat{H}_i^{\sim}\check{M}_v - \check{V}\check{M}_y \in \mathbb{H}^\infty$ . Therefore,

$$\hat{S}_{\alpha, \text{opt}} = \text{proj}_{\mathbb{H}^2}(\hat{H}_i^{\sim}\check{N}_v\check{N}_y^{\sim}).$$

We find that causal  $\hat{S}_{\text{opt}}$  in time domain is a sampler given by

$$y[n] = \int_{-\infty}^{\infty} \psi_0(nh - t)u(t)dt$$

where sampling function

$$\psi_0(t) := \frac{1}{h} \begin{cases} 0 & t \leq -h \\ -\frac{1}{2}e^{\frac{t}{\epsilon}} - \frac{1}{2}e^{-\frac{t+2h}{\epsilon}} + 1 & -h < t \leq 0 \\ \frac{1}{2}e^{-\frac{t}{\epsilon}} - \frac{1}{2}e^{-\frac{t+2h}{\epsilon}} & t > 0 \end{cases}$$

See Figure 5.3.

### 3. 1-causal optimization

Similar to causal optimization, we take  $\hat{V} = 0$ . Therefore,

$$\hat{S}_{\alpha, \text{opt}} = \text{proj}_{\mathcal{J}_2^1 \text{H}^2}(\hat{H}_1 \check{N}_v \check{N}_y).$$

We find that 1-causal  $\hat{S}_{\text{opt}}$  in time domain is a sampler given by

$$y[n] = \int_{-\infty}^{\infty} \psi_1(nh - t)u(t)dt$$

where sampling function

$$\psi_1(t) := \frac{1}{h} \begin{cases} 0 & t \leq -2h \\ \frac{1}{2}(-e^{\frac{t}{\epsilon}} + e^{\frac{t+h}{\epsilon}} - e^{-\frac{4h+t}{\epsilon}} + e^{-\frac{3h+t}{\epsilon}}) & -2h < t \leq -h \\ \frac{1}{2}(-e^{\frac{t}{\epsilon}} - e^{-\frac{t+h}{\epsilon}} - e^{-\frac{4h+t}{\epsilon}} + e^{-\frac{3h+t}{\epsilon}}) + 1 & -h < t \leq 0 \\ \frac{1}{2}(e^{-\frac{t}{\epsilon}} - e^{-\frac{t+h}{\epsilon}} - e^{-\frac{4h+t}{\epsilon}} + e^{-\frac{3h+t}{\epsilon}}) & t > 0 \end{cases}$$

See Figure 5.3.

### 4. Non-causal optimization (i.e. $l = \infty$ )

In this case,

$$\hat{S}_{\alpha, \text{opt}} = \hat{H}_1 \check{N}_v \check{N}_y.$$

We find that non-causal  $\hat{S}_{\text{opt}}$  in time domain is a sampler given by

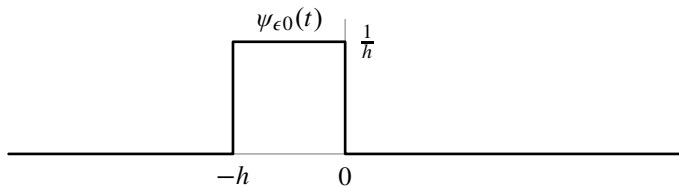
$$y[n] = \int_{-\infty}^{\infty} \psi_{\infty}(nh - t)u(t)dt$$

where sampling function

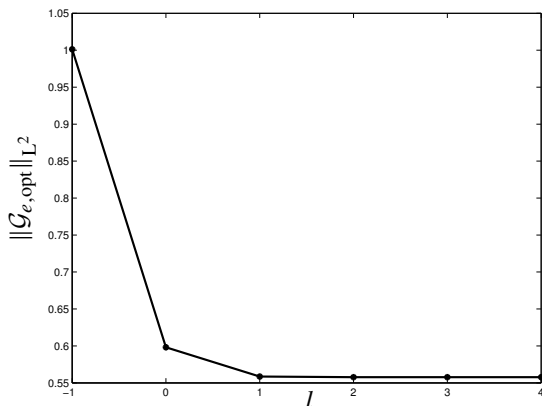
$$\psi_{\infty}(t) := \frac{1}{h} \begin{cases} -\frac{1}{2}e^{\frac{t}{\epsilon}} + \frac{1}{2}e^{\frac{t+h}{\epsilon}} & t \leq -h \\ -\frac{1}{2}e^{\frac{t}{\epsilon}} - \frac{1}{2}e^{-\frac{t+h}{\epsilon}} + 1 & -h < t \leq 0 \\ \frac{1}{2}e^{-\frac{t}{\epsilon}} - \frac{1}{2}e^{-\frac{t+h}{\epsilon}} & t > 0 \end{cases}$$

See Figure 5.3.





**Figure 5.4:** Sampling function of optimal  $l$ -causal (where  $l \geq 0$ ) sampler in Example 5.4.44 when  $\epsilon \rightarrow 0$ .



**Figure 5.5:** Optimal error  $\|\mathcal{G}_{e,\text{opt}}\|_{L^2} := \|\mathcal{G}_v - \mathcal{H}\mathcal{S}_{\text{opt}}\mathcal{G}_y\|_{L^2}$  for different  $l$  and  $\epsilon = 0.5$ .

If  $\epsilon \rightarrow 0$  in this example, then  $\psi_0(t)$ ,  $\psi_1(t)$ ,  $\psi_\infty(t)$  converge to the following

$$\psi_{\epsilon 0}(t) = \begin{cases} 1/h & -h < t \leq 0 \\ 0 & \text{elsewhere} \end{cases}$$

see Figure 5.4.

Figure 5.5 shows the optimal error  $\|\mathcal{G}_{e,\text{opt}}\|_{L^2} := \|\mathcal{G}_v - \mathcal{H}\mathcal{S}_{\text{opt}}\mathcal{G}_y\|_{L^2}$  for different optimal relaxed causal samplers. As expected the optimal error decreases with increasing non-causality.

## 5.5 Conclusions

In this chapter we obtained a stable and optimal  $l$ -causal sampler given hold and a LCTI model  $\mathcal{G}$ . The presence of hold  $\mathcal{H}$  complicates the question of existence of such a sampler when  $\mathcal{G}$  is unstable. We also provided the conditions of existence of optimal  $l$ -causal samplers, in (lifted) frequency domain as well as in state space

with two point boundary condition (STPBC). We also gave the optimal  $l$ -causal sampler in STPBC. Use of STPBC allows an easy and clear framework to solve our problem.

## 5.A Proofs of the results in Section 5.2

*Proof of Proposition 5.2.2.* We first prove that these three criteria are necessary. Let us assume that there exists an  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e := \check{G}_v - \check{H}\check{S}\check{G}_y \in L^\infty$ .

*Condition 1:* Since  $R := \begin{bmatrix} \check{H}_i^\sim \\ I - \check{H}_i \check{H}_i^\sim \end{bmatrix}$  satisfies  $R^\sim R = I$ , we have

$$\begin{aligned} \|\check{G}_e\|_{L^\infty} &= \|\check{G}_v - \check{H}\check{S}\check{G}_y\|_{L^\infty} = \|R\check{G}_v - R\check{H}\check{H}_0\check{S}\check{G}_y\|_{L^\infty} \\ &= \left\| \begin{bmatrix} \check{H}_i^\sim \check{G}_v - \check{H}_0\check{S}\check{G}_y \\ \Pi_{\check{H}} \check{G}_v \end{bmatrix} \right\|_{L^\infty} \end{aligned} \quad (5.80)$$

By assumption  $\check{G}_e \in L^\infty$ , so  $\|\Pi_{\check{H}} \check{G}_v\|_{L^\infty}$  is finite.

*Condition 2:* Since the hold  $\check{H}$  is stable and casual, we have that  $\check{K} := \check{H}\check{S} \in z^l H^\infty$  and  $\check{G}_v - \check{K}\check{G}_y \in L^\infty$ . According to [40, proposition 2.1], the existence of a  $\check{K} \in z^l H^\infty$  that renders  $\check{G}_v - \check{K}\check{G}_y \in L^\infty$  is equivalent to existence of a factorization over  $H^\infty$  of  $\check{G}$  of the form (5.1) with  $\check{M}_y, \check{N}_y$  left coprime. Note that the coprime factorization is over  $H^\infty$  not  $z^l H^\infty$  as we expect. This is related to the causality of  $\check{G}$  (see [29, remark 20.4] for detail).

*Condition 3:* Let  $\check{V} := (\check{H}_i^\sim \check{M}_v + \check{H}_0\check{S})\check{M}_y^{-1}$ . Then,  $\check{H}_i^\sim \check{M}_v - \check{V}\check{M}_y = -\check{H}_0\check{S} \in z^l H^\infty$  as required. Hence,

$$\check{V}\check{M}_y = \check{H}_i^\sim \check{M}_v + \check{H}_0\check{S} \in L^\infty.$$

Also,

$$\check{V}\check{N}_y = (\check{H}_i^\sim \check{M}_v + \check{H}_0\check{S})\check{G}_y \in L^\infty$$

because

$$\begin{aligned} &\check{H}_i^\sim \check{G}_v - \check{H}_0\check{S}\check{G}_y \in L^\infty, && \text{follows from (5.80)} \\ \Rightarrow &\check{H}_i^\sim \check{N}_v - \check{H}_i^\sim \check{M}_v \check{G}_y - \check{H}_0\check{S}\check{G}_y \in L^\infty \\ &\Rightarrow \check{H}_i^\sim \check{M}_v \check{G}_y + \check{H}_0\check{S}\check{G}_y \in L^\infty, && \text{as } \check{H}_i^\sim \check{N}_v \in L^\infty. \end{aligned}$$

Here, we used  $\check{G}_v = \check{N}_v - \check{M}_v \check{G}_y$  and  $\check{G}_y = \check{M}_y^{-1} \check{N}_y$  which follows from (5.1). Therefore, we have

$$\check{V} \begin{bmatrix} \check{N}_y \\ \check{M}_y \end{bmatrix} \in L^\infty$$

This further implies that  $\check{V} \in L^\infty$  as  $\begin{bmatrix} \check{N}_y \\ \check{M}_y \end{bmatrix}$  is right invertible in  $H^\infty$ .

Now assume that Conditions 1-3 are satisfied. We show that there exists a sampler  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e \in L^\infty$ . From (5.80), a sampler  $\check{S} \in z^l H^\infty$  achieves  $\check{G}_e \in L^\infty$  iff  $\|\Pi_{\check{H}} \check{G}_v\|_{L^\infty} < \infty$  and  $\|\check{H}_i \check{G}_v - \check{H}_o \check{S} \check{G}_y\|_{L^\infty} < \infty$ . From Condition 2 and 3, it is clear that we have a  $\check{V} \in L^\infty$  such that  $\check{M}_h := \check{H}_i \check{M}_v - \check{V} \check{M}_y \in z^l H^\infty$  and  $\check{G}_v = \check{N}_v - \check{M}_v \check{G}_y$  and  $\check{G}_y = \check{M}_y^{-1} \check{N}_y$ . Now,  $\check{S} := -\check{H}_o^{-1} \check{M}_h$  does the job because  $\check{S} \in z^l H^\infty$  and

$$\begin{aligned} \check{H}_i \check{G}_v - \check{H}_o \check{S} \check{G}_y &= \check{H}_i \check{N}_v + (-\check{H}_i \check{M}_v + \check{M}_h) \check{G}_y \\ &= \check{H}_i \check{N}_v - \check{V} \check{M}_y \check{G}_y = \check{H}_i \check{N}_v - \check{V} \check{N}_y \in L^\infty. \end{aligned}$$

□

*Proof of Lemma 5.2.3.* We have

$$[\check{G}_y \quad I] = \check{M}_y^{-1} [\check{N}_y \quad \check{M}_y]$$

The right invertibility of  $[\check{N}_y \quad \check{M}_y]$  implies that  $\check{M}_y^{-1}$  is causal if  $\check{G}_y$  is causal.

Now, it is given that  $\check{H}_i \check{M}_v - \check{V}_1 \check{M}_y \in z^l H^\infty$  and  $\check{H}_i \check{M}_v - \check{V}_2 \check{M}_y \in z^l H^\infty$ . This implies that  $(\check{V}_2 - \check{V}_1) \check{M}_y \in z^l H^\infty$  and so that  $\check{V}_2 - \check{V}_1$  is  $l$ -causal as  $\check{M}_y^{-1}$  is causal.

Since  $\check{V}_1$  and  $\check{V}_2$  are in  $L^\infty$  and are samplers, this implies  $\check{V}_2 - \check{V}_1 \in L^2$  [30]. So,  $\check{V}_2 - \check{V}_1 \in L^2$  and  $l$ -causal. It means  $\check{V}_2 - \check{V}_1 \in z^l H^2$ . Hence,  $\text{proj}_{L^2 \setminus z^l H^2}(\check{V}_1 - \check{V}_2) = 0$ . □

*Proof of Proposition 5.2.4.* Given condition 1 of Proposition 5.2.2, it follows from Equation (5.80) that  $\|\check{G}_e\|_{L^\infty}$  is finite iff  $\check{G}_{eh} := \check{H}_i \check{G}_v - \check{H}_o \check{S} \check{G}_y \in L^\infty$ .

Given condition 1-3 of Proposition 5.2.2,  $\check{N}_h := \check{H}_i \check{N}_v - \check{V} \check{N}_y \in L^\infty$ . We now show that every solution  $\check{S} \in z^l H^\infty$  has the form (5.2). Let  $\check{S} \in z^l H^\infty$  such that  $\check{G}_e \in z^l H^\infty$  (see Remark 5.2.1). Let  $\check{G}_{eh0} := \check{H}_i \check{G}_v - \check{H}_o \check{S} \check{G}_y$ . Clearly  $\check{G}_{eh0} \in L^\infty$  and  $\check{H}_i \check{G}_v = \check{H}_i \check{N}_v - \check{H}_i \check{M}_v \check{G}_y = \check{N}_h - \check{M}_h \check{G}_y$  by Proposition 5.2.2. Therefore,

$$\begin{aligned} \check{G}_{eh0} &= \check{H}_i \check{G}_v - \check{H}_o \check{S} \check{G}_y \\ \check{G}_{eh0} - \check{N}_h &= -\check{M}_h \check{G}_y - \check{H}_o \check{S} \check{G}_y \end{aligned}$$

Since  $-\check{M}_h \check{G}_y - \check{H}_o \check{S} \check{G}_y$  is  $l$ -causal (as  $\check{M}_h + \check{H}_o \check{S} \in z^l H^\infty$ ) and  $\check{G}_{eh0} - \check{N}_h \in L^\infty$ , we have  $\check{G}_{eh0} - \check{N}_h \in z^l H^\infty$ . Hence,

$$\check{M}_h \check{G}_y + \check{H}_o \check{S} \check{G}_y \in z^l H^\infty$$

Using  $\check{G}_y = \check{M}_y^{-1} \check{N}_y$  and  $\check{M}_h + \check{H}_o \check{S} \in z^l H^\infty$ , the above implies

$$(\check{M}_h + \check{H}_o \check{S}) [\check{G}_y \quad I] = (\check{M}_h + \check{H}_o \check{S}) \check{M}_y^{-1} [\check{N}_y \quad \check{M}_y] \in z^l H^\infty$$

This implies  $(\dot{M}_h + \bar{H}_0 \dot{S}) \dot{M}_y^{-1} \in z^l H^\infty$  as  $[\check{N}_y \quad \check{M}_y]$  is right invertible in  $H^\infty$ . Hence, we have that the  $\dot{S}_\alpha$  defined as

$$\dot{S}_\alpha := (\dot{M}_h + \bar{H}_0 \dot{S}) \dot{M}_y^{-1}$$

is in  $z^l H^\infty$ . From this  $\dot{S}_\alpha$ , the sampler  $\dot{S}$  follows as

$$\dot{S} = \bar{H}_0^{-1}(-\dot{M}_h + \dot{S}_\alpha \dot{M}_y)$$

On the other hand, if  $\dot{S} := \bar{H}_0^{-1}(\dot{S}_\alpha \dot{M}_y - \dot{M}_h)$  where  $\dot{S}_\alpha \in z^l H^\infty$  then clearly  $\dot{S} \in z^l H^\infty$ ,  $\check{G}_{eh} := \check{H}_i \check{G}_v - \bar{H}_0 \dot{S} \check{G}_y = \check{N}_h - \dot{S}_\alpha \check{N}_y \in L^\infty$  and  $\check{G}_e \in L^\infty$  (given all conditions of Proposition 5.2.2 are satisfied).

In the end using  $\dot{S} := \bar{H}_0^{-1}(\dot{S}_\alpha \dot{M}_y - \dot{M}_h)$ , we have

$$\begin{aligned} \check{G}_e &:= \check{G}_v - \check{H} \dot{S} \check{G}_y = \check{G}_v - \check{H}_i (\dot{S}_\alpha \dot{M}_y - \dot{M}_h) \check{G}_y \\ &= \check{G}_v + \check{H}_i \dot{M}_h \check{G}_y - \check{H}_i \dot{S}_\alpha \check{N}_y \end{aligned}$$

Note that  $\dot{M}_h := \check{H}_i \check{M}_v - \check{V} \check{M}_y$  depends upon  $\check{V}$  which is not fixed. Therefore,  $\dot{S}$  and  $\check{G}_e$  depend upon  $\check{V}$  and  $\dot{S}_\alpha$ . This may (wrongly) suggests that the parameterization of  $\dot{S}$  and  $\check{G}_e$  given in (5.2) and (5.3) is in two parameters. However, we will show now that the parameterization of  $\dot{S}$  and  $\check{G}_e$  is in single parameter. Any  $\check{V} \in L^\infty$  (also implies  $\check{V} \in L^2$  [30]) satisfying  $\dot{M}_h := \check{H}_i \check{M}_v - \check{V} \check{M}_y \in z^l H^\infty$  can be represented as  $\check{V} = \text{proj}_{L^2 \setminus z^l H^2} \check{V} + \text{proj}_{z^l H^2} \check{V}$ . Therefore,

$$\begin{aligned} \dot{S} &:= \bar{H}_0^{-1}(\dot{S}_\alpha \dot{M}_y - \dot{M}_h) \\ &= \bar{H}_0^{-1}(\dot{S}_\alpha \dot{M}_y - \check{H}_i \check{M}_v + \check{V} \check{M}_y) \\ &= \bar{H}_0^{-1}(\dot{S}_\alpha \dot{M}_y - \check{H}_i \check{M}_v + (\text{proj}_{L^2 \setminus z^l H^2} \check{V} + \text{proj}_{z^l H^2} \check{V}) \check{M}_y) \\ &= \bar{H}_0^{-1}((\dot{S}_\alpha + \text{proj}_{z^l H^2} \check{V}) \dot{M}_y - \check{H}_i \check{M}_v + \text{proj}_{L^2 \setminus z^l H^2} \check{V} \check{M}_y) \end{aligned}$$

Since  $\text{proj}_{L^2 \setminus z^l H^2} \check{V}$  is unique by Lemma 5.2.3,  $\dot{S}$  is still parameterized by a new single parameter  $\dot{S}_\alpha + \text{proj}_{z^l H^2} \check{V} \in z^l H^\infty$ .  $\square$

*Proof of Proposition 5.2.5.*  $\check{N}_y \check{N}_y^\sim$  is stable (by construction) and rational (Assumption  $\mathcal{A}_2$ ). Also  $\check{N}_y \check{N}_y^\sim$  has no unit circle zeros (Assumption  $\mathcal{A}_4$ ). Therefore,  $\check{N}_y \check{N}_y^\sim$  has a spectral co-factorization  $\check{W} \check{W}^\sim$  where  $\check{W}$  is bistable and bicausal in  $H^\infty$  (see e.g. [68]). This means  $\check{W}^{-1} \check{N}_y$  is co-inner. As  $\check{G}_y = \dot{M}_y^{-1} \check{N}_y = (\check{W}^{-1} \check{M}_y)^{-1} (\check{W}^{-1} \check{N}_y)$ , we have a coprime factorization of  $\check{G}_y$  with coprime factors  $\check{W}^{-1} \check{M}_y$  and co-inner  $\check{W}^{-1} \check{N}_y$ . In the rest of the proof we assume that  $\check{N}_y$  is co-inner without loss of generality.

