# FAST IDENTIFICATION FOR ROBUST ADAPTIVE CONTROL—A METRIC COMPLEXITY APPROACH

Lin Lin

B.Eng. (Tsinghua University), 1984

M.Eng. (Tsinghua University), 1987

Department of Electrical Engineering

McGill University

Montreal, Quebec, Canada

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

October 1993

ISBN 0-315-94669-5

# Fast identification for adaptive control
# —a metric complexity approach

*To my parents*

# Abstract

In this thesis, the problem of fast identification is formulated in the framework of the theory of metric complexity. Several complexity issues of fast identification are investigated.

Experiment design and model selection, two important components for achieving fast identification, are separated by splitting the estimation error into inherent and representation errors, which are functions of the experiment and the model sets respectively. The optimal inherent error, a measure of the time complexity of identification, is introduced as a notion of n-width (time n-width) related to the Gel'fand n-width. The optimal representation error is related to the Kolmogorov n-width. Estimates of the various n-widths are obtained systematically for a class of data sets covering many cases encountered in practical control problems.

The input design problem is further explored in cases where the input can be designed only to the extent of modifying its ensemble properties. The identifying capability of an input ensemble is characterized in general in terms of the gap metric. This general characterization is reduced to a certain spectrum flatness property of the input in the case of finite impulse response models. Bounds on the inherent error are given in terms of the n-width and spectrum flatness. Several robust identification algorithms are proposed as well.

It is shown that in the continuous-time case, although it is possible to identify a system arbitrarily accurately on an arbitrarily short time interval by increasing the

sampling rate and signal-to-noise ratio, the identification speed is limited for a fixed signal-to-noise ratio. An asymptotically accurate lower bound on the optimal identification speed is given, again in terms of Gel'fand n-width. A logarithmic integral characterization for the optimal inputs is obtained via the theory of quasianalytic functions. The representation and estimation problems for continuous-time systems are reduced to a discrete-time case. A causal reconstruction procedure is given, together with an error estimate.

Finally, the results on fast identification are applied to systems in which the law governing the evolution of the uncertain elements is not time invariant. Such systems can not be identified accurately. The inherent error is bounded in the case of slow time-variation and shown to increase with the variation rate.

# Résumé


Dans cette thèse, le problème d'idendification rapide est formulé dans le cadre de la théorie de complexité métrique. Quelques problèmes de complexité reliés à l'identification rapide sont étudiés.

La conception d'expériences et la sélection de modèles, deux composantes importantes pour arriver à une identification rapide, sont séparées en décomposant l'erreur d'estimation en l'erreur inhérente et celle de représentation qui sont respectivement des fonction de l'expérience et du modèle. L'erreur inhérente optimale, une mesure de la complexité en temps d'identification, est introduite comme une notion de n-ième épaisseur (n-ième épaisseur temporelle) relié au n-ième épaisseur de Gel'fand. L'erreur de représentation optimale est reliée au n-ième épaisseur de Kolmogorov. Les estimations de divers n-ièmes épaisseurs sont obtenues systématiquement pour une classe d'ensemble de données qui couvrent plusieurs problèmes pratiques de commande.

Le problème de conception d'entrée est étudié pour les cas où la conception est limitée à la modification des propriétés des ensembles. La capacité d'identification d'un ensemble d'entrées est en général caractérisée par la mesure de distance. Lorsque le modèle à réponse impulsionnelle finie est utilise, cette caractérisation générale est réduite à la propriété d'aplatissement spectral de l'entrée. Les bornes de l'erreur inhérente sont données en termes du n-ième épaisseur et de l'aplatissement spectral. En outre, quelques algorithmes stables sont proposés pour l'identification.

Il est démontré dans ce travail que la vitesse d'identification est limitée pour un

rapport signal sur bruit donné, bien qu'il soit possible d'identifier un système avec n'importe quelle précision en un temps arbitrairement court en augmentant la vitesse d'échantuonnage et le rapport signal sur bruit. En terme du n-ième épaisseur de Gel'fand, une borne inférieure, qui est précise asymptotiquement pour la vitesse optimale d'identification, est obtenue. Via la théorie des fonctions quasi-analytiques, nous avons obtenu une caractérisation en intégral logarithmique pour les entrées optimales. Les problèmes de représentation et d'estimation pour les systèmes en temps continu sont réduits à ceux des systèms en temps discret. Une procédure de reconstruction causale est donnée avec une estimation d'erreur.

Finalement, les résultats obtenus sur l'identification rapide sont appliqués aux systèmes dans lesquels les lois réagissant l'évolution des éléments incertains varient avec le temps. De tels systèmes ne peuvent pas être identifiés d'une façon précise. L'erreur inhérente est limitée dans le cas ou la variation temporelle est lente et elle est démontrée de s'accroître avec la vitesse de variation.

# Acknowledgements

I wish to express my sincere gratitude and appreciation to my supervisor, Prof. George Zames for his excellent guidance and constant encouragement which were essential to the completion of this work. His rigor and high standard in research, and keen insights and ability to pose fundamental questions concerning the essence of system theory proved to be extremely enlightening.

I would also like to thank Prof. Le Yi Wang for his friendship and extensive discussions on my research, especially on the material in Chapter 7. I am very grateful to Prof. Paul Koosis for insightful discussions on the theory of quasianalytic functions, and to Prof. Peter Caines for teaching me stochastic identification and adaptive control theory and bringing me many interesting papers on this subject.