By Proposition 5.2.2, there exists a solution to the stabilization problem i.e. there exists an  $\dot{S} \in z^l H^\infty$  such that  $\check{G}_e \in L^\infty$  iff all conditions of Proposition 5.2.2 are satisfied.

As  $R := \begin{bmatrix} \tilde{H}_i \\ I - \tilde{H}_i \tilde{H}_i \end{bmatrix}$  satisfies  $R \tilde{R} = I$ , we have

$$\|\check{G}_v - \check{H} \check{S} \check{G}_y\|_{L^2} = \|R \check{G}_v - R \check{H}_i \check{H}_0 \check{S} \check{G}_y\|_{L^2} = \left\| \begin{bmatrix} \check{H}_i \check{G}_v - \check{H}_0 \check{S} \check{G}_y \\ \Pi_{\check{H}} \check{G}_v \end{bmatrix} \right\|_{L^2}$$

Therefore,  $\|\check{G}_v - \check{H} \check{S} \check{G}_y\|_{L^2}$  is finite iff  $\|\check{H}_i \check{G}_v - \check{H}_0 \check{S} \check{G}_y\|_{L^2}$  and  $\|\Pi_{\check{H}} \check{G}_v\|_{L^2}$  are finite. Since  $\check{H}_i \check{G}_v - \check{H}_0 \check{S} \check{G}_y \in L^\infty$  (by (5.80)) and is a sampler, it belongs to  $L^2$  also (by Lemma 2.4.6). Therefore  $\check{G}_e \in L^2$  iff  $\Pi_{\check{H}} \check{G}_v \in L^2$ .

Since  $\check{S} \in z^l H^\infty$  and all conditions of Proposition 5.2.2 are satisfied, by Proposition 5.2.4 we can parameterize  $\check{S}$  and  $\check{G}_e \in L^\infty$  in terms of  $\check{S}_\alpha \in z^l H^\infty$  as given in (5.2) and (5.3) respectively. Define

$$\check{S}_o := \arg \inf_{\check{S}_\alpha \in z^l H^\infty} \|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y - \check{H}_i \check{S}_\alpha \check{N}_y\|.$$

Since  $\check{G}_e \in L^2$ , we use projections to say that  $\check{S}_o$  must satisfy the following:

$$\begin{aligned} \langle \check{G}_e, \check{H}_i \check{S}_\alpha \check{N}_y \rangle &= \langle \check{G}_v + \check{H}_i \check{M}_h \check{G}_y - \check{H}_i \check{S}_o \check{N}_y, \check{H}_i \check{S}_\alpha \check{N}_y \rangle \\ &= \langle \check{H}_i \check{N}_v \check{N}_y^\sim - \check{V} - \check{S}_o, \check{S}_\alpha \rangle = 0 \end{aligned}$$

for all  $\check{S}_\alpha \in z^l H^2$ . This can be achieved if we take  $\check{S}_o = \check{S}_{\alpha, \text{opt}}$ . In particular

$$\langle \check{G}_{e, \text{opt}}, \check{H}_i \check{S}_{\alpha, \text{opt}} \check{N}_y \rangle = \langle \check{G}_v + \check{H}_i \check{M}_h \check{G}_y - \check{H}_i \check{S}_{\alpha, \text{opt}} \check{N}_y, \check{H}_i \check{S}_{\alpha, \text{opt}} \check{N}_y \rangle = 0,$$

Therefore (by Pythagoras theorem),

$$\begin{aligned} \|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y\|_{L^2}^2 &= \|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y - \check{H}_i \check{S}_{\alpha, \text{opt}} \check{N}_y\|_{L^2}^2 + \|\check{H}_i \check{S}_{\alpha, \text{opt}} \check{N}_y\|_{L^2}^2 \\ &= \|\check{G}_{e, \text{opt}}\|_{L^2}^2 + \|\check{H}_i \check{S}_{\alpha, \text{opt}} \check{N}_y\|_{L^2}^2 \end{aligned}$$

Since  $\check{H}_i$  is inner and  $\check{N}_y$  is co-inner, we have (5.6).

Since  $\check{H}_i \check{N}_v \check{N}_y^\sim - \check{V} \in L^\infty \cap L^2$ , the  $\check{S}_{\alpha, \text{opt}}$  of (5.5) is in  $L^\infty \cap L^2$  as well. Hence  $\check{S}_{\text{opt}} \in L^\infty \cap L^2$ .  $\square$

## 5.B Proofs of the results in Section 5.3

*Proof of (5.15).* . We first write  $K_G(\tau, \sigma)$  in such a form where we do not have to worry about order of  $\sigma$  and  $\tau$  as required in (5.14). Note that

$$\begin{aligned} C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} \Omega e^{-A\sigma} B(\sigma) &= C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} (\Omega + \Upsilon e^{Ah} - \Upsilon e^{Ah}) e^{-A\sigma} B(\sigma) \\ &= C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} (\Xi_{\mathcal{G}} - \Upsilon e^{Ah}) e^{-A\sigma} B(\sigma) \\ &= C(\tau) e^{A(\tau-\sigma)} B(\sigma) - C(\tau) e^{A\tau} \Xi_{\mathcal{G}}^{-1} \Upsilon e^{A(h-\sigma)} B(\sigma) \end{aligned}$$

Similarly,

$$C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Upsilon e^{A(h-\sigma)}B(\sigma) = C(\tau)e^{A(\tau-\sigma)}B(\sigma) - C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Omega e^{-A\sigma}B(\sigma)$$

Hence,

$$\begin{aligned} \int_0^h K_{\mathcal{G}}(\tau, \sigma)u(\sigma)d\sigma &= \int_0^\tau C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Omega e^{-A\sigma}B(\sigma)u(\sigma)d\sigma \\ &\quad - \int_\tau^h C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Upsilon e^{A(h-\sigma)}B(\sigma)u(\sigma)d\sigma \\ &= \int_0^\tau C(\tau)e^{A(\tau-\sigma)}B(\sigma)u(\sigma)d\sigma \\ &\quad - \int_0^\tau C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Upsilon e^{A(h-\sigma)}B(\sigma)u(\sigma)d\sigma \\ &\quad - \int_\tau^h C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Upsilon e^{A(h-\sigma)}B(\sigma)u(\sigma)d\sigma \\ &= \int_0^\tau C(\tau)e^{A(\tau-\sigma)}B(\sigma)u(\sigma)d\sigma \\ &\quad - \int_0^h C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Upsilon e^{A(h-\sigma)}B(\sigma)u(\sigma)d\sigma. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_0^h K_{\mathcal{G}}(\tau, \sigma)u(\sigma)d\sigma &= - \int_\tau^h C(\tau)e^{A(\tau-\sigma)}B(\sigma)u(\sigma)d\sigma \\ &\quad + \int_0^h C(\tau)e^{A\tau}\Xi_{\mathcal{G}}^{-1}\Omega e^{-A\sigma}B(\sigma)u(\sigma)d\sigma. \end{aligned}$$

Therefore, using the above expression of  $K_{\mathcal{G}}(\tau, \sigma)$ , the output  $y(\tau)$  can be written as (5.15).  $\square$

*Proof of Lemma 5.3.6.* For any  $y = \mathcal{G}u$ ,  $y_1 \in L^2[0, h)$ , we have

$$\begin{aligned} \langle y_1, y \rangle &= \langle y_1, \mathcal{G}u \rangle \\ &= \int_0^h \left\langle y_1(\tau), \left( Du(\tau) + \int_0^h K_{\mathcal{G}}(\tau, \sigma)u(\sigma)d\sigma \right) \right\rangle d\tau \\ &= \int_0^h \langle y_1(\tau), Du(\tau) \rangle d\tau + \int_0^h \left\langle y_1(\tau), \int_0^h K_{\mathcal{G}}(\tau, \sigma)u(\sigma)d\sigma \right\rangle d\tau \\ &= \int_0^h \langle D^*y_1(\tau), u(\tau) \rangle d\tau + \int_0^h \int_0^h \langle y_1(\tau), K_{\mathcal{G}}(\tau, \sigma)u(\sigma) \rangle d\sigma d\tau \\ &= \int_0^h \langle D^*y_1(\tau), u(\tau) \rangle d\tau + \int_0^h \int_0^h \langle \overline{K_{\mathcal{G}}(\tau, \sigma)}y_1(\tau), u(\sigma) \rangle d\sigma d\tau \end{aligned}$$

$$\begin{aligned}
&= \int_0^h \langle D^* y_1(\tau), u(\tau) \rangle d\tau + \int_0^h \left\langle \int_0^h \overline{K_{\mathcal{G}}(\tau, \sigma)} y_1(\tau) d\tau, u(\sigma) \right\rangle d\sigma \\
&= \int_0^h \langle D^* y_1(\tau), u(\tau) \rangle d\tau + \int_0^h \left\langle \int_0^h \overline{K_{\mathcal{G}}(\sigma, \tau)} y_1(\sigma) d\sigma, u(\tau) \right\rangle d\tau \\
&= \int_0^h \left\langle D^* y_1(\tau) + \int_0^h \overline{K_{\mathcal{G}}(\sigma, \tau)} y_1(\sigma) d\sigma, u(\tau) \right\rangle d\tau \\
&= \langle \mathcal{G}^* y_1, u \rangle
\end{aligned}$$

where  $\mathcal{G}^* y_1(\tau) := D^* y_1(\tau) + \int_0^h \int_0^h \overline{K_{\mathcal{G}}(\sigma, \tau)} y_1(\sigma) d\sigma$ . Clearly,

$$\overline{K_{\mathcal{G}}(\sigma, \tau)} = \begin{cases} B^*(\sigma) e^{-A^* \sigma} \Omega^* \Xi_{\mathcal{G}}^{-*} e^{A^* \tau} C^*(\tau) & 0 \leq \tau < \sigma \leq h \\ -B^*(\sigma) e^{A^*(h-\sigma)} \Upsilon^* \Xi_{\mathcal{G}}^{-*} e^{A^* \tau} C^*(\tau) & 0 \leq \sigma < \tau \leq h \end{cases}$$

where  $\Xi_{\mathcal{G}}^* = \Omega^* + e^{A^* h} \Upsilon^*$ . Hence,

$$\overline{K_{\mathcal{G}}(\sigma, \tau)} = \begin{cases} B^*(\sigma) e^{-A^* \sigma} \Omega^* \Xi_{\mathcal{G}}^{-*} e^{A^* h} e^{-A^*(h-\tau)} C^*(\tau) & 0 \leq \tau < \sigma \leq h \\ -B^*(\sigma) e^{-A^* \sigma} e^{A^* h} \Upsilon^* \Xi_{\mathcal{G}}^{-*} e^{A^* \tau} C^*(\tau) & 0 \leq \sigma < \tau \leq h \end{cases}$$

Now,

$$\mathcal{G}^* = \left[ \begin{array}{c|c} -A^* & C^* \\ \hline -B^* & D^* \end{array} \right] \begin{bmatrix} e^{A^* h} \Upsilon^* \Xi_{\mathcal{G}}^{-*} & \Omega^* \Xi_{\mathcal{G}}^{-*} e^{A^* h} \end{bmatrix}$$

as

$$\Xi_{\mathcal{G}}^* := e^{A^* h} \Upsilon^* \Xi_{\mathcal{G}}^{-*} + \Omega^* \Xi_{\mathcal{G}}^{-*} e^{A^* h} (e^{-A^* h}) = \Xi_{\mathcal{G}}^* \Xi_{\mathcal{G}}^{-*} = I.$$

This implies  $\mathcal{G}^*$  is a well-posed system. By Corollary 5.3.4, the above  $\mathcal{G}^*$  is equivalent to the following system

$$\mathcal{G}_n^* = \left[ \begin{array}{c|c} -A^* & C^* \\ \hline -B^* & D^* \end{array} \right] \begin{bmatrix} \Upsilon_d^* & \Omega_d^* \end{bmatrix}$$

where  $[\Upsilon_d^* \quad \Omega_d^*]$  has full row rank, iff there exists an invertible matrix  $S$  such that

$$\Upsilon_d^* = S e^{A^* h} \Upsilon^* \Xi_{\mathcal{G}}^{-*} \quad (5.81)$$

$$\Omega_d^* = S \Omega^* \Xi_{\mathcal{G}}^{-*} e^{A^* h}. \quad (5.82)$$

Now,

$$\Upsilon_d^* + \Omega_d^* e^{-A^* h} = S (e^{A^* h} \Upsilon^* \Xi_{\mathcal{G}}^{-*} + \Omega^* \Xi_{\mathcal{G}}^{-*}) = S$$

To show that  $S$  is invertible, consider  $[\Upsilon_d^* \quad \Omega_d^*]$  which has full row rank. This implies there exists a non-zero matrix  $K$  such that

$$[\Upsilon_d^* \quad \Omega_d^*] K = I$$

This implies

$$S \begin{bmatrix} e^{A^*h} \Upsilon^* \Xi_{\mathcal{G}}^{-*} & \Omega^* \Xi_{\mathcal{G}}^{-*} e^{A^*h} \end{bmatrix} K = I$$

Since  $S$  is square matrix, the above equation says that  $S$  is invertible.

Substituting the value of  $S$  in (5.81), we have

$$\Upsilon_d^* \Omega^* = \Omega_d^* \Upsilon^*.$$

Therefore,  $\mathcal{G}_n^* \equiv \mathcal{G}^*$  iff  $\Omega \Upsilon_d = \Upsilon \Omega_d$ . We now show the existence of such  $\Upsilon_d$  and  $\Omega_d$ . Since  $[\Omega \ \Upsilon] \in \mathbb{C}^{n \times 2n}$  has maximal row rank,  $\dim(\text{Ker} [\Omega \ \Upsilon]) = n$ . Hence there exists a matrix  $M \in \text{Ker} [\Omega \ \Upsilon] \subset \mathbb{C}^{2n \times n}$  such that

$$[\Omega \ \Upsilon] M = 0$$

Partitioning  $M$ , we have

$$\begin{bmatrix} \Upsilon_d \\ -\Omega_d \end{bmatrix} = M$$

□

*Proof of Lemma 5.3.16.* If  $\check{u} \in L^2(\mathbb{T})$ , we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \int_0^h \|\check{u}(e^{j\theta}; \sigma)\|_2^2 d\sigma d\theta < \infty$$

This implies at almost all  $\theta$ ,

$$\|\check{u}(e^{j\theta})\|_2^2 := \int_0^h \|\check{u}(e^{j\theta}; \sigma)\|_2^2 d\sigma < \infty$$

Using [15, theorem 0.10], we have

$$\|\check{f}(e^{j\theta})\|_2^2 := \int_0^h \|\check{f}(e^{j\theta}; \tau)\|_2^2 d\tau \leq C \|\check{u}(e^{j\theta})\|_2^2$$

at almost all  $\theta$ . Now, we have

$$\begin{aligned} \|f\|_2^2 &= \|\check{f}\|_2^2 := \frac{1}{2\pi} \int_{-\pi}^{\pi} \|\check{f}(e^{j\theta})\|_2^2 d\theta \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} C \|\check{u}(e^{j\theta})\|_2^2 d\theta \\ &= C \|\check{u}\|_2^2 \\ &= C \|u\|_2^2 \end{aligned}$$



Clearly,

$$\frac{\|f\|_2^2}{\|u\|_2^2} \leq C$$

Therefore  $\mathcal{T} \in L^\infty$ .

Using Lemma 2.5.8,  $\mathcal{T}$  is causal. Therefore,  $\mathcal{T} \in L^\infty$  implies that  $\mathcal{T} \in H^\infty$ .  $\square$

*Proof of Lemma 5.3.17.* Since  $B \in C^2([0, h), \mathbb{C}^{n \times m})$ , we have that the integral  $\int_0^h B(\sigma)B(\sigma)^* d\sigma$  is well-defined in  $\mathbb{C}^{n \times n}$ . Therefore,

$$\int_0^h e^{A(h-\sigma)} B(\sigma)B(\sigma)^* e^{A^*(h-\sigma)} d\sigma$$

is a well-defined matrix (note that  $e^{A(h-\sigma)}$  is a bounded function). Similarly, it can be proved that

$$\int_0^h e^{A^*\tau} C^*(\tau)C(\tau)e^{A\tau} d\tau$$

is a well-defined matrix. Since both of the integral above are non-negative, there exist matrices  $\bar{B}$  and  $\bar{C}$  of the form given in (5.29).

Since  $D$  is a constant matrix, we have that

$$\check{G} \in H^\infty \iff \check{G}_1 \in H^\infty$$

where  $\check{G}_1$  is a system with STPBC

$$\check{G}_1(z) := \left[ \begin{array}{c|c} A & B(\sigma) \\ \hline C(\tau) & 0 \end{array} \right] \begin{bmatrix} \Omega(z) & \Upsilon(z) \end{bmatrix}$$

Now using (5.15), we have

$$\check{G}_1 \in H^\infty \iff \check{X} - \check{Y} \in H^\infty$$

where

$$\check{X}(z)\check{u}(z) := C(\tau) \int_0^\tau e^{A(\tau-\sigma)} B(\sigma)\check{u}(z; \sigma) d\sigma, \quad \tau \in [0, h)$$

$$\check{Y}(z)\check{u}(z) := C(\tau)e^{A\tau}(\Omega(z) + \Upsilon(z)e^{Ah})^{-1}\Upsilon(z) \int_0^h e^{A(h-\sigma)} B(\sigma)\check{u}(z; \sigma) d\sigma.$$

As  $e^{A\tau}$  is a continuous function, therefore,  $e^{A(\tau-\sigma)}\mathbb{1}(\tau-\sigma)$  is bounded for all  $\tau \in [0, h)$  and  $\sigma \in [0, h)$ . Here the step function  $\mathbb{1}(\tau-\sigma)$  is 1 whenever  $\tau \geq \sigma$  otherwise it is zero. Along with boundedness of  $B(\sigma)\forall \sigma \in [0, h)$  and  $C(\tau)\forall \tau \in [0, h)$  this implies  $\check{X} \in H^\infty$  by Lemma 5.3.16. Hence,

$$\check{G}_1 \in H^\infty \iff \check{Y} \in H^\infty$$

To check whether  $\check{Y} \in H^\infty$ , define a zero order hold  $\check{C}$  as

$$\check{y}(z) = \check{C}(z)\check{x}(z) : \check{y}(z; \tau) = C(\tau)e^{A\tau}\check{x}(z) \quad (5.83)$$

Clearly,  $\check{C}^*\check{C}$  is a static discrete system given by

$$\check{C}^*\check{C} = \bar{C}^*\bar{C}$$

Define a zero order hold  $\check{V}_L := \check{C}\bar{C}^+$  where  $\bar{C}^+$  is the pseudo-inverse of the matrix  $\bar{C}$ . Clearly  $\check{V}_L^*\check{V}_L = \bar{C}\bar{C}^+$  i.e it is an orthogonal projection onto  $\text{Im } \bar{C}$ . Hence,  $\check{V}_L$  is stable. Since  $\check{V}_L$  is causal (see Lemma 2.5.8), we have that  $\check{V}_L \in H^\infty$ . Also, we have  $\check{V}_L\check{V}_L^*\check{C} = \check{C}$ .

Define a sampler  $\check{B}$  as

$$\bar{y}(z) = \check{B}(z)\check{x}(z) : \bar{y}(z) = \int_0^h e^{A(h-\sigma)}B(\sigma)\check{x}(z; \sigma)d\sigma \quad (5.84)$$

Also, define a sampler  $\check{V}_R := \bar{B}^+\check{B}$  where  $\bar{B}^+$  is the pseudo-inverse of the matrix  $\bar{B}$ . Clearly,  $\check{V}_R\check{V}_R^* = \bar{B}^+\bar{B}$  i.e. it is an orthogonal projection onto  $(\text{Ker } \bar{B})^\perp$ . Hence,  $\check{V}_R$  is stable. Since  $\check{V}_R$  is causal (see Lemma 2.5.8), we have that  $\check{V}_R \in H^\infty$ . Also, we have  $\check{B}\check{V}_R^*\check{V}_R = \check{B}$ .

Therefore, using  $\check{V}_L\check{V}_L^*\check{C} = \check{C}$  and  $\check{B}\check{V}_R^*\check{V}_R = \check{B}$ , we have

$$\check{Y} = \check{V}_L\check{Y}\check{V}_R$$

where  $\bar{Y} := \check{V}_L^*\check{Y}\check{V}_R^*$  is discrete system given in state-space as

$$\bar{Y} = \left( \begin{array}{c|c} e^{Ah} & \bar{B} \\ \hline \bar{C} & 0 \end{array} \right)$$

Here, we used

$$\begin{aligned} (\bar{C}^+)^*\check{C}^*\check{C} &= (\bar{C}^+)^*\bar{C}^*\bar{C} = (\bar{C}\bar{C}^+)^*\bar{C} = (\bar{C}\bar{C}^+)\bar{C} = \bar{C} \\ \check{B}\check{B}^*(\bar{B}^+)^* &= \bar{B}\bar{B}^*(\bar{B}^+)^* = \bar{B}(\bar{B}^+\bar{B})^* = \bar{B}(\bar{B}^+\bar{B}) = \bar{B}. \end{aligned}$$

which follows from the fact that  $\bar{C}\bar{C}^+$  and  $\bar{B}^+\bar{B}$  are orthogonal projections.

Since  $\check{V}_L^*\check{V}_L = \bar{C}\bar{C}^+$ ,  $\check{V}_R\check{V}_R^* = \bar{B}^+\bar{B}$  and  $\check{V}_L, \check{V}_R \in H^\infty$  (follows from lemmas 2.5.3 and 2.5.6 and stability of  $\check{V}_L, \check{V}_R$ ), we have

$$\check{Y} \in H^\infty \iff \bar{Y} \in H^\infty$$

□

*Proof of Lemma 5.3.25.* Using (5.15), we have

$$\check{G} = \check{D} + \check{Y}$$

where

$$\check{Y}(z)\check{u}(z) := C(\tau)e^{A\tau}(zI - \Upsilon e^{Ah})^{-1}\Upsilon \int_0^h e^{A(h-\sigma)}B(\sigma)\check{u}(\sigma) d\sigma.$$

Since  $e^{A(\tau-\sigma)}\mathbb{1}(\tau-\sigma) \in L^2[0, h] \times L^2[0, h]$ ,  $B \in L^2[0, h]$  and  $C \in L^2[0, h]$ , we have that  $\check{D} \in \mathbb{H}^2$  by Lemma 5.3.21. Since  $\check{D}$  is static, we have that  $\|\check{D}\|_{\mathbb{H}^2}^2 = \frac{1}{h}\|\check{D}\|_{HS}^2$ .

Since  $\check{G}(z)$  is causal,  $\check{Y}(z)$  is also causal. The formal series of causal  $\check{Y}(z)$  has the constant term

$$\lim_{z \rightarrow \infty} \check{Y}(z) = 0.$$

Therefore, if  $\check{G} \in \mathbb{H}^2$  then by orthogonality between  $\check{D}$  and  $\check{Y}$  in the space  $\mathbb{H}^2$ , we have

$$\|\check{G}\|_{\mathbb{H}^2}^2 = \frac{1}{h}\|\check{D}\|_{HS}^2 + \|\check{Y}\|_{\mathbb{H}^2}^2.$$

The integral form of  $\|\check{D}\|_{HS}^2$  can be obtained by using  $\|\check{D}\|_{HS}^2 = \text{tr } \check{D}\check{D}^*$ .

Define a static hold  $\check{V}_L := \check{C}\bar{C}^+$  and a static sampler  $\check{V}_R := \bar{B}^+\check{B}$  where  $+$  denotes the pseudo-inverse operation,  $\check{C}$  is defined in (5.83) and  $\check{B}$  is defined in (5.84). Now proceeding as in the proof of Lemma 5.3.17, we have

$$\|\check{Y}\|_{\mathbb{H}^2}^2 = \|\check{V}_L\check{Y}\check{V}_R\|_{\mathbb{H}^2}^2 = \|\bar{Y}\|_{\mathbb{H}^2}^2.$$

□

## 5.C Proofs of the results in Section 5.4

*Proof of Lemma 5.4.2.* Due to observability of  $(C_y, A)$ , there exists an  $L$  such that  $A + LC_y$  is Hurwitz [71, theorem 3.1]. Define  $L_1 := \begin{bmatrix} 0 & L \end{bmatrix}$ . Now using output

injection, the left coprime factorization of  $\check{G} := \begin{bmatrix} \check{G}_v \\ \check{G}_y \end{bmatrix}$  is  $\check{M}^{-1}\check{N}$  where

$$[\check{N} \quad \check{M}] = \left[ \begin{array}{c|cc} A + L_1C & B + L_1D & L_1 \\ \hline ZC & ZD & Z \end{array} \right]$$

where  $C$  and  $D$  are defined in (5.40) and  $Z$  is an invertible complex matrix. As a special case we choose  $Z = \text{diag}\{I, Z_y\}$  where  $Z_y$  is an invertible complex matrix. Partitioning the output according to  $\check{G}_v$  and  $\check{G}_y$  and substituting the value of  $L_1 = \begin{bmatrix} 0 & L \end{bmatrix}$ , we have

$$\left[ \begin{array}{c|cc} \check{N}_v & & \\ \check{N}_y & & \\ \hline & \check{M} & \end{array} \right] = \left[ \begin{array}{c|cc} A + LC_y & B + LD_y & \begin{matrix} 0 & L \\ I & 0 \\ 0 & Z_y \end{matrix} \\ \hline C_v & \begin{matrix} 0 & I \\ Z_y D_y & 0 \end{matrix} & \end{array} \right].$$

Clearly  $\check{M}$  has a form  $\begin{bmatrix} I & \check{M}_v \\ 0 & \check{M}_y \end{bmatrix}$  and

$$\check{M} = \begin{bmatrix} I & \check{M}_v \\ 0 & \check{M}_y \end{bmatrix} = \left[ \begin{array}{c|cc} A + LC_y & 0 & L \\ C_v & I & 0 \\ \hline Z_y C_y & 0 & Z_y \end{array} \right].$$

The controllability of the pair  $(A, B)$  helps in the construction of Bezout factors of  $\check{N}$  and  $\check{M}$ . The construction of the Bezout factors is quite standard (see [70, Chapter 5]).  $\check{M}_y, \check{M}_v, \check{N}_y$  and  $\check{N}_v$  belong to  $H^\infty$  because  $e^{(A+LC_y)h}$  is Schur (see Corollary 5.3.20).  $\square$

*Proof of Lemma 5.4.3.* Note that  $\mathcal{N}_y$  is an LCTI system with state-space

$$\mathcal{N}_y = \left[ \begin{array}{c|c} A + LC_y & B + LD_y \\ \hline Z_y C_y & Z_y D_y \end{array} \right].$$

where  $A + LC_y$  is Hurwitz. Now, if we have  $\check{N}_y(e^{j\theta})\check{N}_y^\sim(e^{j\theta}) > 0 \forall \theta \in [-\pi, \pi]$  then  $\mathcal{N}_y\mathcal{N}_y^\sim > 0$  and  $N_y(j\omega)N_y(j\omega)^\sim > 0$  for all  $\omega \in \mathbb{R}$  including  $\infty$ . Here,  $\mathcal{N}_y$  is the system  $\check{N}_y$  in time domain and  $N_y$  is the system  $\mathcal{N}_y$  in the (classic) frequency domain. As  $Z_y$  is invertible and  $A + LC_y$  is Hurwitz,  $N_y(j\omega)N_y(j\omega)^\sim > 0$  for all  $\omega \in \mathbb{R}$  including  $\infty$  iff  $D_y$  has full row rank and

$$\begin{bmatrix} A + LC_y - j\omega I & B + LD_y \\ Z_y C_y & Z_y D_y \end{bmatrix} = \begin{bmatrix} I & L \\ 0 & Z_y \end{bmatrix} \begin{bmatrix} A - j\omega I & B \\ C_y & D_y \end{bmatrix}$$

has full row rank for all  $\omega \in \mathbb{R}$ .  $\square$

*Proof of Lemma 5.4.5.* The proof is quite standard, for completeness it is given below. Let  $A_L := A + LC_y$  and  $B_L := B + LD_y$ .

$$\begin{aligned} \check{N}_v\check{N}_y^\sim &= \begin{bmatrix} A_L & B_L \\ C_v & 0 \end{bmatrix} \begin{bmatrix} -A_L^* & (Z_y C_y)^* \\ -B_L^* & (Z_y D_y)^* \end{bmatrix} \\ &= \begin{bmatrix} A_L & -B_L B_L^* & B_L (Z_y D_y)^* \\ 0 & -A_L^* & (Z_y C_y)^* \\ C_v & 0 & 0 \end{bmatrix} \end{aligned}$$

Applying a state-transformation  $\begin{bmatrix} I & X \\ 0 & I \end{bmatrix}$  and using  $A_L X + X A_L^* + B_L B_L^* = 0$  and  $B_L (Z_y D_y)^* + X (Z_y C_y)^* = 0$  (follows from Lemma 5.4.4), we have

$$\begin{aligned} \check{N}_v\check{N}_y^\sim &= \left[ \begin{array}{cc|c} A_L & A_L X + X A_L^* + B_L B_L^* & B_L (Z_y D_y)^* + X (Z_y C_y)^* \\ 0 & -A_L^* & (Z_y C_y)^* \\ \hline C_v & -C_v X & 0 \end{array} \right] \\ &= \left[ \begin{array}{cc|c} A_L & 0 & 0 \\ 0 & -A_L^* & (Z_y C_y)^* \\ \hline C_v & -C_v X & 0 \end{array} \right] \end{aligned}$$

$$= \left[ \begin{array}{c|c} -A_L^* & (Z_y C_y)^* \\ \hline -C v X & 0 \end{array} \right]$$

□

*Proof of Proposition 5.4.6.* Define  $\bar{H}_{cs}(z) := \bar{C}_H \bar{H}_s(z)$ . Note that

$$\hat{H}^\sim(z) \hat{H}(z) = \bar{H}_s(z) \bar{C}_H^* \bar{C}_H \bar{H}_s(z) = \bar{H}_{cs}^\sim(z) \bar{H}_{cs}(z).$$

As  $E e^{A_H h}$  is Schur (Assumption  $\mathcal{A}_6$ ),  $\bar{H}_{cs}^\sim(e^{j\theta}) \bar{H}_{cs}(e^{j\theta}) > 0 \forall \theta \in [-\pi, \pi]$  is equivalent to say that the matrix

$$R(e^{j\theta}) := \begin{bmatrix} E e^{A_H h} - e^{j\theta} I & E e^{A_H h} B_H \\ \bar{C}_H & \bar{C}_H B_H \end{bmatrix}$$

has full column rank for every  $\theta \in [-\pi, \pi]$ . Now,  $R(e^{j\theta})$  can be written as

$$R(e^{j\theta}) = \begin{bmatrix} E e^{A_H h} - e^{j\theta} I & B_H \\ \bar{C}_H & 0 \end{bmatrix} \begin{bmatrix} I & B_H \\ 0 & e^{j\theta} I \end{bmatrix}.$$

Since  $\begin{bmatrix} I & B_H \\ 0 & e^{j\theta} I \end{bmatrix}$  is invertible for all  $\theta \in [-\pi, \pi]$ , we have that

$$\text{rank } R(e^{j\theta}) = \text{rank} \begin{bmatrix} E e^{A_H h} - e^{j\theta} I & B_H \\ \bar{C}_H & 0 \end{bmatrix} \quad (5.85)$$

for all  $\theta \in [-\pi, \pi]$ . Now, the proof of equivalence of condition 1 and 2 is well known and the proof of equivalence of condition 2 and 3 is essentially given in [32, Theorem 4.1]. Note that if  $\hat{H} = \hat{H}_1 \bar{H}_0$ , then a spectral factor  $W(z)$  of  $\hat{H}^\sim(z) \hat{H}(z)$  is  $\bar{H}_0$  and if  $W(z)$  is a spectral factor of  $\hat{H}^\sim(z) \hat{H}(z)$  then  $\bar{H}_0(z) = W(z)$  and  $\hat{H}_1(z) = \hat{H}(z) W^{-1}(z)$ .

Now, we prove the equivalence of condition 2 and 4. Note that

$$\begin{aligned} \bar{H}_{cs}(z) &= \bar{C}_H B_H + \bar{C}_H (zI - E e^{A_H h})^{-1} E e^{A_H h} B_H \\ &= \bar{C}_H (zI - E e^{A_H h})^{-1} (zI - E e^{A_H h} + E e^{A_H h}) B_H \\ &= z \bar{C}_H (zI - E e^{A_H h})^{-1} B_H. \end{aligned}$$

This implies,

$$\bar{H}_{cs}^\sim(z) \bar{H}_{cs}(z) = B_H^* \left( \frac{1}{z} I - (E e^{A_H h})^* \right)^{-1} \bar{C}_H^* \bar{C}_H (zI - E e^{A_H h})^{-1} B_H.$$

Using the above equation and [32, Theorem 4.1], the existence of the spectral factorization of  $\hat{H}^\sim(z) \hat{H}(z) = \bar{H}_{cs}^\sim(z) \bar{H}_{cs}(z)$  is equivalent to 4.

Now, we prove the equivalence of condition 5 with the rest. Note that  $\hat{H} \in H^\infty$  is the same as  $\mathcal{H} \in H^\infty$ . Condition 1 says that  $\hat{H}$  is left invertible in  $L^\infty$ . Therefore by Parseval,  $\mathcal{H} : \ell^2 \rightarrow L^2$  is left invertible. Now,

$$\|\bar{u}\|_2 = \|\mathcal{H}^{-L} \mathcal{H} \bar{u}\| \leq \|\mathcal{H}^{-L}\|_{L^\infty} \|\mathcal{H} \bar{u}\|_2$$

where  $\mathcal{H}^{-L}$  is a left inverse of  $\mathcal{H}$ . In this case, condition 5 holds for  $\epsilon = \|\mathcal{H}^{-L}\|_{L^\infty}$ . On the other hand, if there exists an  $\epsilon$  such that

$$\|\mathcal{H}\bar{u}\|_2 \geq \epsilon\|\bar{u}\|_2 \quad \forall \bar{u} \in \ell^2,$$

then by taking lifted  $z$ -transform we have

$$\|\mathring{H}\mathring{\mathfrak{Z}}(\bar{u})\|_2 \geq \epsilon\|\mathring{\mathfrak{Z}}(\bar{u})\|_2.$$

This implies that  $\mathring{H}$  is left invertible in  $L^\infty$  (see [47, lemma 4.47, 4.48(b)]). Hence,  $\bar{H}_{cs}^\sim(e^{j\theta})\bar{H}_{cs}(e^{j\theta}) = \mathring{H}^\sim(e^{j\theta})\mathring{H}(e^{j\theta}) > 0$  (condition 3). This again is equivalent to condition 1 as proved earlier.  $\square$

*Proof of Lemma 5.4.8.* If  $\mathcal{A}_6$  is satisfied, then  $F = 0$  makes  $(E + B_H F)e^{A_H h}$  Schur. However, there may be many such  $F$  other than zero. The state equation of the Hold  $\mathring{H}$  can be written as

$$\begin{aligned} \dot{\check{x}}(\tau) &= A_H \check{x}(\tau) + B_H \mathcal{J}_{0^+} \bar{u}, & z\check{x}(0) &= E\check{x}(h^-), \quad \tau \in [0, h) \\ \check{y}(\tau) &= C_H \check{x}(\tau) \end{aligned}$$

Using the standard trick of state feedback for constructing coprime factors, we define

$$\begin{aligned} \bar{v} &:= \bar{u} - F\check{x}(0^-) \\ &= \bar{u} - \frac{1}{z} F \mathcal{J}_h \check{x}(\tau). \end{aligned}$$

Therefore,

$$\begin{aligned} \dot{\check{x}}(\tau) &= A_H \check{x}(\tau) + B_H \mathcal{J}_{0^+} (\bar{v} + F\check{x}(0^-)), & z\check{x}(0) &= E\check{x}(h^-), \quad \tau \in [0, h) \\ \check{y}(\tau) &= C_H \check{x}(\tau) \end{aligned}$$

Now using Lemma 5.3.31, we have

$$\begin{aligned} \dot{\check{x}}_1(\tau) &= A_H \check{x}_1(\tau) + B_H \mathcal{J}_{0^+} \bar{v}, & z(\check{x}_1(0) - B_H F \check{x}_1(0^-)) &= E\check{x}(h^-) \\ \check{y}(\tau) &= C_H \check{x}_1(\tau) \end{aligned}$$

Now, consider the boundary condition

$$\begin{aligned} z(\check{x}_1(0) - B_H F \check{x}_1(0^-)) &= E\check{x}_1(h^-) \\ \iff z(\check{x}_1(0) - \frac{1}{z} B_H F \check{x}_1(h^-)) &= E\check{x}_1(h^-) \\ \iff z\check{x}_1(0) &= E\check{x}_1(h^-) + B_H F \check{x}_1(h^-) = (E + B_H F)\check{x}_1(h^-) \end{aligned}$$

Now,  $\bar{u} = \bar{M}_H \bar{v}$ ,  $\check{y} = \hat{N}_H \bar{v}$ . Since  $F$  is such that  $(E + B_H F)e^{A_H h}$  is Schur, therefore  $\bar{M}_H, \hat{N}_H \in H^\infty$  (see Corollary 5.3.20).