Sincere thanks also go to my colleagues and friends in the System and Control group and Center of Intelligent Machines at McGill, who shared with me their time and made my stay a pleasant experience. Particularly, I would like to thank James Owen, Finn Wredenhagen, Wei Yuanjun, Mark Readman, Wu Kenong, Liu Zheng, Yair Matas, and Kaouthar Benameur.

Finally, I wish to extend my deepest thanks to my wife for her love, support, patience, and computer drafting skills.

# Claim of Originality

- The fast identification problem is posed. A link between identification and the metric complexity theory is established. Two notions of time n-width, related to the Gel'fand n-width, are introduced to characterize the time complexity of identification. The optimal estimation error is related to the maximum of the time and Kolmogorov n-widths.

- The various n-widths are estimated for a class of data sets and shown to be derivable from a common principle. For these data sets, the n-widths coincide, and the optimal estimation error equals the maximum of the n-widths in the sense of Gel'fand and Kolmogorov.

- Ensemble input design is investigated for fast identification. A general characterization of the identifying capability of the input is obtained. Upper and lower bounds on the inherent error are given in terms of a certain spectrum flatness property of the input in the case where a finite impulse response model is used.

- Two robust identification algorithms are proposed on the basis of the analytic center.

- It is shown that, in the noise-free case, it is possible to identify a stable LTI system exactly on an arbitrarily short continuous-time interval. A logarithmic integral characterization of the optimal inputs is obtained via the theory of quasianalytic functions.

- In the case where only corrupted output samples are available, an arbitrarily accurate identification for a system in a compact set can be achieved on an arbitrarily short interval by increasing the sampling rate and signal-to-noise ratio. This compactness condition is not dispensable. For a fixed sampling rate and signal to noise ratio, an asymptotically accurate lower bound on the inherent error is given in terms of the Gel'fand n-width.

- The representation and estimation problems for continuous-time systems are reduced to a discrete-time case. A causal reconstruction procedure is given with error estimation.

- Uncertainty principles for time varying system identification are obtained by using the results on fast identification.

# Contents

# List of Figures

# Notation

$\mathbb{C}$, $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{Z}_+$ denote the complexes, reals, integers and non-negative integers.

$\ell^p[a,b]$, $1 \le p \le \infty$, $-\infty < a \le b < \infty$, denotes the space of sequences of real numbers $f(t)$, $t$ being an integer in the interval $a \le t \le b$, satisfying $\|f\|_p := [\sum_{t=a}^{b} |f(t)|^p]^{1/p} < \infty$ for $1 \le p < \infty$, and $\|f\|_\infty := \sup_{t\in[a,b]} |f(t)| < \infty$.

$\mathbf{L}^p[a,b]$, $1 \le p \le \infty$, $-\infty < a \le b < \infty$, denotes the space of the real Lebesgue measurable functions on $[a,b]$, $f(t)$, for which $\|f\|_p := [\int_a^b |f(t)|^p]^{1/p} < \infty$ for $1 \le p < \infty$, and $\|f\|_\infty := ess\ \sup_{t\in[a,b]} |f(t)| < \infty$.

$\ell^p[a,b]$ and $\mathbf{L}^p[a,b]$ are extended in the usual way to the cases where the interval has one (or both) end points missing, such as $[a,b)$, or is (semi) infinite, such as $[a,\infty)$.

$\mathbf{H}^p(D_r)$ denotes the Hardy space on $D_r := \{z \in \mathbb{C} : |z| < r\}$, $r > 0$. $\|K\|_{\infty,r} := \sup_{z\in D_r} |K(z)|$ is the norm defined on $\mathbf{H}^\infty(D_r)$.

$\mathbf{H}^\infty(D) := \mathbf{H}^\infty(D_r)|_{r=1}$.

$\mathbf{H}^p(\Omega)$ denotes the Hardy space on the right half complex plane $\Omega$. $\|K\|_{\mathbf{H}^\infty} := \sup_{z\in\Omega} |K(z)|$ is the norm defined on $\mathbf{H}^\infty(\Omega)$.

$\mathbf{P}_{[n,m]}$ is the truncation operator on $\ell^p$, defined by $\left(\mathbf{P}_{[n,m]}f\right)(t) := f(t)$ for $t \in [n,m]$, and 0 otherwise.

$\|\mathbf{S}\|_{\mathbf{L}}$ is the norm of the largest function in the subset $\mathbf{S}$ of a normed space $\mathbf{L}$, i.e., $\|\mathbf{S}\|_{\mathbf{L}} := \sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}\}$.

$\mathbf{S}|_{[t_1,t_2)}$ is the subset of functions of $\mathbf{S}$ mapping $\mathbb{Z}$ to $\mathbb{R}$ with support in the interval $[t_1,t_2)$ of $\mathbb{Z}$, i.e., $\mathbf{S}|_{[t_1,t_2)} := \mathbf{S} \bigcap \mathbf{P}_{[t_1,t_2)}\mathbf{S}$

$Int(\cdot) : \mathbb{R} \to \mathbb{Z}$ is defined as the smallest integer strictly greater than an argument.