The mapping  $\bar{M}_H^{-1}$  from  $\bar{u}$  to  $\bar{v}$  is given by

$$\bar{M}_H^{-1}(z) = \left[ \begin{array}{c|c} A_H & B_H \mathcal{J}_{0^+} \\ \hline -\frac{1}{z} \mathcal{J}_h^* F & I \end{array} \right]_{[zI \quad -E]}$$

Since  $Ee^{A_H h}$  is Schur (see Assumption  $\mathcal{A}_6$ ), discrete system  $\bar{M}_H^{-1}$  is in  $H^\infty$  (as  $\bar{M}_H^{-1}(z)$  is analytic and bounded in  $\mathbb{C} \setminus \mathbb{D}$ ). Now, we have

$$\bar{M}_H^{-1} \bar{M}_H + 0 \dot{N}_H = I$$

Therefore,  $\bar{M}_H$  and  $\dot{N}_H$  are right coprime.  $\square$

*Proof of Lemma 5.4.10.* For some invertible complex matrix  $Z$ , our aim here is to make  $\dot{N}_H Z$  an inner matrix i.e.  $(\dot{N}_H Z) \sim \dot{N}_H Z = I$ . By Lemma 5.3.6, the conjugate of the  $\dot{N}_H$  is the sampler  $\dot{N}_H^\sim$  with

$$\dot{N}_H^\sim(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} -A_H^* & C_H^* \\ \hline -B_H^* & 0 \end{array} \right]_{[z(E + B_H F)^* \quad -I]}.$$

To find the  $Z$  such that  $(\dot{N}_H^\sim Z) \sim \dot{N}_H Z = I$ , we first consider

$$\dot{N}_H^\sim(z) \dot{N}_H(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & C_H^* C_H & 0 \\ 0 & A_H & B_H \\ \hline -B_H^* & 0 & 0 \end{array} \right]_{[z\Omega_0 \quad \Upsilon_0]} \mathcal{J}_{0^+}$$

where  $\Omega_0 := \begin{bmatrix} (E + B_H F)^* & 0 \\ 0 & I \end{bmatrix}$  and  $\Upsilon_0 := -\begin{bmatrix} I & 0 \\ 0 & E + B_H F \end{bmatrix}$ .

Applying a time varying state transform  $T(t) = \begin{bmatrix} I & Q(t) \\ 0 & I \end{bmatrix}$  where  $Q(t)$  satisfy the differential Lyapunov equation

$$\dot{Q}(t) = -A_H^* Q(t) - Q(t) A_H - C_H^* C_H, \quad (5.86)$$

we have

$$\dot{N}_H^\sim(z) \dot{N}_H(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & 0 & Q(t) B_H \\ 0 & A_H & B_H \\ \hline -B_H^* & B_H^* Q(t) & 0 \end{array} \right]_{[z\Omega_0 T^{-1}(0) \quad \Upsilon_0 T^{-1}(h)]} \mathcal{J}_{0^+}$$

Now, the boundary condition of  $\dot{N}_H^\sim(z) \dot{N}_H(z)$  is given by

$$\begin{bmatrix} z \\ z \end{bmatrix} \begin{bmatrix} (E + B_H F)^* & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & -Q_0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} I & 0 \\ 0 & E + B_H F \end{bmatrix} \begin{bmatrix} I & -Q_h \\ 0 & I \end{bmatrix}$$

which can be rearranged as

$$\begin{bmatrix} z \left[ \begin{array}{cc} (E + B_H F)^* & -(E + B_H F)^* Q_0 \\ 0 & I \end{array} \right] & - \begin{bmatrix} I & -Q_h \\ 0 & E + B_H F \end{bmatrix} \end{bmatrix}.$$

Here  $Q_0 := Q(0)$  and  $Q_h := Q(h)$ . To decouple the boundary condition (meaning block diagonal here), pre-multiply both side the above by

$$S := \begin{bmatrix} I & (E + B_H F)^* Q_0 \\ 0 & I \end{bmatrix}$$

Pre-multiplication with  $S$  does not change the system (see Corollary 5.3.4), therefore we have the boundary condition

$$\begin{bmatrix} z \left[ \begin{array}{cc} (E + B_H F)^* & 0 \\ 0 & I \end{array} \right] & - \begin{bmatrix} I & -Q_h + (E + B_H F)^* Q_0 (E + B_H F) \\ 0 & E + B_H F \end{bmatrix} \end{bmatrix}$$

The condition

$$(E + B_H F)^* Q_0 (E + B_H F) - Q_h = 0 \quad (5.87)$$

will lead to decoupled STPBC.

From Lemma 5.4.9, the solution of (5.86) with initial condition  $Q(0) = Q_0$  is given by

$$Q(t) = R_1(t) Q_0 R_1^*(t) - R_3(t)$$

where  $R_1(t) := e^{-A_H^* t}$ , and  $R_3(t) := \int_0^t e^{-A_H^* s} C_H^* C_H e^{-A_H s} ds$ . Now using decoupling condition (5.87), we have

$$Q(h) = R_1(h) Q_0 R_1^*(h) - R_3(h) = (E + B_H F)^* Q_0 (E + B_H F).$$

The above equation can be written in  $Q_0$  alone as

$$\begin{aligned} Q_0 &= R_1(h)^{-1} \left( (E + B_H F)^* Q_0 (E + B_H F) + R_3(h) \right) R_1(h)^{-*} \\ &= R_1(h)^{-1} (E + B_H F)^* Q_0 (E + B_H F) R_1(h)^{-*} + \bar{C}_H^* \bar{C}_H \end{aligned} \quad (5.88)$$

As  $(A_H, C_H)$  is assumed observable (Assumption  $\mathcal{A}_8$ ),  $\bar{C}_H^* \bar{C}_H > 0$  [71, theorem 3.3]. Also,  $(E + B_H F) R_1(h)^{-*}$  is Schur because  $F$  is assumed to be chosen that way. Therefore, there exists a unique solution  $Q_0 > 0$  of discrete Lyapunov equation (5.88).

This value of  $Q_0$  renders

$$\dot{\tilde{N}}_H(z) \dot{\tilde{N}}_H(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & 0 & Q(t) B_H \\ 0 & A_H & B_H \\ \hline -B_H^* & B_H^* Q(t) & 0 \end{array} \right] \begin{bmatrix} z \Omega_0 & \Upsilon_0 \end{bmatrix} \mathcal{J}_{0^+}$$



with decoupled boundary condition. Next, we look for a matrix  $F$  such that  $\tilde{N}_H(z)\hat{N}_H(z)$  is a static (and invertible also) discrete system. Using Lemma 5.3.30

$$\begin{aligned}\tilde{N}_H(z)\hat{N}_H(z) &= -B_H^*(z(E + B_H F)^* - e^{-A_H^* h})^{-1}z(E + B_H F)^* Q_0 B_H \\ &\quad + B_H^* Q_0(zI - (E + B_H F)e^{A_H h})^{-1}z B_H\end{aligned}$$

Since,

$$\begin{aligned}B_H^* Q_0(zI - (E + B_H F)e^{A_H h})^{-1}z B_H \\ &= B_H^* Q_0(I + z^{-1}(E + B_H F)e^{A_H h}(I - z^{-1}(E + B_H F)e^{A_H h})^{-1})B_H \\ &= B_H^* Q_0 B_H + B_H^* Q_0(E + B_H F)e^{A_H h}(zI - (E + B_H F)e^{A_H h})^{-1} B_H\end{aligned}$$

This implies that if we choose  $F$  such that  $B_H^* Q_0(E + B_H F) = 0$ , then

$$\tilde{N}_H(z)\hat{N}_H(z) = B_H^* Q_0 B_H$$

which is a static discrete system.

Since  $Q_0 > 0$  and  $B_H$  has full column-rank, the matrix  $B_H^* Q_0 B_H$  is invertible, hence

$$F = -(B_H^* Q_0 B_H)^{-1} B_H^* Q_0 E$$

and  $\hat{N}_H(z)(B_H^* Q_0 B_H)^{-\frac{1}{2}}$  is the desired inner factor.

To prove that  $F = -(B_H^* Q_0 B_H)^{-1} B_H^* Q_0 E$  makes  $(E + B_H F)e^{A_H h}$  Schur, we substitute this value of  $F$  in (5.88). Hence,

$$\begin{aligned}(E + B_H F)^* Q_0 (E + B_H F) &= E^* Q_0 E - E^* Q_0 B_x Q_0 E \\ &\quad - E^* Q_0 B_x Q_0 E + E^* Q_0 B_x Q_0 B_x Q_0 E \\ &= E^* Q_0 E - E^* Q_0 B_x Q_0 E\end{aligned}$$

where we used  $B_x := B_H(B_H^* Q_0 B_H)^{-1} B_H^*$  and that  $B_x Q_0 B_x = B_x$ . Therefore,

$$Q_0 = R_1(h)^{-1} E^* (Q_0 - Q_0 B_H (B_H^* Q_0 B_H)^{-1} B_H^* Q_0) E R_1(h)^{-*} + \bar{C}_H^* \bar{C}_H$$

By Lemma 5.4.7, Assumption  $\mathcal{A}_3$  is satisfied as assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied. Now, if Assumption  $\mathcal{A}_3$  is satisfied then there exists a unique solution  $Q_0$  such that  $(E + B_H F)e^{A_H h}$  is Schur matrix (see Proposition 5.4.6).  $\square$

*Proof of Lemma 5.4.11.* Note that  $\hat{C}_H$  is in  $H^\infty$  as it is causal and its hold function is in  $L^2$  (see lemmas 2.4.7 and 2.5.6). Also,  $\tilde{C}_H \hat{C}_H$  is a discrete system given by

$$\tilde{C}_H \hat{C}_H = \bar{C}_H^* \bar{C}_H = \int_0^h e^{A_H^* \tau} C_H^* C_H e^{A_H \tau} d\tau. \quad (5.89)$$

Hence,  $\hat{V}_H = \hat{C}_H \bar{C}_H^+$ . Note that matrix  $\bar{C}_H$  is a finite domain operator (hence closed), therefore  $(\bar{C}_H^+)^* = (\bar{C}_H^*)^+$  and  $\bar{C}_H = (\bar{C}_H^*)^+(\bar{C}_H^* \bar{C}_H)$  [5, lemma 4.1]. Now  $\hat{V}_H \hat{V}_H$  is a discrete system given by

$$\hat{V}_H \hat{V}_H = (\bar{C}_H^+)^* \hat{C}_H \bar{C}_H \bar{C}_H^+ = (\bar{C}_H^+)^* \bar{C}_H^* \bar{C}_H \bar{C}_H^+ = \bar{C}_H \bar{C}_H^+$$

Therefore,  $\hat{V}_H \hat{V}_H$  is an orthogonal projection onto the space  $(\text{Ker } \hat{V}_H \hat{V}_H)^\perp = \text{Im } \bar{O}_{\bar{C}_H}$ . Since  $\bar{O}_{\bar{C}_H}$  is a static discrete system, it is in  $H^\infty$ . Similarly, we have

$$\begin{aligned} \hat{V}_H \hat{V}_H \hat{C}_H &= \hat{C}_H \bar{C}_H^+ (\bar{C}_H^+)^* \hat{C}_H \bar{C}_H = \hat{C}_H \bar{C}_H^+ (\bar{C}_H^+)^* \bar{C}_H^* \bar{C}_H \\ &= \hat{C}_H \bar{C}_H^+ \bar{C}_H. \end{aligned}$$

It follows from (5.89) that  $(\text{Ker } \bar{O}_{\bar{C}_H})^\perp = (\text{Ker } \hat{C}_H)^\perp$ . Since the static discrete system  $\bar{C}_H^+ \bar{C}_H$  is an orthogonal projection onto  $(\text{Ker } \bar{O}_{\bar{C}_H})^\perp = (\text{Ker } \hat{C}_H)^\perp$ , we have

$$\hat{V}_H \hat{V}_H \hat{C}_H = \hat{C}_H \bar{C}_H^+ \bar{C}_H = \hat{C}_H. \quad (5.90)$$

Therefore,  $\hat{V}_H \hat{V}_H$  maps the  $\text{Im } \hat{C}_H$  to itself. Now, from (5.46) and (5.90) we have

$$\hat{V}_H \hat{V}_H \hat{H} = \hat{V}_H \hat{V}_H \hat{C}_H \bar{H}_s = \hat{C}_H \bar{H}_s = \hat{H}.$$

Now,

$$\hat{V}_H \hat{H} = (\bar{C}_H^+)^* \hat{C}_H \hat{H} = (\bar{C}_H^+)^* \bar{C}_H^* \bar{C}_H \bar{H}_s = \bar{C}_H \bar{H}_s = \bar{H}_{cs}.$$

□

*Proof of Lemma 5.4.14.* If assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$  are satisfied then  $\hat{H}_i$  and  $\bar{H}_o$  in  $H^\infty$  exist (see Lemma 5.4.7). Now, using Lemma 5.4.11, we have

$$\hat{H} = \hat{V}_H \bar{H}_{cs}.$$

Similarly, using lemma 5.4.10 and 5.4.13, we have

$$\hat{H} = \hat{H}_i \bar{H}_o = \hat{V}_H \bar{H}_{ics} \bar{H}_o.$$

Therefore, we have

$$\hat{V}_H \bar{H}_{cs} = \hat{V}_H \bar{H}_{ics} \bar{H}_o.$$

Since  $\hat{V}_H \hat{V}_H$  is an orthogonal projection onto the space  $(\text{Ker } \hat{V}_H \hat{V}_H)^\perp = \text{Im } \bar{C}_H$  (see Lemma 5.4.11), we have

$$\bar{H}_{cs} = \bar{H}_{ics} \bar{H}_o.$$

□

*Proof of Lemma 5.4.16.* Using (5.55), we have

$$(I - \dot{H}_1 \dot{H}_1^\sim) \check{G}_v \in L^\infty \iff (I - \dot{V}_H \bar{H}_{cs} \bar{H}_{cs}^+ \dot{V}_H^\sim) \check{G}_v \in L^\infty$$

Since  $\dot{V}_H \dot{V}_H^\sim$  is an orthogonal projection onto the space  $(\text{Ker } \dot{V}_H \dot{V}_H^\sim)^\perp = \text{Im } \bar{C}_H$  and  $\begin{bmatrix} \dot{V}_H^\sim \\ I - \dot{V}_H \dot{V}_H^\sim \end{bmatrix}$  is inner, we have

$$\begin{aligned} (I - \dot{H}_1 \dot{H}_1^\sim) \check{G}_v \in L^\infty &\iff \begin{bmatrix} \dot{V}_H^\sim \\ I - \dot{V}_H \dot{V}_H^\sim \end{bmatrix} (I - \dot{V}_H \bar{H}_{cs} \bar{H}_{cs}^+ \dot{V}_H^\sim) \check{G}_v \in L^\infty \\ &\iff \begin{bmatrix} \dot{V}_H^\sim - \bar{H}_{cs} \bar{H}_{cs}^+ \dot{V}_H^\sim \\ I - \dot{V}_H \dot{V}_H^\sim \end{bmatrix} \check{G}_v \in L^\infty \end{aligned}$$

Since  $\dot{V}_H$  is a zero order hold,  $\dot{V}_H$  and  $\dot{V}_H^\sim$  are causal (see Example 2.5.5). Therefore, if  $\check{G}_v$  is causal then  $(I - \dot{V}_H \dot{V}_H^\sim) \check{G}_v \in L^\infty$  implies  $(I - \dot{V}_H \dot{V}_H^\sim) \check{G}_v \in H^\infty$ .  $\square$

*Proof of Lemma 5.4.21.* Since  $\bar{D}^\S$ , defined in Lemma 5.4.19, is invertible and satisfies  $\bar{D}^\S \bar{C}_H B_H = \begin{bmatrix} I \\ 0 \end{bmatrix}$ , there exists a matrix  $X$  such that

$$(\bar{D}^\S)^{-1} = \begin{bmatrix} \bar{D}^+ \\ \bar{D}^\perp \end{bmatrix}^{-1} = [\bar{C}_H B_H \quad X].$$

Similarly, as  $\bar{D}^\S$  is invertible, the system  $\begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix}$  has an inverse

$$\begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix}^{-1} = \left( \begin{array}{c|cc} Ee^{A_H h} & B_H & L_{cs} \\ \hline \bar{C}_H & \bar{C}_H B_H & X \end{array} \right) = [\bar{H}_{cs} \quad \bar{P}]$$

where  $\bar{P} := \left( \begin{array}{c|c} Ee^{A_H h} & L_{cs} \\ \hline \bar{C}_H & X \end{array} \right)$ . The systems  $\begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix}^{-1}$  and  $\bar{P}$  are in  $H^\infty$  because  $Ee^{A_H h}$  is Schur (Assumption  $\mathcal{A}_6$ ). Multiplying both sides of the above by  $I - \bar{H}_{cs} \bar{H}_{cs}^+$ , we have

$$(I - \bar{H}_{cs} \bar{H}_{cs}^+) \begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix}^{-1} = (I - \bar{H}_{cs} \bar{H}_{cs}^+) [\bar{H}_{cs} \quad \bar{P}]$$

Note that existence of  $\bar{H}_{cs}^+$  in  $L^\infty$  is guaranteed by assumptions  $\mathcal{A}_6$ – $\mathcal{A}_8$ . As  $(I - \bar{H}_{cs} \bar{H}_{cs}^+) \bar{H}_{cs} = 0$ , the above implies

$$(I - \bar{H}_{cs} \bar{H}_{cs}^+) \begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix}^{-1} = [0 \quad \bar{W}_{cs}]$$

where  $\bar{W}_{cs} := (I - \bar{H}_{cs} \bar{H}_{cs}^+) \bar{P}$ . Therefore,

$$(I - \bar{H}_{cs} \bar{H}_{cs}^+) = \begin{bmatrix} 0 & \bar{W}_{cs} \end{bmatrix} \begin{bmatrix} \bar{H}_{cs}^L \\ \bar{H}_{cs}^\perp \end{bmatrix} = \bar{W}_{cs} \bar{H}_{cs}^\perp$$

Since  $\bar{P}$  and  $(I - \bar{H}_{cs} \bar{H}_{cs}^+)$  are in  $L^\infty$ , we have that  $\bar{W}_{cs}$  is in  $L^\infty$ .

To prove that  $\bar{W}_{cs}$  is left invertible consider

$$\begin{aligned} \bar{H}_{cs}^\perp \bar{W}_{cs} &= \bar{H}_{cs}^\perp (I - \bar{H}_{cs} \bar{H}_{cs}^+) \bar{P} \\ &= \bar{H}_{cs}^\perp \bar{P} = I \end{aligned}$$

This implies  $\bar{H}_{cs}^\perp \in L^\infty$  is a left inverse of  $\bar{W}_{cs}$ . □

*Proof of Theorem 5.4.23.* From Lemma 5.4.16, we have

$$(I - \check{H}_i \check{H}_i^\sim) \check{G}_v \in L^\infty \iff \begin{bmatrix} (I - \bar{H}_{cs} \bar{H}_{cs}^+) \check{V}_H^\sim \\ I - \check{V}_H \check{V}_H^\sim \end{bmatrix} \check{G}_v \in L^\infty$$

Using Lemma 5.4.21, we have

$$\begin{aligned} (I - \bar{H}_{cs} \bar{H}_{cs}^+) \check{V}_H^\sim \check{G}_v \in L^\infty &\iff \bar{W}_{cs} \bar{H}_{cs}^\perp \check{V}_H^\sim \check{G}_v \in L^\infty \\ &\iff \bar{H}_{cs}^\perp \check{V}_H^\sim \check{G}_v \in L^\infty \end{aligned}$$

where we used the invertibility of the system  $\bar{W}_{cs}$  in  $L^\infty$  in the last step.

Note that  $\check{V}_H^\sim$  and  $\check{V}_H$  are causal as  $\check{V}_H$  is a zero order hold (see Example 2.5.5). Since  $\check{G}_v$ ,  $\check{V}_H$ ,  $\check{V}_H^\sim$  and  $\bar{H}_{cs}^\perp$  are causal, the systems  $\bar{H}_{cs}^\perp \check{V}_H^\sim \check{G}_v$  and  $(I - \check{V}_H \check{V}_H^\sim) \check{G}_v$  are causal as well. Now the results follows from the fact that a causal and stable system belong to  $H^\infty$ . □

*Proof of Lemma 5.4.26.* Using (5.15), we have

$$\check{G}_u(z) = \check{T}_a(z) + \check{T}_b(z)$$

for the mapping  $\check{T}_a$  and  $\check{T}_b$  defined as

$$\begin{aligned} \check{T}_a(z) \check{u}(z) &:= \int_0^\tau C_u e^{A_u(\tau-\sigma)} B \check{u}(z; \sigma) d\sigma \\ \check{T}_b(z) \check{u}(z) &:= C_{vu} e^{A_u \tau} (zI - e^{A_u h})^{-1} \int_0^h e^{A_u(h-\sigma)} B_u \check{u}(z; \sigma) d\sigma \end{aligned}$$

Therefore,

$$\begin{aligned} \check{V}_H^\sim(z) \check{T}_b(z) \check{u}(z) &= (\bar{C}_H^*)^+ \int_0^h e^{A_H^* \tau} C_H^* C_{vu} e^{A_u \tau} d\tau \\ &\quad \times (zI - e^{A_u h})^{-1} \int_0^h e^{A_u(h-\sigma)} B_u \check{u}(z; \sigma) d\sigma \end{aligned}$$

$$\begin{aligned}
&= (\bar{C}_H^*)^+ P_u (zI - e^{A_u h})^{-1} \int_0^h e^{A_u(h-\sigma)} B_u \check{u}(z; \sigma) d\sigma \\
&= \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_u & B_u \\ \hline -(\bar{C}_H^*)^+ P_u & 0 \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix} \check{u}(z)
\end{aligned}$$

Clearly  $e^{A_u(\tau-\sigma)} \mathbb{1}(\tau - \sigma)$  is bounded on  $\tau \in [0, h)$  and  $\sigma \in [0, h)$ . By Lemma 5.3.16 this implies,  $\check{T}_a \in H^\infty$ . Since  $\check{T}_a \in H^\infty$  and  $\check{V}_H \in H^\infty$  (as  $\check{V}_H$  is a zero order hold), we have that  $\check{V}_H \check{T}_a$  and  $(I - \check{V}_H \check{V}_H) \check{T}_a$  belongs to  $H^\infty$ . Therefore,

$$(I - \check{V}_H \check{V}_H) \check{G}_u \in H^\infty \iff \check{T}_b - \check{V}_H \check{V}_H \check{T}_b \in H^\infty$$

Now using (5.15) again, we have

$$(\check{T}_b(z) - \check{V}_H(z) \check{V}_H(z) \check{T}_b(z)) \check{u}(z) = \check{T}_2(z) \check{u}(z)$$

where

$$\begin{aligned}
\check{T}_2(z) \check{u}(z) &:= \left( C_{vu} e^{A_u \tau} - C_H e^{A_H \tau} (\bar{C}_H)^+ (\bar{C}_H^*)^+ P_u \right) \\
&\quad (zI - e^{A_u h})^{-1} \int_0^h e^{A_u(h-\sigma)} B_u \check{u}(z; \sigma) d\sigma
\end{aligned}$$

To check whether  $\check{T}_2$  belongs to  $H^\infty$  or not, we aim here to find a discrete system which is in  $H^\infty$  iff  $\check{T}_2$  is in  $H^\infty$ . To this end, define a zero order hold  $\check{C}_L$  as

$$\check{y}(z) = \check{C}_L(z) \check{x}(z) : \quad \check{y}(z; \tau) = \left( C_{vu} e^{A_u \tau} - C_H e^{A_H \tau} (\bar{C}_H)^+ (\bar{C}_H^*)^+ P_u \right) \check{x}(z)$$

Clearly,  $\check{C}_L^* \check{C}_L$  is a static discrete system given by

$$\begin{aligned}
\check{C}_L^* \check{C}_L &= \bar{C}_{vu}^* \bar{C}_{vu} - P_u^* \bar{C}_H^+ (\bar{C}_H^*)^+ P_u - P_u^* \bar{C}_H^+ (\bar{C}_H^*)^+ P_u \\
&\quad + P_u^* \bar{C}_H^+ (\bar{C}_H^*)^+ (\bar{C}_H^* \bar{C}_H) \bar{C}_H^+ (\bar{C}_H^*)^+ P_u
\end{aligned}$$

Since  $\bar{C}_H^+ \bar{C}_H$  is an orthogonal projection onto the space  $(\text{Ker } \bar{C}_H)^\perp$  and  $(\bar{C}_H^*)^+ = (\bar{C}_H^+)^*$ , we have that

$$\bar{C}_H^+ \bar{C}_H (\bar{C}_H^+ (\bar{C}_H^*)^+) = \bar{C}_H^+ (\bar{C}_H^*)^+ \text{ and } ((\bar{C}_H^*)^+ \bar{C}_H^*) (\bar{C}_H \bar{C}_H^+) = \bar{C}_H \bar{C}_H^+.$$

Therefore,

$$\begin{aligned}
\bar{C}_H^+ (((\bar{C}_H^*)^+ \bar{C}_H^*) (\bar{C}_H \bar{C}_H^+)) (\bar{C}_H^*)^+ &= \bar{C}_H^+ \bar{C}_H \bar{C}_H^+ (\bar{C}_H^*)^+ \\
&= (\bar{C}_H^+ \bar{C}_H) \bar{C}_H^+ (\bar{C}_H^*)^+ \\
&= \bar{C}_H^+ (\bar{C}_H^*)^+
\end{aligned}$$

Hence,

$$\check{C}_L^* \check{C}_L = \bar{C}_L^* \bar{C}_L$$

where  $\bar{C}_L$  is a matrix which satisfies

$$\bar{C}_L^* \bar{C}_L = \bar{C}_{vu}^* \bar{C}_{vu} - P_u^* \bar{C}_H^+ (\bar{C}_H^*)^+ P_u.$$

Define a zero order hold  $\hat{V}_L := \hat{C}_L \bar{C}_L^+$  where  $\bar{C}_L^+$  is the pseudo-inverse of the matrix  $\bar{C}_L$ . Clearly  $\hat{V}_L \hat{V}_L^\sim = \bar{C}_L \bar{C}_L^+$  i.e it is an orthogonal projection onto  $\text{Im } \bar{C}_L$ . Also, we have  $\hat{V}_L \hat{V}_L^\sim \hat{C}_L = \hat{C}_L$ .

Define a sampler  $\hat{B}_u$  as

$$\bar{y}(z) = \hat{B}_u(z) \check{x}(z) : \quad \bar{y}(z) = \int_0^h e^{A_u(h-\sigma)} B_u \check{x}(z; \sigma) d\sigma$$

Also, define a sampler  $\hat{V}_R := \bar{B}_u^+ \hat{B}_u$  where  $\bar{B}_u^+$  is the pseudo-inverse of the matrix  $\bar{B}_u$ . Clearly,  $\hat{V}_R \hat{V}_R^\sim = \bar{B}_u^+ \bar{B}_u$  i.e. it is an orthogonal projection onto  $(\text{Ker } \bar{B}_u)^\perp$ . Also, we have  $\hat{B}_u \hat{V}_R^\sim \hat{V}_R = \hat{B}_u$ .

Therefore, using  $\hat{V}_L \hat{V}_L^\sim \hat{C}_L = \hat{C}_L$  and  $\hat{B}_u \hat{V}_R^\sim \hat{V}_R = \hat{B}_u$ , we have

$$\check{T}_2 = \hat{V}_L \bar{T}_2 \hat{V}_R$$

where  $\bar{T}_2 := \hat{V}_L^\sim \check{T}_2 \hat{V}_R^\sim$  is discrete system given in state-space as

$$\bar{T}_2 = \left( \begin{array}{c|c} e^{A_u h} & \bar{B}_u \\ \hline \bar{C}_L & 0 \end{array} \right)$$

Here, we used

$$\begin{aligned} (\bar{C}_L^+)^* \bar{C}_L \hat{C}_L &= (\bar{C}_L^+)^* \bar{C}_L^* \bar{C}_L = (\bar{C}_L \bar{C}_L^+)^* \bar{C}_L = (\bar{C}_L \bar{C}_L^+) \bar{C}_L = \bar{C}_L \\ \hat{B}_u \hat{B}_u^\sim (\bar{B}_u^+)^* &= \bar{B}_u \bar{B}_u^* (\bar{B}_u^+)^* = \bar{B}_u (\bar{B}_u^+ \bar{B}_u)^* = \bar{B}_u (\bar{B}_u^+ \bar{B}_u) = \bar{B}_u. \end{aligned}$$

which follows from the fact that  $\bar{C}_L \bar{C}_L^+$  and  $\bar{B}_u^+ \bar{B}_u$  are orthogonal projections.

Clearly,  $\hat{V}_L^\sim, \hat{V}_R^\sim \in L^\infty$  and they are causal (see lemmas 2.5.3 and 2.5.6), therefore  $\hat{V}_L^\sim, \hat{V}_R^\sim \in H^\infty$ .

Since  $\hat{V}_L \hat{V}_L^\sim = \bar{C}_L \bar{C}_L^+$  and  $\hat{V}_R \hat{V}_R^\sim = \bar{B}^+ \bar{B}$  and  $\bar{C}_L \bar{C}_L^+$  and  $\bar{B}_u^+ \bar{B}_u$  are orthogonal projections, we have that

$$\check{T}_2 \in H^\infty \iff \bar{T}_2 \in H^\infty$$

Note that if  $(A, B)$  is controllable then  $(e^{A_u h}, \bar{B}_u)$  is controllable (see Lemma 5.4.25). Then  $\bar{T}_2 \in H^\infty$  iff  $\bar{C}_L = 0$ . Now,  $\bar{C}_L = 0$  iff  $\bar{C}_L^* \bar{C}_L = 0$ , which is equivalent to  $\bar{C}_{vu}^* \bar{C}_{vu} - P_u^* \bar{C}_H^+ (\bar{C}_H^*)^+ P_u = 0$ . Note that  $\bar{C}_H^+ (\bar{C}_H^*)^+ = (\bar{C}_H^* \bar{C}_H)^+$  because  $\bar{C}_H$  is a closed operator [5].  $\square$

*Proof of Lemma 5.4.27.* If assumptions  $\mathcal{A}_7$ – $\mathcal{A}_8$  are satisfied then  $\bar{H}_{cs}^\perp \in H^\infty$  exists (see Lemma 5.4.19). Now, similar to the proof of Lemma 5.4.26, we have

$$\bar{H}_{cs}^\perp \hat{V}_H^\sim \check{G}_u \in H^\infty \iff \bar{H}_{cs}^\perp \hat{V}_H^\sim \check{T}_b \in H^\infty$$

where

$$\check{V}_H^\sim(z)\check{T}_b(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} \frac{A_u}{- (\bar{C}_H^*)^+ P_u} & \frac{B_u}{0} \\ \hline zI & -I \end{array} \right]$$

Also similar to the proof of Lemma 5.4.26, we have

$$\bar{H}_{cs}^\perp \check{V}_H^\sim \check{G}_u \in H^\infty \iff \bar{H}_{cs}^\perp \bar{T} \in H^\infty.$$

where  $\bar{T}$  is a discrete system with state space

$$\bar{T} := \left( \begin{array}{c|c} e^{A_u h} & \bar{B}_u \\ \hline - (\bar{C}_H^*)^+ P_u & 0 \end{array} \right)$$

Using the above and the state-space representation of  $\bar{H}_{cs}^\perp$  given in Lemma 5.4.19, we have

$$\bar{H}_{cs}^\perp \bar{T} = \left( \begin{array}{cc|c} A_l & -B_l(\bar{C}_H^*)^+ P_u & 0 \\ 0 & e^{A_u h} & \bar{B}_u \\ \hline -\bar{D}^\perp \bar{C}_H & -\bar{D}^\perp (\bar{C}_H^*)^+ P_u & 0 \end{array} \right)$$

where  $A_l := Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H - L_{cs} \bar{D}^\perp \bar{C}_H$  and  $B_l := Ee^{A_H h} B_H \bar{D}^+ + L_{cs} \bar{D}^\perp$ . Apply a state transform  $\begin{bmatrix} I & X_l \\ 0 & I \end{bmatrix}$  such that

$$A_l X_l - X_l e^{A_u h} + B_l (\bar{C}_H^*)^+ P_u = 0 \quad (5.91)$$

Note that the eigenvalues of  $Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H$  which are unobservable from  $\bar{D}^\perp \bar{C}_H$  are the invariant zeros of the  $\bar{H}_{cs}$  (see Lemma 5.4.19(3)). Therefore, by Assumption  $\mathcal{A}_9$  and the suitable choice of  $L_{cs}$  (see Lemma 5.4.19),  $A_l$  and  $e^{A_u h}$  have no common eigenvalues. Hence, a unique solution of the Sylvester equation (5.91) exists. The state transformation leads to the following state-space representation of  $\bar{H}_{cs}^\perp \bar{T}$ .

$$\bar{H}_{cs}^\perp \bar{T} = \left( \begin{array}{cc|c} A_l & 0 & X_l \bar{B}_u \\ 0 & e^{A_u h} & \bar{B}_u \\ \hline -\bar{D}^\perp \bar{C}_H & \bar{D}^\perp \bar{C}_H X_l - \bar{D}^\perp (\bar{C}_H^*)^+ P_u & 0 \end{array} \right).$$

Note that if  $(A, B)$  is controllable then  $(e^{A_u h}, \bar{B}_u)$  is controllable (see Lemma 5.4.25). Then the right hand side of the above is in  $H^\infty$  iff all unstable modes are un-observable. As all unstable modes of  $A_l$  are already unobservable from  $\bar{D}^\perp \bar{C}_H$ , we have that  $\bar{H}_{cs}^\perp \check{V}_H^\sim \check{G}_u \in H^\infty$  iff

$$\bar{D}^\perp \bar{C}_H X_l - \bar{D}^\perp (\bar{C}_H^*)^+ P_u = 0. \quad (5.92)$$

$X_l$  happens to be independent of  $L_{cs}$  if  $\bar{H}_{cs}^\perp \check{V}_H^\sim \check{G}_u \in H^\infty$ . This is shown below.

$$A_l X_l - X_l e^{A_u h} + B_l (\bar{C}_H^*)^+ P_u$$

$$\begin{aligned}
&= \left( Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H - L_{cs} \bar{D}^\perp \bar{C}_H \right) X_l - X_l e^{A_u h} \\
&\quad + (Ee^{A_H h} B_H \bar{D}^+ + L_{cs} \bar{D}^\perp) (\bar{C}_H^*)^+ P_u \\
&= \left( Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H \right) X_l + Ee^{A_H h} B_H \bar{D}^+ (\bar{C}_H^*)^+ P_u \\
&\quad - L_{cs} \left( \bar{D}^\perp \bar{C}_H X_l - \bar{D}^\perp (\bar{C}_H^*)^+ P_u \right).
\end{aligned}$$

If  $\bar{H}_{cs}^\perp \bar{V}_H \check{G}_u \in H^\infty$  then using (5.92),  $A_l X_l - X_l e^{A_u h} + B_l (\bar{C}_H^*)^+ P_u$  equals

$$\left( Ee^{A_H h} - Ee^{A_H h} B_H \bar{D}^+ \bar{C}_H \right) X_l + Ee^{A_H h} B_H \bar{D}^+ (\bar{C}_H^*)^+ P_u.$$

□

*Proof of Lemma 5.4.36.* We have

$$\dot{H}_i^\sim(z) \check{M}_v(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & C_H^* C_v & 0 \\ 0 & A + LC_y & L \\ \hline -(B_H Z)^* & 0 & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix}$$

and

$$-\dot{V}(z) \check{M}_y(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & -P(\tau)LC_y & -P(\tau)L \\ 0 & A + LC_y & L \\ \hline -(B_H Z)^* & 0 & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix}$$

where  $\Omega_p := \begin{bmatrix} (E + B_H F)^* & 0 \\ 0 & I \end{bmatrix}$ . Hence,

$$\dot{M}_h(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & C_H^* C_v - P(\tau)LC_y & -P(\tau)L \\ 0 & A + LC_y & L \\ \hline -(B_H Z)^* & 0 & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix}$$

Using a time-varying state transform  $T(\tau) = \begin{bmatrix} I & P_1(\tau) \\ 0 & I \end{bmatrix}$ , where  $P_1(\tau)$  satisfy

$$\dot{P}_1(\tau) + A_H^* P_1(\tau) + P_1(\tau)(A + LC_y) + C_H^* C_v - P(\tau)LC_y = 0,$$

Now, we have the boundary condition

$$\begin{bmatrix} z \begin{bmatrix} (E + B_H F)^* & -(E + B_H F)^* P_1(0) \\ 0 & I \end{bmatrix} & - \begin{bmatrix} I & -P_1(h) \\ 0 & I \end{bmatrix} \end{bmatrix}$$

Decoupling the boundary condition by premultiplying with matrix  $\begin{bmatrix} I & P_1(h) \\ 0 & I \end{bmatrix}$ ,

we have

$$\begin{bmatrix} z \begin{bmatrix} (E + B_H F)^* & -(E + B_H F)^* P_1(0) - P_1(h) \\ 0 & I \end{bmatrix} & - \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \end{bmatrix}$$



Therefore  $(E + B_H F)^* P_1(0) = -P_1(h)$ , would guarantee de-coupled states.