$\underline{\sigma}(\cdot)$ and $\overline{\sigma}(\cdot)$ denote the minimum and the maximum singular values of a matrix.

$Null(\cdot)$ denotes the null space of an operator, i.e., $Null\,(\mathbf{\Phi}) := \{k \in \mathbf{X} : \ \mathbf{\Phi}(\mathbf{k}) = \mathbf{0}\}$.

$\mathbf{U}^*$ is the dual space of $\mathbf{U}$.

$\rho(\cdot)$ denotes the variation rate of a time-varying system.

# Chapter 1

# Introduction

The essential property of an adaptive system is that it self-adjusts its parameters and structure to adapt to a changing environment and to improve its performance, using the information collected while it is evolving. How well the system can adapt to the changing environment is limited by how quickly it can identify these changes. Therefore, fast identification is of crucial importance to all adaptive systems.

Despite the long history of research on adaptive control and identification, and various adaptive control and identification theories and algorithms associated with the names of Astrom [2], Caines [5], Goodwin [11], Ljung [26], Morse [33, 34], etc, fast identification has received little attention in the past. It is the objective of this thesis to formulate and solve the fast identification problem using metric complexity theory introduced into control by Zames [51, 52].

## 1.1 What Is Involved in Fast Identification?

An identification procedure can be divided into two stages consisting of information acquisition and information processing. In the first stage, an input-output experiment is carried out and input-output data is collected. After the experiment, uncertainty

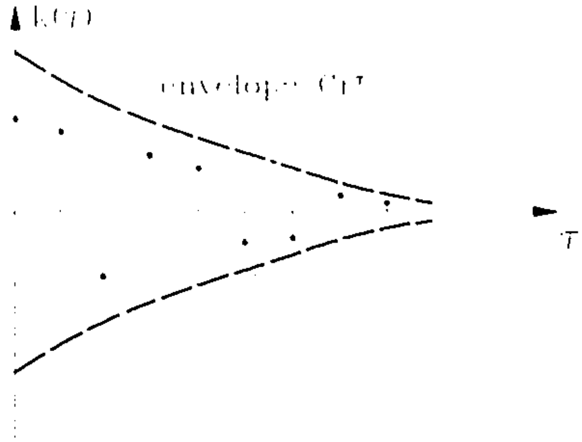


Figure 1.1: A priori information locates $k$ between two exponential envelopes

about the system is reduced. In the second stage, a representation of the available information is obtained. For robust controller design, the information is represented by a nominal system with a surrounding ball in a suitable normed space, e.g., $\mathbf{H}^\infty$ or $\ell^1$. To achieve fast identification, an identification procedure should acquire information quickly in the first stage and process information promptly in the second stage.

In the following we will use a typical example to illustrate how an experiment should be designed to maximize the amount of information contained in each observation, and how a nominal model should be chosen to reduce the complexity of an algorithm.

Consider the identification of a discrete-time stable LTI system. The system is considered as an operator on $\ell^\infty(-\infty,\infty)$. Its impulse response $k$ is in $\ell^1[0,\infty)$. The objective of identification is to obtain an estimate $k_{est}$ of $k$, and a bound on the estimation error in the $\ell^1$ norm.

Assume that the a priori information locates the impulse response $k$ in a given subset $\mathbf{S}_{prior}$ of $\ell^1$,

$$\mathbf{S}_{prior} := \left\{ k \in \ell^1[0,\infty) : \ |k(\tau)| \leq C r^\tau \ \forall \tau \in \mathbb{Z}_+ \right\},$$

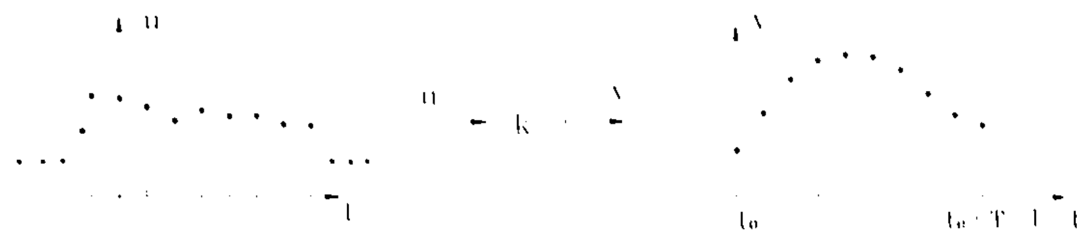

Figure 1.2: An identification experiment

as shown in Figure 1.1. An identification experiment as shown in Figure 1.2 is performed. A known input $u$ is applied to the system and the output observations start at time $t_0$. For simplicity, assume there is no additive noise on the output observations, i.e.,

$$y(t) = \sum_{\tau=0}^{\infty} k(\tau)u(t-\tau).$$

After $T-1$ sampling periods, all we know is that the true system $k_{true}$ must be in the set

$$\mathbf{S}^T(u) := \left\{ k \in \mathbf{S}_{prior} : \sum_{\tau=0}^{\infty} k(\tau)u(t-\tau) = y(t), \ \forall t \in [t_0, t_0+T) \right\}.$$

Since all the systems in $\mathbf{S}^T(u)$ are consistent with the a priori information and the input-output observations, they are not distinguishable from each other. Therefore, *$\mathbf{S}^T(u)$ represents the a posteriori uncertainty about the system at time $t_0+T-1$. For fast identification, the input should be designed so that $\mathbf{S}^T(u)$ shrinks quickly as time progresses.*

The most accurate nominal-ball type representation of the information at time $t_0+T-1$ (i.e., $k_{true} \in \mathbf{S}^T(u)$) is given by the smallest ball in $\ell^1$ covering $\mathbf{S}^T(u)$, with its center as the nominal and its radius as the estimation error. Unfortunately, in general this most accurate representation is impossible to realize in practice, as the nominal model can be any function in $\ell^1$ which is an infinite dimensional space, and therefore an infinite number of parameters might have to be identified to determine the nominal or to represent it.