If we take  $P_1(\tau) = P(\tau)$ , we have

$$\dot{P}(\tau) + A_H^* P(\tau) + P(\tau) A + C_H^* C_v = 0,$$

with boundary condition  $(E + B_H F)^* P(0) = -P(h)$ , which we know is true by (5.64). Hence

$$\begin{aligned} \dot{M}_h(z) &= \mathcal{J}_{0^+}^* \left[ \begin{array}{cc|c} -A_H^* & 0 & 0 \\ 0 & A + LC_y & L \\ \hline -(B_H Z)^* & -(B_H Z)^* P(\tau) & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix} \\ &= \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A + LC_y & L \\ \hline -(B_H Z)^* P_0 & 0 \end{array} \right] \begin{bmatrix} zI & -I \end{bmatrix} \end{aligned}$$

where we used  $T(\tau) \begin{bmatrix} -P(\tau)L \\ L \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix}$ . Since  $A + LC_y$  is Hurwitz,  $e^{(A+LC_y)h}$  is Schur. Therefore, it follows from Corollary 5.3.20 that  $\dot{M}_h \in H^\infty$ .  $\square$

*Proof of Corollary 5.4.37.* Using Lemma 5.4.5, we have

$$\begin{aligned} &\dot{H}_i^{\sim}(z) \check{N}_v(z) \check{N}_y^{\sim}(z) - \dot{V}(z) \\ &= \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & \begin{bmatrix} 0 \\ (Z_y C_y)^* \end{bmatrix} \\ \hline C_p & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix} - \dot{V}(z) \\ &= \mathcal{J}_{0^+}^* \left[ \begin{array}{ccc|c} -A_H^* & 0 & 0 & -P(\tau)LZ_y^{-1} \\ 0 & -A_H^* & -C_H^* C_v X & 0 \\ 0 & 0 & -A_L^* & (Z_y C_y)^* \\ \hline -(B_H Z)^* & -(B_H Z)^* & 0 & 0 \end{array} \right] \begin{bmatrix} z\Omega_1 & -I \end{bmatrix} \end{aligned}$$

where

$$\Omega_1 := \begin{bmatrix} (E + B_H F)^* & 0 & 0 \\ 0 & (E + B_H F)^* & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Applying a transform  $T = \begin{bmatrix} I & 0 & 0 \\ I & I & 0 \\ 0 & 0 & I \end{bmatrix}$ , we have

$$\begin{aligned} &\dot{H}_i^{\sim}(z) \check{N}_v(z) \check{N}_y^{\sim}(z) - \dot{V}(z) \\ &= \mathcal{J}_{0^+}^* \left[ \begin{array}{ccc|c} -A_H^* & 0 & 0 & -P(\tau)LZ_y^{-1} \\ 0 & -A_H^* & -C_H^* C_v X & -P(\tau)LZ_y^{-1} \\ 0 & 0 & -A_L^* & (Z_y C_y)^* \\ \hline 0 & -(B_H Z)^* & 0 & 0 \end{array} \right] \begin{bmatrix} z\Omega_2 & -T^{-1} \end{bmatrix} \end{aligned}$$

where

$$\Omega_2 := \begin{bmatrix} (E + B_H F)^* & 0 & 0 \\ -(E + B_H F)^* & (E + B_H F)^* & 0 \\ 0 & 0 & I \end{bmatrix}$$

Multiplying the boundary condition by a invertible matrix does not change system, therefore we multiply with  $S := T$ , therefore we have

$$\begin{aligned} & \tilde{H}_1^{\sim}(z) \tilde{N}_v(z) \tilde{N}_y^{\sim}(z) - \tilde{V}(z) \\ &= \mathcal{J}_{0^+}^* \left[ \begin{array}{ccc|c} -A_H^* & 0 & 0 & -P(\tau) L Z_y^{-1} \\ 0 & -A_H^* & -C_H^* C_d X & -P(\tau) L Z_y^{-1} \\ 0 & 0 & -A_L^* & (Z_y C_y)^* \\ \hline 0 & -(B_H Z)^* & 0 & 0 \end{array} \right] \begin{bmatrix} z\Omega_1 & -I \end{bmatrix} \\ &= \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix} \end{aligned}$$

□

*Proof of Theorem 5.4.38.* To obtain  $\hat{S}_{\alpha, \text{opt}}$  consider the STPBC realization of  $\hat{Y} := \tilde{H}_1^{\sim} \tilde{N}_v \tilde{N}_y^{\sim} - \tilde{V}$  given in Corollary 5.4.37 i.e.

$$\hat{Y}(z) = \mathcal{J}_{0^+}^* \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix}$$

where all eigenvalues of  $e^{-A_p h} \Omega_p$  are in the region  $|z| > 1$ . Define a sampler

$$\hat{T}(z) \check{u}(z) := \int_0^h e^{A_p(h-\sigma)} B \check{u}(z; \sigma) d\sigma.$$

To calculate  $\hat{S}_{\alpha, \text{opt}} = \text{proj}_{z^l \mathbb{H}^2} \hat{Y}$  consider

$$\begin{aligned} \hat{Y} \check{u}(z) &= C_p (z\Omega_p - e^{A_p h})^{-1} \hat{T}(z) \check{u}(z) \\ &= C_p (ze^{-A_p h} \Omega_p - I)^{-1} \hat{T} \check{u}(z) \\ &= -C_p \left( I + ze^{-A_p h} \Omega_p + \dots + (ze^{-A_p h} \Omega_p)^l \right) \hat{T} \check{u}(z) \\ &\quad - C_p (ze^{-A_p h} \Omega_p)^{l+1} \left( I + (ze^{-A_p h} \Omega_p) + \dots \right) \hat{T} \check{u}(z) \\ &= \text{proj}_{z^l \mathbb{H}^2} \hat{Y}(z) \check{u}(z) + C_p (ze^{-A_p h} \Omega_p)^{l+1} (ze^{-A_p h} \Omega_p - I)^{-1} \hat{T} \check{u}(z) \\ &= \text{proj}_{z^l \mathbb{H}^2} \hat{Y}(z) \check{u}(z) + C_p (ze^{-A_p h} \Omega_p)^{l+1} (z\Omega_p - e^{A_p h})^{-1} \hat{T} \check{u}(z) \end{aligned}$$

Now,

$$\text{proj}_{z^l \mathbb{H}^2} \hat{Y}(z) = \hat{Y}(z) - C_p (ze^{-A_p h} \Omega_p)^{l+1} (z\Omega_p - e^{A_p h})^{-1} \hat{T} \check{u}(z)$$

$$= \mathcal{J}_{0^+} \left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p(1 - (ze^{-A_p h} \Omega_p)^{l+1}) & 0 \end{array} \right] \begin{bmatrix} z\Omega_p & -I \end{bmatrix}$$

The rest of the proof follows from Lemma 5.4.4, Lemma 5.4.10, Theorem 5.4.28, and Corollary 5.4.31.  $\square$

*Proof of Lemma 5.4.41.* The integral equalities in (5.74) and (5.76) follows from Lemma 5.3.29.

Since  $\check{G}_v + \check{H}_i \check{M}_h \check{G}_y$  is causal and in  $L^\infty$ , it is in  $H^\infty$ . By Lemma 5.3.23, this further implies  $\check{G}_v + \check{H}_i \check{M}_h \check{G}_y \in H^2$  as its STPBC has no feed through term. Using Lemma 5.3.25, we have

$$\|\check{G}_v + \check{H}_i \check{M}_h \check{G}_y\|_{H^2}^2 = \frac{1}{h} \|\check{D}\|_{HS}^2 + \|\bar{Y}\|_{H^2}^2$$

where

$$\bar{Y} = \left( \begin{array}{c|c} \Upsilon_e e^{A_e h} & \Upsilon_e \bar{B}_e \\ \hline \bar{C}_e & 0 \end{array} \right) = \left( \begin{array}{cc|c} A_{ms} & Q_m & \bar{B}_{ms} \\ 0 & e^{A_u h} & \bar{B}_{mu} \\ \hline \bar{C}_{ms} & \bar{C}_{vu} & 0 \end{array} \right)$$

where  $Q_m := \begin{bmatrix} M_{1u} e^{A_u h} \\ 0 \\ 0 \end{bmatrix}$ . Now,  $\bar{Y} \in H^2$  as  $\check{G}_v + \check{H}_i \check{M}_h \check{G}_y \in H^2$  and  $\|\check{D}\|_{HS}$  is finite.

Since  $A_{ms}$  is Schur and  $e^{-A_u h}$  has all its poles in the closed unit disk of the complex plane, therefore the Sylvester equation (5.79) has a unique solution  $X_m$ .

Now, applying a state transform  $T := \begin{bmatrix} I & -X_m \\ 0 & I \end{bmatrix}$ , we have

$$\bar{Y} = \left( \begin{array}{cc|c} A_{ms} & 0 & \bar{B}_{ms} - X_m \bar{B}_{mu} \\ 0 & e^{A_u h} & \bar{B}_{mu} \\ \hline \bar{C}_{ms} & \bar{C}_{ms} X_m + \bar{C}_{mu} & 0 \end{array} \right).$$

Since  $\bar{Y} \in H^2$ ,  $\left( \begin{array}{c|c} e^{A_u h} & \bar{B}_{mu} \\ \hline \bar{C}_{ms} X_m + \bar{C}_{mu} & 0 \end{array} \right) = 0$ . Therefore,

$$\bar{Y} = \left( \begin{array}{c|c} A_{ms} & \bar{B}_{ms} - X_m \bar{B}_{mu} \\ \hline \bar{C}_{ms} & 0 \end{array} \right)$$

This shows that the STPBC of the system  $\check{G}_v + \check{H}_i \check{M}_h \check{G}_y$  given in (5.73) contains unobservable or uncontrollable poles that lie in the region  $|z| \geq 1$  of the complex plane.

Since  $A_{ms}$  is Schur, the rest of the proof follows from Lemma 5.3.28.  $\square$

*Proof of Lemma 5.4.42.* For a given integer  $k$ , we have  $\|z^k \check{G}\|_{L^2} = \|\check{G}\|_{L^2}$  for a system  $\check{G}$ . Therefore,

$$\|\text{proj}_{L^2 \setminus z^l H^2}(\check{H}_i \check{N}_v \check{N}_y - \check{V})\|_{L^2} = \|\check{P}\|_{L^2}$$

where

$$\check{P}(z) := \frac{1}{z^{(l+1)}} \text{proj}_{L^2 \setminus z^l H^2}(\check{H}_i \check{N}_v \check{N}_y - \check{V}).$$

Let  $\check{Y} := \check{H}_i \check{N}_v \check{N}_y - \check{V}$ . Then using (5.69) and (5.70), the adjoints are

$$\begin{aligned} \check{Y}^\sim(z) &= \left[ \begin{array}{c|c} -A_p^* & -C_p^* \\ \hline B_p^* & 0 \end{array} \right] \begin{bmatrix} zI & -\Omega_p^* \end{bmatrix} \mathcal{J}_{0^+} \\ \check{P}^\sim(z) &= \left[ \begin{array}{c|c} -A_p^* & -(C_p(e^{-A_p h} \Omega_p)^{l+1})^* \\ \hline B_p^* & 0 \end{array} \right] \begin{bmatrix} zI & -\Omega_p^* \end{bmatrix} \mathcal{J}_{0^+}. \end{aligned}$$

The samplers  $\check{Y}$  and  $\check{P}$  are anti-causal systems in  $L^\infty$ , therefore their conjugates are holds in  $H^\infty$ . Using Lemma 5.3.23 this further implies that holds  $\check{Y}^\sim$  and  $\check{P}^\sim$  are in  $H^2$  as well. Therefore,  $\|\check{Y}^\sim\|_{H^2} = \|\check{Y}\|_{L^2}$  and  $\|\check{P}^\sim\|_{H^2} = \|\check{P}\|_{L^2}$  can be obtained by Lemma 5.3.26.

To calculate the norms we need to evaluate the integral

$$\bar{B}_p \bar{B}_p^* = \int_0^h e^{-A_p \tau} B_p(\tau) B_p(\tau)^* e^{-A_p^* \tau} d\tau$$

The above integral is not straight forward as  $B(\tau)$  is not constant but a function of  $\tau$ . The rest of the proof is mainly devoted to evaluation of the above integral.

Using [28, Theorem 1] and  $A_p = -\begin{bmatrix} A_H^* & C_H^* C_v X \\ 0 & A_L^* \end{bmatrix}$ , we have that

$$e^{-A_p \tau} = e^{\begin{bmatrix} A_H^* & C_H^* C_v X \\ 0 & A_L^* \end{bmatrix} \tau} = \begin{bmatrix} e^{A_H^* \tau} & T_1(\tau) \\ 0 & e^{A_L^* \tau} \end{bmatrix}$$

where  $T_1(\tau) := \int_0^\tau e^{A_H^*(\tau-\sigma)} C_H^* C_v X e^{A_L \sigma} d\sigma$ . Recall  $A_L := A + LC_y$ . Therefore, using  $P(\tau)$  given in Lemma 5.4.34 as

$$P(\tau) = e^{-A_H^* \tau} P_0 e^{-A \tau} - \int_0^\tau e^{-A_H^* \sigma} C_H^* C_v e^{-A \sigma} d\sigma,$$

and  $B_p(\tau) = \begin{bmatrix} -P(\tau) L Z_y^{-1} \\ (Z_y C_y)^* \end{bmatrix}$ , we have that

$$e^{-A_p \tau} B_p(\tau) = e^{-\begin{bmatrix} -A_H^* & -C_H^* C_v X \\ 0 & -A_L^* \end{bmatrix} \tau} \begin{bmatrix} -P(\tau) L Z_y^{-1} \\ (Z_y C_y)^* \end{bmatrix}$$

$$= \begin{bmatrix} -P_0 e^{-A\tau} L Z_y^{-1} + T_2(\tau) L Z_y^{-1} + T_1(\tau) (Z_y C_y)^* \\ e^{A_L \tau} (Z_y C_y)^* \end{bmatrix}$$

where  $T_2(\tau) := e^{A_H^* \tau} \int_0^\tau e^{-A_H^* \sigma} C_H^* C_v e^{-A \sigma} d\sigma$ . Using [28, Theorem 1] again, we can show that the above is equal to

$$e^{-A_p \tau} B_p(\tau) = P_z e^{A_z \tau} B_z$$

Using the above, we have that

$$\begin{aligned} \int_0^h e^{-A_p \tau} B_p(\tau) B_p(\tau)^* e^{-A_p \tau} d\tau &= P_z \left( \int_0^h e^{A_z \tau} B_z B_z^* e^{A_z \tau} d\tau \right) P_z^* \\ &= P_z \Lambda_{22}^*(A_z^*, B_z^*) \Lambda_{12}(A_z^*, B_z^*) P_z^* \end{aligned}$$

where  $\Lambda_{22}$  and  $\Lambda_{12}$  are defined in Lemma 5.3.29. Now, using Lemma 5.3.26 we have

$$\begin{aligned} \|\dot{Y}^\sim\|_{\mathbb{H}^2} &= \left\| \left( \begin{array}{c|c} \Omega_p^* e^{-A_p^* h} & -C_p^* \\ \hline \bar{B}_p^* & 0 \end{array} \right) \right\|_{\mathbb{H}^2}, \\ \|\dot{P}^\sim\|_{\mathbb{H}^2} &= \left\| \left( \begin{array}{c|c} \Omega_p^* e^{-A_p^* h} & -C_{pm}^* \\ \hline \bar{B}_p^* & 0 \end{array} \right) \right\|_{\mathbb{H}^2} \end{aligned}$$

The rest of the proof follows from Lemma 5.3.28.  $\square$

# Chapter 6

## Conclusions and Recommendations

A system theoretic analysis of the signal processing problem lead to solutions with maximum generality. In this thesis, we used the sampled data system theory to analyze the signal processing problems. The lifting is a main technical tool in the sampled data system theory. The lifting technique enables us to treat discrete and analog signals in a common framework. In this way we avoid any prior assumption on the input analog signal. In fact, we never have to worry about bandlimitedness of input signals and Shannon's sampling theorem during our analysis and design processes. We also used system norm to measure our reconstruction performance.

In this thesis, we obtained a computationally efficient way to compute frequency truncated norms. We also obtained optimal non-causal downsamplers-and-holds, and optimal relaxed causal samplers. The proofs of optimality are done by using sampled-data system theory.

In Chapter 3, we presented a method to compute the frequency truncated norms. These norms are required for reconstruction error calculation in the sampled-data signal processing. We expressed these norms in terms of the matrix logarithm for stable and unstable proper rational linear continuous time invariant system (LCTI) systems. We also showed a relationship of the frequency truncated norms with the  $H^2$  norm of an LCTI system. The result of this chapter can be applied to other areas of system theory like model reduction.

In Chapter 4, we obtained a solution to the optimal non-causal downsampling problem using sampled-data system theory. Specifically, we obtained an optimal non-causal stable downsampler-and-hold given the fast sampler and the signal model in the sampled-data setup. Here we used  $L^2$  and  $L^\infty$  optimality criteria. In the end, we also described a way to do optimal non-causal downsampling in the presence of noise. We also obtained expressions for the error for different optimality criteria in the optimal non-causal downsampling problem.

In Chapter 4, we obtained a solution to the optimal relaxed causal sampling

problem i.e. we obtained a stable optimal relaxed-causal sampler given a hold and a LCTI signal model. We provided the conditions of existence of the solution of the optimal relaxed causal sampling problem. We also gave some fundamentals of state space representation with two point boundary conditions (STPBC). STPBC representations are a useful way to represent a large class  $h$ -time shift invariant system. It includes all proper rational LCTI system as well. We also used this representation to solve optimal sampling problems and we obtained an easily computable optimal relaxed causal sampler and the optimal error.

## 6.1 Recommendations

From the topics that are discussed in the thesis, there are lot of potential improvements and problems for further research. These are discussed below.

1. A closed form expression for frequency truncated norm provided in Chapter 3 is sufficient if frequency truncation happens at finite number of frequencies. However, to obtain a closed form expression of reconstruction error in the  $L^2$  downsampling problem, we need to calculate the norm of a frequency response that is truncated at infinite number of frequencies. A closed form expression for such a response can be good objective for further research.
2. In Chapter 4, we just obtained the optimal hybrid interpolator in the lifted frequency domain at each  $\theta \in [-\pi, \pi]$ . However, we ignored the question of the Lebesgue measurability of the optimal hybrid interpolator. Thus a measure theoretic formulation of the optimal downsampling solutions in Chapter 4 can be investigated further.
3. In Chapter 4, downsampling of a discrete signal is done by an integer downsampling factor. A natural enhancement is to do optimal downsampling by a rational downsampling factor.
4. Also, the method we provided in Chapter 5 to compute the optimal relaxed causal sampler and the error for the systems given by the STPBC representation, have some numerical issues. We have not explored these issues in this thesis. Hence, there are some serious numerical issues to be looked at.
5. In Chapter 5, we took a causal signal model and hold. The effect of relaxing this causality constraint on the optimal relaxed causal sampler needs further investigation.

# Appendix A

## State-space representation

This appendix is a short review of state-space representations. State-space representation is an internal description of systems where we can define the system outputs at a given time in terms of some internal variables (known as state-variables) and inputs of the systems up to the given time. If number of state variables are finite then such systems are known as finite-dimensional systems. For details see [71, chapter 3,21], [57, appendix A] and [27, chapter 13].

### A.1 Continuous time system

If an analog system  $\mathcal{G}$  is a finite-dimensional linear continuous time invariant system then we can write the description of system in term of the following differential equations

$$y = \mathcal{G}u \quad : \quad \begin{cases} \frac{dx}{dt} = Ax + Bu \\ y = Cx + Du \end{cases} \quad (\text{A.1})$$

where  $x(t) \in \mathbb{R}^n$  is called *state*,  $u(t) \in \mathbb{R}^m$  and  $y(t) \in \mathbb{R}^p$ . Here the  $A$ ,  $B$ ,  $C$  and  $D$  are constant matrices.

The system  $\mathcal{G}$  given by (A.1) is represented by the following notation in the thesis

$$\mathcal{G} = \left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right].$$

Also in the Laplace domain the system  $\mathcal{G}$  is represented as

$$G(s) = D + C(sI - A)^{-1}B.$$

The above is also known as the *transfer-function* of the system  $\mathcal{G}$ .



## A.2 Discrete time system

If a discrete system  $\bar{\mathcal{G}}$  is a finite-dimensional linear shift invariant system then we can write the description of system in term of the following difference equations

$$\bar{y} = \bar{\mathcal{G}}\bar{u} \quad : \quad \begin{cases} \bar{x}[n+1] = A\bar{x}[n] + B\bar{u}[n] & n \in Z \\ \bar{y}[n] = C\bar{x}[n] + D\bar{u}[n] \end{cases} \quad (\text{A.2})$$

where  $\bar{x}[n] \in \mathbb{R}^n$  is called *state*,  $u[n] \in \mathbb{R}^m$  and  $y[n] \in \mathbb{R}^p$ . Here the  $A$ ,  $B$ ,  $C$  and  $D$  are constant matrices.

The system  $\bar{\mathcal{G}}$  given by (A.2) is represented by the following notation in the thesis

$$\mathcal{G} = \left( \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right).$$

Also in the  $z$ -domain the system  $\bar{\mathcal{G}}$  is represented as

$$\bar{G}(z) = D + C(zI - A)^{-1}B.$$

The above is also known as *transfer-function* of the system  $\mathcal{G}$ .

The *Invariant zeros* of the discrete system  $\bar{\mathcal{G}}$  given by state-space (A.2) are those values of  $z \in \mathbb{C}$  where the matrix

$$\begin{bmatrix} A - zI & B \\ C & D \end{bmatrix}$$

loses its normal rank (see [14, §8.4.2] and [71, Chapter 3]). The above matrix is known as the *Rosenbrock system matrix*.

For the definition of controllability, observability, detectability and stabilizability of systems represented in state-space see [57, appendix A] or [71, chapter 3,21].

# List of notations

- $\delta$  Dirac delta function
- $\bar{\delta}$  For all integer  $n$ ,  $\bar{\delta}[n] := \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$
- $\lceil x \rceil$  the smallest integer greater than or equal to  $x \in \mathbb{R}$
- $\lfloor x \rfloor$  the largest integer less than or equal to  $x \in \mathbb{R}$
- $\mathbb{1}(t)$   $\mathbb{1}(t) := \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases}$
- $\mathbb{1}_{\mathfrak{A}}(t)$  For a given set  $\mathfrak{A}$ , indicator function  $\mathbb{1}_{\mathfrak{A}}(t) := \begin{cases} 1 & t \in \mathfrak{A} \\ 0 & t \notin \mathfrak{A} \end{cases}$
- Arg principal argument function
- $\mathbb{C}$  the set of complex numbers
- $\mathbb{C}^-$  open left half complex plane
- $\mathbb{C}^n$  For a positive integer  $n$ ,  $\mathbb{C}^n := \left\{ \begin{bmatrix} w_0 \\ \vdots \\ w_{n-1} \end{bmatrix} \mid w_0, \dots, w_{n-1} \in \mathbb{C} \right\}$
- $C^2(\mathbb{B}, \mathcal{H})$   $\{x : \mathbb{B} \rightarrow \mathcal{H} \mid x \text{ is continuous \& } \int_{\mathbb{B}} \|x(t)\|_{\mathcal{H}}^2 dt < \infty\}$  (here  $\mathbb{B} \subseteq \mathbb{R}$  and  $\mathcal{H}$  is a Hilbert space)
- $\mathbb{D}$  unit disk  $\{z \in \mathbb{C} : |z| < 1\}$
- $\bar{\mathbb{D}}$  closed unit disk  $\{z \in \mathbb{C} : |z| \leq 1\}$
- $\exp\{A\}$   $e^A$
- $\|A\|_{HS}$  the Hilbert-Schmidt norm of operator  $A$
- Im  $A$  image of operator  $A$

$\ A\ $	induced norm of operator $A$
$\ A\ _\infty$	induced norm of operator $A$
$\overline{\text{Im}}A$	closure of image of operator $A$
$j$	$\sqrt{-1}$
$\text{Ker } A$	kernel of operator $A$
$L^2[0, h)$	space of square integrable functions defined on $[0, h)$
LCTI	linear continuous time invariant
LDTI	Linear discrete time invariant
$\mathbb{M}$	the set of integers $\{0, 1, \dots, M - 1\}$ for any $M \in \mathbb{Z}^+$
$\mathbb{N}$	set of positive integers and zero
$\mathbb{N}_A$	For an operator $A$ , $\mathbb{N}_A := \begin{cases} \{0, 1, \dots, \text{rank } A - 1\} & \text{If rank } A \text{ is finite} \\ \mathbb{N} & \text{If rank } A \text{ is infinite} \end{cases}$
$\text{proj}_{\mathcal{A}}$	orthogonal projector on Hilbert space $\mathcal{A}$
$\bar{\mathbb{R}}^-$	$\mathbb{R}^- \cup \{0\}$
$\bar{\mathbb{R}}^+$	$\mathbb{R}^+ \cup \{0\}$
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^-$	the set negative real numbers
$\mathbb{R}^+$	the set positive real numbers
$\text{rank } A$	rank of operator $A$
ROC	Region of convergence
$\text{sinc}$	$\text{sinc}(x) := \frac{\sin(\pi x)}{\pi x}$
$\text{sinc}_h$	$\text{sinc}_h(x) := \frac{\sin(\frac{\pi}{h}x)}{\frac{\pi}{h}x}$
SVD	singular value decomposition
$\mathbb{T}$	unit circle $\{z \in \mathbb{C} :  z  = 1\}$
$\omega_k$	$\frac{\theta + 2\pi k}{h}$ for all $\theta \in [-\pi, \pi]$ and $k \in \mathbb{Z}$
$\mathbb{Z}$	set of integers
$\mathbb{Z}^+$	set of positive integers

$\mathbb{Z}_l^+$  set of all integers greater than or equal to integer  $l$

$\mathbb{Z}^-$  set of negative integers

$\mathbb{Z}_l^-$  set of all integers smaller than integer  $l$



# Bibliography

- [1] S. Ashida, M. Nagahara, and Y. Yamamoto. Audio signal compression via sampled-data control theory. In *SICE Annual Conference*, August 2003.
- [2] B.A. Bamieh and J.B. Pearson. A general framework for linear periodic systems with applications to  $H^\infty$  sampled-data control. *IEEE Trans. Automat. Control*, 37:418–435, 1992.
- [3] B.A. Bamieh and J.B. Pearson. The  $H^2$  problem for sampled-data systems m for sampled-data systems. *Syst. Control Lett.*, 19:1–12, July 1992.
- [4] H. Bart, I. Gohberg, M.A. Kaashoek, and A.C.M. Ran. *A State Space Approach to Canonical Factorization with Applications*. Birkhäuser/Springer, 2010.
- [5] R. Bouldin. The pseudo-inverse of a product. *SIAM Journal on Applied Mathematics*, 24(4):pp. 489–495, June 1973.
- [6] J. Braslavsky, G. Meinsma, R. Middleton, and J. Freudenberg. On a key sampling formula relating the laplace and transforms. *SCL*, 29:181–190, 1997.
- [7] J.V. Candy. *Model-Based Signal Processing*. IEEE Press, 2006.
- [8] J.B. Conway. *A Course in Functional Analysis*. Springer-Verlag, 1985.
- [9] R.E. Crochiere and L.E. Rabiner. Interpolation and decimation of digital signals- a tutorial review. *Proc. IEEE*, 69(3):300–331, March 1981.
- [10] P.I. Davies and N.J. Higham. A schur-parlett algorithm for computing matrix functions. *Siam J. Matrix Anal. Appl.*, 25(2), 2003.
- [11] J. Doyle, B. Francis, and A. Tannenbaum. *Feedback Control Theory*. Macmillan, 1992.
- [12] G.E. Dullerud. *Control of Uncertain Sampled-Data Systems*. PhD thesis, University of Cambridge, 1994.

- [13] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer academic, 1996.
- [14] S. Fadali and A. Visioli. *Digital Control Engineering: Analysis and Design*. Academic/Elsevier, 2009.
- [15] G.B. Folland. *Introduction to Partial Differential Equations*. Princeton University Press, second edition, 1995.
- [16] B.A. Francis. *A Course in  $H_\infty$  Control Theory*. Springer-Verlag, 1987.
- [17] W. Gawronski and J.N. Juang. Model reduction in limited time and frequency intervals. *Int. J. Systems Sciences*, 21(2):349–376, 1990.
- [18] I. Gohberg and M.A. Kaashoek. Time varying linear systems with boundary conditions and integral operators. I. the transfer operator and its properties. *Integral Equations and Operator Theory*, 7:325–391, 1984.
- [19] N.J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, 2008.
- [20] H. Ishii, Y. Yamamoto, and B.A. Francis. Sample-rate conversion via sampled-data  $H^\infty$  control. In *IEEE-CDC*, volume 4, pages 3440 –3445, 1999.
- [21] A.B. Israel and T.N.E. Greville. *Generalized Inverses: Theory and Applications*. Springer, second edition, 2003.
- [22] H. Kakemizu, M. Nagahara, A. Kobayashi, and Y. Yamamoto. Noise reduction of jpeg images by sampled-data  $H^\infty$  optimal  $\epsilon$  filters. In *SICE Annual Conference*, August 2005.
- [23] P.P. Khargonekar and Y. Yamamoto. Delayed signal reconstruction using sampled-data control. In *IEEE-CDC*, volume 2, pages 1259 –1263, December 1996.
- [24] M.M. Konstantinov. *Perturbation theory for matrix equations*. Elsevier, 2003.
- [25] A.J. Krener. Boundary value linear systems. *Asterisque*, 75:149 – 165, 1980.
- [26] M. Kristalny. *Exploiting Previewed Information in Estimation and Control*. PhD thesis, Technion, 2010.
- [27] B.P. Lathi. *Signal processing and linear systems*. Oxford University Press, 2000.
- [28] C.V. Loan. Computing integrals involving the matrix exponential. *IEEE-AC*, 23(3):395 – 404, jun 1978.

- [29] G. Meinsma and L. Mirkin. Sampling from a system-theoretic viewpoint. Memorandum, Department of Applied Mathematics, University of Twente, 2009.
- [30] G. Meinsma and L. Mirkin. Sampling from a system-theoretic viewpoint: Part I—concepts and tools. *IEEE Trans. Signal Processing*, 58(7):3578–3590, July 2010.
- [31] G. Meinsma and L. Mirkin. Sampling from a system-theoretic viewpoint: Part II—non-causal solutions. *IEEE Trans. Signal Processing*, 58(7):3591–3606, July 2010.
- [32] G. Meinsma and L. Mirkin. Yet another discrete-time  $H^\infty$  fixed-lag smoothing solution. In *Proc. of 20th MTNS Symposium*, Melbourne, Australia, July 2012.
- [33] G. Meinsma and H. S. Shekhawat. Frequency-truncated system norms. *Automatica*, 47(8):1842 – 1845, 2011.
- [34] G. Meinsma and H.S. Shekhawat. Truncated norms and limitations on signal reconstruction. In *Proc. of 19th MTNS Symposium*, pages 1135–1140, Budapest, Hungary, July 2010.
- [35] L. Mirkin. Transfer functions of sampled-data systems in the lifted domain. In *CDC-ECC*, pages 5180 – 5185, dec. 2005.
- [36] L. Mirkin and Z.J. Palmor. A new representation of the parameters of lifted systems. *IEEE Trans. Automat. Control*, 44(4):833 –840, apr 1999.
- [37] L. Mirkin, H. Rotstein, and Z.J. Palmor.  $H^2$  and  $H^\infty$  design of sampled-data systems using lifting. Part I: General framework and solutions. *SIAM Journal on Control and Optimization*, 38(1):175–196, 1999.
- [38] L. Mirkin, H. Rotstein, and Z.J. Palmor.  $H^2$  and  $H^\infty$  design of sampled-data systems using lifting. Part II: Properties of systems in the lifted domain. *SIAM Journal on Control and Optimization*, 38(1):197–218, 1999.
- [39] L. Mirkin and G. Tadmor. On geometric and analytic constraints in the  $H^\infty$  fixed-lag smoothing. *IEEE-AC*, 52(8), Aug 2007.
- [40] L. Mirkin and R. Zaslavsky. A frequency-domain solution to the sampled-data  $H^2$  smoothing problem. In *CDC-ECC*, pages 5186 – 5191, dec. 2005.
- [41] M. Nagahara. *Multirate Digital Signal Processing via Sampled-Data  $H$ -Infinity Optimization*. PhD thesis, Kyoto University, 2003.
- [42] M. Nagahara. YY filter - a paradigm of digital signal processing. In *Perspectives in Mathematical System Theory, Control, and Signal Processing*, volume 398 of *Lecture Notes in Control and Information Sciences*, pages 331–340. Springer Berlin / Heidelberg, 2010.



- [43] M. Nagahara and Y. Yamamoto. A new design for sample-rate converters. In *IEEE-CDC*, volume 5, pages 4296–4301, 2000.
- [44] B.S. Nagy, H. Bercovici, C. Foiaş, and L. Kérchy. *Harmonic Analysis of Operators on Hilbert Space*. Springer, 2010.
- [45] S. Parrott. On a quotient norm and the sz.-nagy-foias lifting theorem. *Journal of Functional Analysis*, 30(3):311 – 328, 1978.
- [46] J.G. Proakis and D.G. Manolakis. *Digital signal processing: Principles, algorithms, and applications*. Prentice-Hall, 2001.
- [47] B.P. Rynne and M.A. Youngson. *Linear functional analysis*. Springer-Verlag, first edition, 2000.
- [48] E.B. Saff and A.D. Snider. *Fundamentals of Complex Analysis with Applications to Engineering, Science, and Mathematics*. Pearson Education, third edition, 2003.
- [49] C.E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.
- [50] H.S. Shekhawat and G. Meinsma. A sampled-data approach to optimal non-causal downsampling. In preparation.
- [51] H.S. Shekhawat and G. Meinsma. A sampled-data approach to optimal relaxed-causal sampling. In preparation.
- [52] H.S. Shekhawat and G. Meinsma.  $L^2$  and  $L^\infty$  optimal downsampling from system theoretic viewpoint. In *Proc. of 20th MTNS Symposium*, Melbourne, Australia, July 2012.
- [53] H.S. Shekhawat and G. Meinsma. Optimal relaxed causal sampler using sampled-data system theory. In *Proc. of 20th MTNS Symposium*, Melbourne, Australia, July 2012.
- [54] G. Strang. *Linear algebra and its applications*. Thomson learning, third edition, 1988.
- [55] W. Sun, K.M. Nagpal, and P.P. Khargonekar.  $H_\infty$  control and filtering for sampled-data systems. *IEEE-AC*, 38(8):1162 –1175, aug 1993.
- [56] C. Tongwen and B.A. Francis. Design of multirate filter banks by  $H^\infty$  optimization. *IEEE Trans. Signal Processing*, 43(12):2822 –2830, dec 1995.
- [57] C. Tongwen and B.A. Francis. *Optimal Sampled-Data Control Systems*. Springer-Verlag New York, Inc., 1995.
- [58] M.K. Tsatsanis and G.B. Giannakis. Principal component filter banks for optimal multiresolution analysis. *IEEE-SP*, 43(8):1766 –1777, aug 1995.

- [59] M. Unser. On the optimality of ideal filters for pyramid and wavelet signal approximation. *IEEE-SP*, 41(12):3591–3596, dec 1993.
- [60] M. Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, april 2000.
- [61] P.P. Vaidyanathan. Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial. *Proc. IEEE*, 78(1):56–93, January 1990.
- [62] M. Vidyasagar. *Control System Synthesis: A Factorization Approach*. MIT Press, Cambridge, MA, 1985.
- [63] C.R. Vogel. *Computational methods for inverse problem, volume 10*. SIAM, 2002.
- [64] S. Weiland. *Theory of Approximation and Disturbance Attenuation for Linear Systems*. PhD thesis, Rijksuniversiteit Groningen, 1991.
- [65] Y. Yamamoto, H. Fujioka, and P.P. Khargonekar. Signal reconstruction via sampled-data control with multirate filter banks. In *IEEE-CDC*, volume 4, pages 3395–3400, dec 1997.
- [66] Y. Yamamoto, M. Nagahara, and P.P. Khargonekar. A brief overview of signal reconstruction via sampled-data  $H^\infty$  optimization. *Applied and Computational Mathematics*, 11(1):3–18, 2012.
- [67] K. Yosida. *Functional Analysis*. Springer-Verlag, sixth edition, 1980.
- [68] D. Youla. On the factorization of rational matrices. *IRE Trans. on Information Theory*, 7(3):172–189, july 1961.
- [69] N. Young. *An Introduction to Hilbert Space*. Cambridge University Press, 1988.
- [70] K. Zhou and J.C. Doyle. *Essentials of robust control*. Prentice-Hall, 1998.
- [71] K. Zhou, J.C. Doyle, and K. Glover. *Robust and optimal control*. Prentice Hall, 1996.