This raises the question: how should a nominal be chosen to balance the trade-off between its accuracy and the efficiency of the algorithm? Here, the complexity of an algorithm is measured by the number of parameters it has to identify. *For fast identification, an algorithm should identify the least number of parameters possible while keeping the representation within a given error tolerance.*

If the nominal is to be chosen in a given n-parameter model set $\mathbf{X}_n \subset \ell^1[0,\infty)$, the most accurate nominal-ball representation of $\mathbf{S}^T(u)$ is given by the smallest ball centered in $\mathbf{X}_n$ covering $\mathbf{S}^T(u)$. For such a nominal, an algorithm only needs to identify $n$ parameters and the estimation error is

$$\inf_{k_{est}\in\mathbf{X}_n} \sup_{k\in\mathbf{S}^T(u)} \|k - k_{est}\|_{\ell^1}.$$

Note that since $\mathbf{S}^T(u)$ also depends on the true system, this minimum error is also a function of the true system $k_{true}$, which is unknown before the identification. To study the effects of the input and model set on the estimation error, we consider the worst-case error

$$e^T(u, \mathbf{X}_n) := \sup_{k_{true}\in\mathbf{S}_{prior}} \inf_{k_{est}\in\mathbf{X}_n} \sup_{k\in\mathbf{S}^T(u)} \|k - k_{est}\|_{\ell^1}.$$

Given an input $u$ and a model set $\mathbf{X}_n$, $e^T(u, \mathbf{X}_n)$ is the optimal worst-case estimation error an algorithm can achieve at time $t_0 + T - 1$. It is a measure of identification speed. For fast identification, the input and the model set should be designed to minimize the estimation error $e^T$. In this thesis, we will devise a general theory of design for the input and the model set so as to minimize $e^T$ for both the discrete-time and continuous-time cases. The system is assumed to be stable throughout the thesis.

## 1.2 Thesis Organization and Outline of Results

### 1.2.1 Thesis Organization

In the first half of the thesis, Chapter 2 to Chapter 5, the fast identification problem is studied for the discrete-time case. In Chapter 2, the problem is formulated as an input and model design problem, with the worst-case estimation error $e^T$ as the design criterion. This design problem is reduced to two standard optimization problems in the metric complexity theory, and the estimation error $e^T$, optimized over all bounded inputs and all n-parameter affine models, is related to the n-widths in the sense of Gel'fand and Kolmogorov, two standard notions in metric complexity theory. In Chapter 3, the optimal estimation error and several related notions of n-widths are computed for a class of data sets, and shown to be derivable from a common principle. For these data sets, the optimal estimation error is shown to be equal to the maximum of the Gel'fand and Kolmogorov n-widths of the data sets. The input design problem is further explored in Chapter 4 for the case where the input properties that are allowed to vary are constrained to be certain ensemble properties. A spectral characterization of the identifying capability of an input is obtained. In Chapter 5, several robust identification algorithms are proposed on the basis of convex optimization.

In Chapter 6 and 7, the fast identification problem is studied for the continuous-time case. It is shown that although it is possible to identify a continuous-time system arbitrarily accurately on an arbitrarily short time interval by increasing the sampling rate and the signal-to-noise ratio, identification speed is limited in practical control problems. The optimal estimation error is again related to the Gel'fand and Kolmogorov n-widths. This shows that, in both the discrete and continuous time cases, the time needed to identify an LTI system to certain accuracy increases with the metric complexity of the a priori data set. In other words, the identification speed is limited by the metric complexity of the a priori data set.

In the case where the system is not known to be time-invariant beforehand, it is shown in Chapter 8 that there is an irreducible uncertainty in identification if the system changes quickly in comparison with the optimal identification speed.

These results provide tools for input design and model selection for fast identification in practical control problems. More importantly, the establishment of the link between identification and metric complexity theory paves the way for unified theories of identification and feedback and of information-based adaptive control. (See the overview by Zames [57] for details.) The results on time-varying system identification (uncertainty principles) indicate the necessity of *robust* adaptive control for time-varying systems.

### 1.2.2 A More Detailed Outline of the Results

As shown in Section 1.1, the fast identification problem can be formulated as an input design and model selection problem, with the worst-case estimation error $e^T(u, \mathbf{X}_n)$ as the design criterion. With such a formulation in hand, the first question one may ask is: how should this rather complicated input and model design problem be solved? In Chapter 2, we answer this question by reducing the design problem to one of several standard optimization problems in the metric complexity theory. First, the input design and model selection are decoupled by splitting the worst-case estimation error $e^T(u, \mathbf{X}_n)$ into two parts, the *inherent* and *representation* errors, which depend on the input and the model set respectively. Next, the optimal inherent error over all bounded inputs is introduced as a notion of n-width, called the time n-width, which is similar to the Gel'fand n-width; and the optimal representation error over all n-parameter affine models is related to the Kolmogorov n-width. Then, the optimal worst-case estimation error is related to the maximum of the time and Kolmogorov n-widths of the a priori data set. The establishment of the link between the optimal worst-case estimation error and the n-widths not only reduces the fast identification problem to the estimation of the n-widths, but also shows that it is the metric complexity of the data sets that limits the optimal identification speed. These results are illustrated by several examples, in which the optimal worst-case estimation error equals the maximum of the two n-widths.