# Summary

The main objective in this thesis is to design optimal samplers, downsamplers and interpolators (holds) which are required in signal processing. The sampled-data system theory is used to fulfill this objective in a generic setup.

Signal processing, which includes signal transmission, storage and analysis, plays a significant role in a human society. Since ages mankind is trying to find better and better means of signal transmission. Sending signal over a long distance was a challenge in ancient time. However with the advent of electrical/optical signals and electromagnetic waves, signal transmission is a matter of seconds now even over long distances. In the later half of the twentieth century, the digital revolution changed the scenario of signal processing. Digital signal processing provides better quality, low cost, ease of implementation and reconfigurability. In digital signal transmission the original analog signal is sampled using a sampler before transmission. At the receiver side, the received signal is processed digitally to remove unwanted signals like noise and then it is interpolated using a hold to obtain a reconstructed signal. This reconstructed signal must look like our original analog signal. Most of the systems like mobile, TV etc. nowadays use digital signal transmission techniques. Using digital techniques it is also easy to store and analyze signals.

Digital signal processing needs sampling of the original signal. Therefore, a fundamental question is: can we reconstruct the original signal from its samples using a hold? This is known as the signal reconstruction problem. The most famous answer of the signal reconstruction problem is given by Shannon's sampling theorem for the bandlimited analog signals. Bandlimitedness rarely happens in practice therefore Shannon's sampling theorem is not enough unless we apply some filters to make our analog signal bandlimited. Hence, researchers started looking at the signal reconstruction problem as a mathematical optimization problem from system theoretical viewpoint i.e. how to design samplers and holds such that the reconstructed signal resembles the original signal (measured in some norm sense). Here it is assumed that the spectrum of the signals are known. Sampled-data system theory is such an approach i.e. it is used to solve the signal reconstruction problem where the assumption of bandlimitedness is not required. It also enables us to obtain the solution with greater generality. A distinctive feature of the sampled-data system theory is that it optimizes the analog performance. This

approach is much closer to reality as most of the signals we use are analog in nature and utilized in the analog domain. Another distinctive feature of sampled-data system theory is the use of signal model to describe the spectrum of the original analog signals. The choice of signal model depends upon several factors like ease of implementation, accuracy required in signal reconstruction and the information available about signals. One advantage of using sampled-data system theory in the design process is that we can calculate the reconstruction error without any practical implementation.

Calculation of the reconstruction error boils down to calculation of the frequency truncated norm if the signal models are linear continuous time invariant (LCTI). In Chapter 3, we obtained closed form expressions to calculate the frequency truncated norms if the LCTI signal model is given by state-space. These methods are easy to implement in Matlab. The use of these closed form expressions to calculate the frequency truncated norms is not restricted to sampled-data system theory but also to other areas of system theory like model reduction.

Downsampling of the sampled signal is required in several signal processing applications like audio, image etc. This complicates our signal reconstruction problem because there is a downsampler in between the sampler and hold, and we have to work with multiple sampling rates. In Chapter 4 we provide a general formulation and solution of optimal downsampling in the sampled-data setup for all linear continuous time invariant signal models. Here we allow non-causal solutions. The effect of noise on the downsampling is also discussed in this chapter.

Non-causal solutions that have access to the infinite future, provide a theoretical limit to our solutions. However, they are rarely used in practice because of their unrealizability. Most of the time we design sampler and hold with causality or relaxed causality constraint. This is because it is practically impossible to have access to all future inputs at a given point of time. The constraint of causality/relaxed causality makes our problem a bit more difficult, but also more interesting. In Chapter 5, we provide a frequency domain abstract and implementable state-space solution to the optimal sampler design problem with a relaxed causality constraint.

In this thesis, we used sampled data system theory to answer downsampler and (relaxed causal) sampler design problems. However, the sampled-data system theory has the potential to answer many more interesting optimization problems arising in signal processing.

# Samenvatting

Het hoofddoel van dit proefschrift is het ontwerp van optimale samplers, down-samplers en holds die kunnen worden gebruikt in de signaalverwerking. Met behulp van de theorie van sampled-datasystemen wordt deze doelstelling op generieke wijze verwezenlijkt.

Signaalverwerking, waaronder signaaloverdracht, opslag en analyse van signalen, speelt een belangrijke rol in de samenleving. Sinds eeuwen is de mens bezig betere manieren van signaaloverdracht te bedenken. Het verzenden van signalen over een lange afstand was lange tijd een uitdaging. Echter met het gebruik van elektrische / optische signalen en elektromagnetische golven, is de duur van signaaloverdracht, ook over lange afstanden, een kwestie van seconden geworden. In de tweede helft van de twintigste eeuw heeft de digitale revolutie de signaalverwerking fundamenteel veranderd. Digitale signaalverwerking zorgt voor betere kwaliteit, lagere kosten, gemak van implementatie en herconfigureerbaarheid. In digitale signaaltransmissie wordt het oorspronkelijke analoge signaal eerst bemonsterd met behulp van een sampler en dan verzonden. Aan de ontvangtzijde wordt vervolgens het ontvangen signaal digitaal gefilterd om ongewenste componenten zoals ruis te verwijderen en daarna wordt het signaal met behulp van een hold geïnterpoleerd tot een analoog signaal. Dit gereconstrueerde signaal moet er uitzien als het oorspronkelijke analoge signaal. In de meeste systemen van vandaag de dag, zoals mobiele telefonie, TV etc. worden deze digitale technieken toegepast. Deze digitale technieken zijn ook van nut voor het opslaan van signalen en het analyseren ervan.

Digitale signaalverwerking vereist bemonsteren van het onderliggende analoge signaal. Daarom is een fundamentele vraag: “kunnen we het onderliggende signaal reconstrueren aan de hand van de monsters?” Dit staat bekend als het signaalreconstructie probleem. Het bekendste antwoord hierop is Shannon’s bemonsteringsstelling voor band-begrensd analoge signalen. Echter, band-begrensd komt in de praktijk maar nauwelijks voor en Shannon’s stelling volstaat daarom niet, tenzij we met een aantal filters het analoge signaal vooraf band-begrensd maken. Om deze ongewenste situatie te omzeilen zijn onderzoekers gaan kijken naar het algemene signaal-reconstructieprobleem als een wiskundig optimalisatieprobleem vanuit een systeemtheoretisch oogpunt, namelijk hoe kunnen we sampler en hold zodanig ontwerpen dat het gereconstrueerde signaal optimaal goed lijkt op het ori-

ginele signaal (gemeten in een bepaalde norm). In deze opzet wordt aangenomen dat het spectrum van de signalen bekend is. Sampled-data systeemtheorie is een dergelijke aanpak en deze vereist niet dat de signalen bandbegrensd zijn. Het stelt ons in staat om een oplossing te bepalen in grotere algemeenheid. Een onderscheidend kenmerk van de sampled-data systeemtheorie is dat het de analoge prestaties optimaliseert. Deze aanpak komt veel dichterbij werkelijkheid omdat de meeste signalen die we gebruiken analoog van aard zijn en worden gebruikt in het analoge domein. Een ander onderscheidend kenmerk van sampled-data systeemtheorie is het gebruik van een signaalmodel van het spectrum van het oorspronkelijke analoge proces. De keuze van signaalmodel is afhankelijk van verschillende factoren, zoals gemak van implementatie, nauwkeurigheid van de beschrijving en de beschikbaarheid van informatie over de signalen. Een voordeel van het gebruik van deze theorie is dat in het ontwerpproces de reconstructiefout kan worden berekend zonder praktische implementatie.

Berekening van de reconstructiefout komt neer op het berekenen van een frequentie afgekapte norm indien het signaalmodel lineair en continue-tijdinvariant is. In hoofdstuk 3 zijn gesloten uitdrukkingen verkregen van deze frequentie-afgekapte normen, voor de gevallen dat het signaalmodel een eindig-dimensionale toestandsrepresentatie heeft. De methoden zijn eenvoudig te implementeren in Matlab. Het gebruik van deze gesloten uitdrukkingen is niet beperkt tot sampled-dataproblemen, maar is ook op andere gebieden van de systeemtheorie van toepassing, zoals modelreductie.

Downsampling van bemonsterde signalen is nodig in verschillende toepassingen zoals audio, beeld etc. Downsampling bemoeilijkt het signaal-reconstructie probleem, want de aanwezigheid van een downsampler, tussen sampler en hold, betekent dat we moeten werken met meerdere bemonsteringsfrequenties. In hoofdstuk 4 geven we een algemene formulering en oplossing van het optimale downsamplingprobleem voor alle lineaire continue-tijdinvariante signaalmodellen. We laten hier niet-causale oplossingen toe. Het effect van ruis op downsampling wordt ook besproken in dit hoofdstuk.

Niet-causale systemen die toegang hebben tot de oneindige toekomst geven een theoretische limiet aan aan onze ontwerpproblemen, maar ze zijn vaak van weinig praktisch nut omdat ze niet geïmplementeerd kunnen worden. Vaker willen we samplers en holds ontwerpen die causaal zijn of in een beperkte mate kunnen anticiperen. Dit is omdat het praktisch onmogelijk is om toegang te hebben tot de volledige toekomst van ingangssignalen. De beperking tot causale of beperkt anticiperende samplers maakt ons ontwerpprobleem gecompliceerder, maar ook interessanter. In hoofdstuk 5 geven we een abstracte frequentiedomeinoplossing en een toepasbare toestandsrepresentatie-oplossing van het ontwerp van optimale samplers met beperkte anticipatie.

In dit proefschrift hebben we sampled-data systeemtheorie gebruikt om downsamplers en samplers met beperkte anticipatie te ontwerpen. Echter, deze sampled-datatheorie heeft de potentie om nog vele andere interessante optimalisatieproblemen in de signaalverwerking te ontrafelen.

# Acknowledgments

The last four years of my life were spent as a scientific journey where I got invaluable support and encouragement from my family, friends, colleagues and the University. I am grateful to all of them. It is difficult to include the names of all people who helped me in this unforgettable journey in just a few pages. Therefore, my list will be slightly compact. However, my heart is always thankful to each and every one of them.

First of all, I am deeply grateful to my supervisor Gjerrit Meinsma for his encouragement and support during my PhD. I learned a lot from his scientific attitude and patience. His interest and love for mathematics helped me in smooth transition from the field of engineering to the field of applied mathematics. I am also indebted to my promotor Anton Stoorvogel for many motivating discussions and timely involvement to keep my research on track. His helpful nature and patience helped me not only in mathematics but also in solving some administrative issues. I am deeply obliged to Gjerrit and Anton for proofreading the thesis and giving their invaluable comments. I would also like to thank Gjerrit for helping me in translating the summary into the Dutch language.

Furthermore, I am thankful to Leonid Mirkin for stimulating suggestions on my work that helped me in thinking about the problems from a new angle. I am very grateful to my PhD committee members Prof. Dr. S. Weiland, Prof. Dr. A. Bagchi, Dr. R.N.J. Veldhuis, Prof. Dr. H.J. Zwart and Prof. Dr. L. Mirkin, for their insightful comments and suggestions on my thesis. Their invaluable inputs helped in improving the quality of the thesis.

I am also thankful to the Faculty EWI and employees of the Department of Applied Mathematics for support and favors I received during my PhD. It was my pleasure to do research in the Department of Applied Mathematics especially with the Hybrid Systems group. Special thanks go to Prof. H.J. Zwart and Dr. J.W. Polderman for their help and support. I am thankful to Prof. A. Bagchi who always gave me a new perspective to look into the existing problems with his articles. I would also like to thank my colleagues Ove, Niels, Emad, Anindya, Mikael, Wishnu, Rahayu, Felix, David, and Lilya who made my journey more joyful. I am thankful to Saikat, Edson and Svetlana for the wonderful time we spent together as office-mates. I would like to thank Svetlana for her timely advises that helped me in focusing on my research. Thanks Svetlana for all your encouragement and



help. I would like to thank my colleague and friend Briti with whom I had many discussions on every possible issue. I am greatly indebted to Vijay Ambati who not only helped me settling in my office but also in the Netherlands. Thanks Vijay you really eased a lot of things for me.

I am very thankful to our secretary Marja who not only helped me in office work but also in spiritual enlightenment. Thanks Sakhya, you along with Anand Aadhar really helped me in understanding some of the spiritual aspects of life.

Thanks to the Dutch Institute of Systems and Control (DISC) that offered some great courses and gave me an opportunity to meet fellow PhD's of other Universities in the Netherlands. I made some great friends: Sasanka and Fiaz at DISC courses. I am really thankful to Sasanka for many discussions on topics not limited to just courses and research.

I am very indebted to the garden club at the University that gave me the opportunity to learn one of the oldest and fundamental work. Here I met Henny whose positive attitude towards life shall be an inspiration for my whole life.

I would like to say thanks to all members of the Indian Student Association for creating a homely atmosphere at the University. Special thanks to the organization AADHAAR for allowing me to associate with their activities. I have made some great friends during my stay in Enschede, which kept me busy socially. I am thankful to Vijay-Sangeeta, Vishwas-Sarika, Pramod-Vishaka, Sourabh-Neha, Digvijay-Aabha, Jigar-Falguni, Ambarish-Chetna, Chandu-Meenaksi, Abhinender-Shruti, Himanshu-Shabluka, Giri-Varsha, Omkar-Sampada, Pandu-Pallavi, and Shodan-Chaitanya for their support. I am very grateful to Nupur, Ramen-Priyanki and Jitendra-Neelam along with their daughter Bhairavi for their help I received during our stay. I will always cherish the wonderful time spend with you all. Special thanks to Bhairavi for helping me learning a lot of things.

I will be indebted to my cousin Surya Pratap Singh Rathore for helping me in designing the cover of my thesis. His concern for the cover design is deeply appreciated. In the last but not least, I am thankful to my parents for their continuous inspiration and encouragement. I am also very grateful to my grandparent and relatives for their blessings. I would also like to thank my wife Neeru to be on my side during the ups and downs in my PhD life. I am very thankful to my sister Poonam and brother Arun for their love and support.

Hanumant Singh Shekhawat  
November 2012

# Index

- $h'$ -aliased frequencies, 80
- $h$ -time shift invariance, 11
- $l$ -causal, 32
  
- Parseval identity, 23, 70
  
- adjoint, 24
- aliasing, 3
- analog
  - signal, 1
  - signal processing, 1
  - system, 1
- anti-causal, 33
  
- baseband dominant, 39
  
- causality, 27
  - classic, 27
  - input/output, 30
  - lifted, 31
  - relaxed, 32
  - strict, 30, 33
- coinner, 121
- conjugate, 24, 132
- continuous-time
  - signal, 1
  - system, 1
  
- digital
  - signal, 1
  - signal processing, 1
- direct feed-through term, 124, 127
- discrete
  - signal, 1
  - system, 1
- discrete key lifting formula, 66
  
- discrete lifted
  - $z$ -transform, 65
  - Fourier transform, 65
- discrete lifting
  - frequency domain, 65
  - time domain, 62
- dominant index set, 79
- downsampler, 61
  - discrete sampling function, 61
  - ideal, 61
- downsampling factor, 59, 61
- downsampling problem, 72
  
- frequency truncated norms
  - $\|\cdot\|_{\omega_l}$ , 41
  - $\|\cdot\|_{[\omega_l, \omega_u]}$ , 41
  
- Hardy space, 23
- hold, 13
  - hold function, 13, 61
- Hurwitz matrix, 34
- hybrid
  - interpolator, 62
  - signal processor, 136
  - system, 1
  
- impulse operator, 128
- impulse response, 12
- indicator function, 199
- inner-outer factorization, 119
  - inner factor, 119
- interpolator, 61
- Invariant zeros, 198
  
- Jordan canonical form, 43

- key lifting formula, 18
- lifted
  - z-transform, 18
  - downsampler, 64
  - Fourier transform, 18
  - frequency domain downsampler, 68
  - frequency domain hold, 21
  - frequency domain sampler, 21, 68
  - frequency domain system, 20
  - transfer function, 19
- lifting, 14
  - continuous(analog), 14
  - frequency domain, 17
- lifting period, 14
- Matrix functions, 44
- Nyquist
  - frequency, 2
  - interval, 81
- principal logarithm, 45
  - scalar, 43
- rank theorem, 71
- relaxed causal, 32
- Rosenbrock system matrix, 198
- sampled-data system, 11
- sampler, 12
  - ideal, 12
- sampling
  - function, 12
  - period, 12
- sampling operator, 128
- Schur matrix, 34
- shift invariance, 13
- signal
  - $l$ -causal, 28
  - causal, 28
  - discrete causal, 28
  - lifted  $l$ -causal, 32
  - lifted causal, 32
- signal generator, 3, 11
- signal model, 3
- signal norm
  - $C^2(0, h)$ , 128
  - $H^2$ , 23
  - $L^2$ , 22
  - $L^2(\mathbb{B})$ , 22
  - $L^2(\mathbb{B}, \mathbb{C}^n)$ , 22
  - $L^2(\mathbb{T})$ , 23
  - $L^2(\mathbb{T}, \mathcal{H})$ , 23
  - $\ell^2(\mathbb{Z}, L^2[0, h])$ , 22
  - $\ell^2(\mathbb{Z}, \ell^2(\mathbb{M}))$ , 69
  - $\ell^2_{h'}$ , 59
  - $\ell^2$ , 22
  - $\ell^2(\mathbb{B}, \mathcal{H})$ , 21
  - $\ell^2(\mathbb{Z})$ , 22
  - $\ell^2(\mathbb{Z}_l^+)$  and  $\ell^2(\mathbb{Z}_l^-)$ , 22
  - $(H^2)^\perp$ , 23
- singular vectors
  - left, 59
  - right, 59
- stability, 33
- state-space, 197
  - continuous-time, 197
  - discrete, 198
- state-variables, 197, 198
- static discrete system, 14
- STPBC, 124
  - addition, 126
  - adjoint, 126
  - equivalence, 125
  - multiplication, 126
  - similarity transformation, 126
  - well-posed, 124
- strictly (classic) causal, 30
- strictly (lifted) causal, 33
- system, 1
  - linear  $h$ -time shift invariant, 11
  - linear continuous time invariant, 12
  - linear discrete time invariant, 12
- system norm, 25
  - $z^l H^2$ , 27
  - $z^l H^\infty$ , 25
  - $H^2$ , 27

$H^\infty$ , 25

$L^2$ , 26

$L^\infty$ , 25

transfer function, 19, 197

truncation operator, 27, 28, 31

zero order hold, 13

    generalized, 13, 129

    ideal, 13, 129



**ISBN: 978-90-365-3473-4**