Since, the resolution of the fast identification problem involves the estimation of the various n-widths of the a priori data set, the problem would be substantially simplified if these n-widths could be obtained from a common principle. In Chapter 3, we estimate these n-widths and the optimal worst-case estimation error for a class of data sets, and show that they can be obtained from a common principle which captures a monotone decreasing property of these data sets. For such data sets, the optimal worst-case estimation error equals the maximum of the n-widths in Gel'fand and Kolmogorov senses; the optimal input is an impulse at the start of the observation interval; and the optimal affine model is an FIR model. If the observation location cannot be positioned advantageously, there is a loss of optimal identification speed, but the loss never exceeds a factor of $\pi$.

In practical on-line identification, the input is seldom free to be optimized; it can only be modified to the extent of having certain desirable ensemble properties, e.g., flat spectrum, by introduction of a dither signal. In Chapter 4, we will study how the input should be modified to be suitable for fast identification. First, a characterization of the identifying capability of an input is given in the gap metric. Then it is shown that this characterization is related to certain "spectrum flatness" properties of the input, in the case where an FIR (finite impulse response) model is used. In this case, the worst-case inherent error is bounded both above and below by functions of the spectrum flatness. The bounds become large when the spectrum is far from flat, which implies the necessity of a not-far-from- flat spectrum in nonparametric identification.

Several identification algorithms are proposed in Chapter 5. It is shown that the choice of the analytic center of the a posteriori uncertainty set as a nominal has a certain robustness property with respect to the inaccuracies in the a priori information.

In Chapter 6, we study the fast identification problem in the continuous-time case, which is different from its discrete-time counterpart. On a continuous-time interval, it is possible to collect an unlimited amount of sampled data, provided the sampling can be made arbitrarily fast. It remains unclear, however, whether arbitrarily accurate identification can be achieved on the basis of this large amount of data. In Section 6.3, it is shown, that in the noise free case, one can identify a stable continuous-time LTI system exactly on an arbitrarily short time interval, provided the entire segment of the output on the interval is available and the input is chosen properly. Here, the only a priori information is that the system is BIBO stable. No structural information or quantitative information about the system is required. A logarithmic integral condition on the inputs involved is obtained via quasianalyticity theory.

In such a case, however, accurate identification becomes impossible when the measurements are even slightly corrupted by noise. Similarly, the inherent error can be large if only samples of the output on a interval are available. An example is given where the inherent error is the same as the a priori uncertainty no matter how fast the sampling. Nevertheless, it is shown in Section 6.4 that, if the system is known to be in a compact set, (in either the $\mathbf{H}^\infty$ norm or $\mathbf{L}^1$ norm,) then the inherent error can be made arbitrarily small provided there are enough sampling points in an interval in which the noise-to-signal ratio is small enough.

With the above results in mind, one can ask: is identification speed still restricted? Is the metric complexity still a factor limiting identification speed? The answer to these questions is affirmative. It is shown in Section 6.5 that for a fixed noise level, even if the sampling rate is infinitely high, there is an irreducible uncertainty whenever the a priori uncertainty set contains a smooth subset (e.g. a set of low-pass functions,) of positive Gel'fand n-width. The higher the metric complexity, the slower the identification. Finally, the irreducible identification error is obtained for a set of approximately band-limited and time-limited systems in an example.

In Chapter 7, we study the problem of representation and estimation of continuous-time systems by sampling. One of the key questions we consider there is: once a model set and estimate are obtained for the sampled data systems, how should their continuous-time counterparts be constructed? A causal procedure is given for the construction of a continuous-time model set and estimate from the discrete ones. Representation and estimation errors are given in the $\mathbf{L}_1$ norm which is an upper bound on the $\mathbf{H}^\infty$ norm.

In Chapter 8, the results on fast identification are extended to obtain uncertainty principles for the identification of slowly varying systems. Slowly varying linear system are of interest in adaptive control because from a certain point of view they are the most general ones for which an input-output theory is useful. In particular, identification of uncertain elements has predictive value only if their future behavior is like their past or, at worst, approximately like their past. However, if a "black-box" system changes substantially in relation to the length of time needed to identify it, then accurate identification is inherently impossible. This fact is expressed through uncertainty principles, which relate the inherent uncertainty to the n-widths mentioned above.

## 1.3 Literature Review

System identification is a well established area of research. Indeed, there is a large body of literature on system identification. Many impressive results have been obtained and effective identification algorithms have been developed. There are several excellent books on this subject which describe the accomplishments during the last decades. See, for example, the books by Caines [5] and Ljung [26].

System identification has been traditionally formulated as a parameter estimation problem in a stochastic setting. The quality of an estimated model is given in terms of estimated standard deviations for the parameters or alternatively, confidence intervals for them. These parametric error estimations can be obtained even in the case where the true system model is not included in the identification model set by using techniques such as prediction error ( see Chapter 5, 6, and 8 of Cainse [5] for details).

In the robust control theory developed in the past fifteen years, however, the starting point for control system analysis and design is a nominal plant model and an operator norm bound on the model uncertainty, which is different from what is given by classical identification. This has fueled an renewed interest in identification in the control community, aiming at developing a theory of identification that is compatible with robust control. Its objective is to find system identification techniques which provide guaranteed error bounds in the operator norm in addition to a nominal system model. Such identification schemes have been labeled as worst-case deterministic approaches.

Several early papers on this subject appeared in the late 80's and early 90's. Inspired by a plenary lecture by Zames [56], Helmicki et al [17, 18, 19] derived robustly convergent algorithms which give estimates of the system with error bounds in the $\mathbf{H}^{\infty}$ norm from a set of corrupted frequency response measurements. These algorithms are related the work of Parker and Bitmead [37]. A study of asymptotic identification in the $\ell^1$ norm was given by Tse et al in [45]. Gu et al [15] and Makila [28] approached this problem from the system approximation point of view. Since then, the problem has attracted a lot of attention, e.g., [22, 47, 14, 38] and the references therein. Some of these results are closely related to those on set membership identification, a well studied subject, where identification is formulated as a parameter estimation problem under the assumption that the observations are corrupted by unknown but bounded noise. Several survey papers are available on this subject, e.g., [32, 36, 8].

Other efforts in developing an identification theory for robust control also appeared in the last few years, notably, the model validation approach by Smith and Doyle [44] and Poolla et al [40], the stochastic embedding approach by Goodwin et al [12], and the iterative estimation and control approach by Zang et al [61] and Schrama and Van den Hof [43].

The complexity issue of worst-case identification, the main topic of this thesis, was first posed by Zames using metric complexity theory [53, 54, 55, 56]. Based on the observation that both feedback and identification can be used as agents in uncertainty reduction, he pointed out that a unified theory for feedback and identification and a theory of information based adaptive control can be developed on the basis of metric complexity theory. It is this point of view led to the recent work of Zames and Wang [60, 51], and their joint work with the author [58, 24, 25, 23, 48, 50, 49]. The metric complexity of some data sets are also studied in [59, 29] in the context of control. The books by Vitushkin [46] and Pinkus [39] are good references on metric complexity theory.

The time complexity of worst-case identification was also studied by Poolla and Tikku [41] and Kacewicz and Milanese [20]. They showed that in the case where the observations are subject to unknown but bounded noise, the number of samples needed to identify a system of impulse response of length $n$ is of the order $2^n$. Tse et al [45] studied the limitation of worst-case identification in the asymptotic case.

The model selection problem in the stochastic setting have been studied by many people. References can be found in the book by Rissanen [42].

# Chapter 2

# Fast Identification and Metric Complexity

As shown in Section 1.1, the fast identification problem can be formulated as an input design and model selection problem, with the worst-case estimation error $e^T(u, \mathbf{X}_n)$ as the design criterion. In this Chapter the problem is formulated in a similar way in a general setting. With such a formulation in hand, the first question one may ask is: how should this rather complicated input and model design problem be solved? Right after the formulation, we answer the question by reducing the design problem to one of several standard optimization problems in the metric complexity theory. First, we decouple the input design and model selection by splitting the worst-case estimation error $e^T(u, \mathbf{X}_n)$ into two parts, the *inherent* and *representation* errors, which depend on the input and the model set respectively. Next, the optimal inherent error over all bounded inputs is introduced as a notion of n-width, called the time n-width, which is similar to the Gel'fand n-width; and the optimal representation error over all n-parameter affine models is related to the Kolmogorov n-width. Then, the optimal worst-case estimation error is related to the maximum of the time and Kolmogorov n-widths of the a priori data set. The establishment of the link between the optimal worst-case estimation error and the n-widths not only reduces the fast identification problem to the estimation of the n-widths, but also shows that it is the metric complexity of the data sets that limits the optimal identification speed. These results are illustrated by several examples, in which the optimal worst-case estimation error equals the maximum of the Gel'fand and Kolmogorov n-widths.

The material in this chapter has been published in [24, 58, 23].

## 2.1 Fast Identification in Operator-Normed Spaces

We will consider discrete-time systems represented by convolution operators of the form $\mathbf{K}: \mathbf{U} \to \mathbf{Y}$,

$$y(t) = \sum_{\tau=0}^{\infty} k(\tau)u(t-\tau), \quad t \in \mathbb{Z}, \tag{2.1}$$

where $k(\cdot) \in \mathbf{L}$, under the assumptions that:

(i) $\mathbf{U}$, $\mathbf{Y}$ are normed linear spaces of functions $\mathbb{Z} \to \mathbb{R}$ representing inputs and outputs respectively. We assume that the sets $\mathbf{U}$ and $\mathbf{Y}$ are contained in $\ell^{\infty}(-\infty, \infty)$. Inputs start at $-\infty$ to allow situations in which the system is running before observations begin.

(ii) $\mathbf{L}$ is a normed linear space consisting of causal weighting functions $\mathbb{Z}_+ \to \mathbb{R}$ acting on input pasts. The set $\mathbf{L}$ is contained in $\ell^1[0, \infty)$, ensuring that (2.1) is well defined. The norm $\|\cdot\|_{\mathbf{L}}$ can be the $\ell^1$ norm of the weighting functions, or the $\mathbf{H}^{\infty}$ norm of their Fourier transforms. Since $\ell^1[0,\infty)$ is a subspace of $\ell^2[0,\infty)$, the $\ell^2$ norm is also well defined for this class of systems.

For fixed $u \in \mathbf{U}$, the map

$$\mathbf{\Phi}_u : \mathbf{L} \to \mathbf{Y}, \ \ \mathbf{\Phi}_u(k) := \mathbf{K}u = y$$

is a linear map from kernels to outputs.

In general, the observations will be contaminated by noise, i.e.,

$$y(t) = \sum_{\tau=0}^{\infty} k(\tau)u(t-\tau) + v(t), \quad t \in \mathbb{Z}, \tag{2.2}$$

where $v \in \mathbf{Y}$ is the measurement noise. Sometimes relation (2.2) will be written in a more compact form as

$$y = \mathbf{\Phi}_u(k) + v. \tag{2.3}$$

We are given the a priori information that the true kernel lies in a set $\mathbf{S}_a \subset \mathbf{L}$ and the noise lies in a set $\mathbf{V} \subset \mathbf{Y}$. $\mathbf{S}_a$ and $\mathbf{V}$ will be assumed to satisfy the following:

**Assumption 1** $\mathbf{S}_a$ *and* $\mathbf{V}$ *are convex symmetric (i.e.,* $k \in \mathbf{S}_a \Rightarrow -k \in \mathbf{S}_a$ *and* $v \in \mathbf{V} \Rightarrow -v \in \mathbf{V}$*) subsets of* $\mathbf{L}$ *and* $\mathbf{Y}$ *respectively.*

Unlike in the case of parametric system identification, we do not assume any a priori knowledge on the system structure. $\mathbf{S}_a$ is in general a set containing infinite dimensional systems. Since the structure of the true system is not known, the accuracy of an estimate will be measured by its distance from the true system in the $\mathbf{L}$ norm. In fact, this is exactly the right measure to use if the estimate is to be used for robust controller design. (See comments in Section 1.3.)

**The objective of identification is to estimate a system in $\mathbf{S}_a$ from the noise corrupted output observations on a finite length interval.**

Given an input $u \in \mathbf{U}$, on the basis of the observations, $y(t_0), y(t_0+1), \ldots, y(t_0+T-1)$, the location of the true kernel $k_{true}$ is narrowed down from the a priori data set $\mathbf{S}_a$ to a smaller set,

$$\begin{aligned} \mathbf{S}(y) := \{k \in \mathbf{S}_a : \quad & \textstyle\sum_{\tau=0}^{\infty} k(\tau)u(t-\tau) = y(t) + v(t) \\ & \forall t \in [t_0, t_0+T) \ \mathit{for\ some}\ v \in \mathbf{V}\}, \end{aligned} \tag{2.4}$$

(which depends on $y$, $\mathbf{V}$ and $\mathbf{S_a}$,) or in a more compact form

$$\mathbf{S}(y) = \left\{k \in \mathbf{S}_a : \ \mathbf{P}_{[t_0,t_0+T)}\left(\mathbf{\Phi}_u(k) - y\right) = \mathbf{P}_{[t_0,t_0+T)}(v) \ \mathit{for\ some}\ v \in \mathbf{V}\right\} \tag{2.5}$$

where $\mathbf{P}_{[t_0,t_0+T)}$ denotes the truncation operator on $\ell^\infty$, defined by $\left(\mathbf{P}_{[n,m]}f\right)(t) := f(t)$ for $t \in [n,m]$, and 0 otherwise. $\mathbf{S}(y)$ represents the a posteriori information about the true system.

For robust controller design, it is desirable to represent the true system by a nominal system (i.e., an estimate) and describe the uncertainty by the distance between the nominal and the true system. Since all the systems in $\mathbf{S}(y)$ are consistent with the a priori information and the observations, any system in $\mathbf{S}(y)$ could be the true system. Hence the worst-case error between an estimate and the true system in the $\mathbf{L}$ norm is

$$e_{y,k_{est}}(u) := \sup_{k\in\mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}}. \tag{2.6}$$

The true system can be represented by the estimate $k_{est}$ and a ball centered at $k_{est}$ with radius $e_{y,k_{est}}(u)$. We call this a *nominal-ball type representation.*

To obtain the most accurate nominal-ball type representation of the system with the available information, it has been suggested (e.g. in [43, 6]) that the nominal system should be chosen to minimize the above worst-case error. For such an optimally chosen estimated kernel, $k_{est} \in \mathbf{L}$, the minimum error is

$$e_y(u) := \inf_{k_{est}\in\mathbf{L}} \sup_{k\in\mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}}. \tag{2.7}$$

Although an algorithm based on choice would give the most accurate nominal-ball type representation of the available information, it is in general impossible to implement, as the set of possible nominals is infinite dimensional. An infinite number of parameters might have to be identified to determine the nominal system or to represent it. In some special cases as shown in [6], it might be possible to implement such an algorithm by exploiting certain special properties of the a priori data set, but the computational complexity of these algorithms increases combinatorically with the amount of data, and the representation complexity also increases with the amount of data.

To achieve fast identification, we minimize the identification error (2.6) by choosing an estimate in a finite parameter model set, such as a finite dimensional subspace of $\mathbf{L}$, an ARMA model or a state space model. Then, an algorithm only needs to identify and store or print out a finite number of parameters.

**Definition 2.1** *A subset $\mathbf{X}_n$ of $\mathbf{L}$ is called a n-parameter model set if it is in the range of a mapping from $\mathbb{R}^n$ to the set of real sequences.*

If a $n$-parameter set $\mathbf{X}_n$ is chosen to be the model set, the minimum identification error becomes

$$e_y^T(u, \mathbf{X}_n) := \inf_{k_{est} \in \mathbf{X}_n} \sup_{k \in \mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}}. \tag{2.8}$$

This minimum error depends on the actual measurement $y$, which in turn depends on the true system and the true disturbance, as $y = \mathbf{K}_{true}u + v_{true}$ which, however, are not known beforehand. To study the effects of the input and the model set on the identification, we consider *the worst-case identification error*

$$e^T(u, \mathbf{X}_n) := \sup_{k_{true} \in \mathbf{S}_a} \sup_{v_{true} \in \mathbf{V}} \inf_{k_{est} \in \mathbf{X}_n} \sup_{k \in \mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}}, \tag{2.9}$$

as a function of the input and the model set.

Several algorithms for worst-case identification under an operator norm have been developed in the past few years [19, 43, 14, 28]. It is related to the usual set-membership identification [32, 36, 8] in the sense that uncertainty is described by sets. What is new and important in our formulation is that the finite parameter set in which the nominal is chosen is itself a design variable.

For a fixed input, the worst-case identification error $e^T(u, \mathbf{X}_n)$ also depends on the location of the observation interval $t_0$. If the probing capability of the input is not persistent, one may obtain more information on one interval than another of the same length. In the case where the identified model is to be continuously updated on the basis of data from the recent past, as in adaptive control [60, 51], we consider the worst-case of the identification error over all shifts of the observation interval relative to the input:

$$\bar{e}^T(u, \mathbf{X}_n) := \sup_{t_0 \in \mathbf{Z}} e^T(u, \mathbf{X}_n), \tag{2.10}$$

which is a shift-invariant quantity. We call it *the shift-invariant worst-case identification error*. Obviously, $\bar{e}^T \geq e^T$.

$e^T(u, \mathbf{X}_n)$ and $\bar{e}^T(u, \mathbf{X}_n)$ are the two key quantities we will study in this thesis. They are the identification errors given by the input and model set pair $(u, \mathbf{X}_n)$, when the observations are constrained on a interval of length $T$. They represent the speed of an identification procedure. Clearly, to achieve fast identification, the model set and input have to be designed properly to minimize these errors.

## 2.2 Separation of Input Design and Model Selection

To separate input design from model selection, we split the worst-case identification error $e^T(u, \mathbf{X}_n)$ into two parts. One part depends on the input. For this we introduce the notion of *inherent error* (which depends on $u$, $\mathbf{S}_a$ and $\mathbf{V}$)

$$\begin{aligned} \delta^T(u) &:= \sup\left\{\|k\|_{\mathbf{L}} :\ k \in \mathbf{S}_a,\ \mathbf{\Phi}_u(k)(t) = v(t)\ \forall t \in [t_0, t_0 + T),\ v \in \mathbf{V}\right\}, && (2.11)\\ &= \|\mathbf{S}(0)\|_{\mathbf{L}}. && (2.12) \end{aligned}$$

The second part will be called the *representation error*, which is defined as

$$dist(\mathbf{S}_a, \mathbf{X}_n) := \sup_{k \in \mathbf{S}_a} \inf_{g \in \mathbf{X}_n} \|k - g\|_{\mathbf{L}}. \tag{2.13}$$

It depends on the model set only.

The next proposition gives upper and lower bounds on $e^T(u, \mathbf{X}_n)$ in terms of inherent and representation errors, with the lower bound being greater than one third of the upper bound.

**Proposition 2.1** *Under Assumption 1,*

$$\max\left\{\delta^T(u), dist(\mathbf{S}_n, \mathbf{X}_n)\right\} \leq e^T(u, \mathbf{X}_n) \leq 3\max\left\{\delta^T(u), dist(\mathbf{S}_n, \mathbf{X}_n)\right\} \qquad (2.14)$$

**Remark** The worst-case identification error $e$ can be decomposed into two terms, i.e., inherent error and representation error. The inherent error is generated in the information collecting stage of identification, due to lack of data and inaccurate measurements; it is irreducible no matter what identification algorithm is used in the second stage. The representation error is due to inaccurate representation of the a priori uncertainty set; it represents the loss of information in the information processing stage.

If a model set $\mathbf{X}_{n,0}$ and an input $u_0$ are chosen to minimize the representation error and the inherent error respectively, then the worst-case identification error $e^T(u_0, \mathbf{X}_{n,0})$ is within a factor of three of the optimal one. It will be shown later that in many cases such a model set and input pair are also optimal for minimizing the worst-case identification error $e$. This implies that model selection and input design can be done independently. Moreover, the model selection is not affected by the experiment conditions, e.g., noise, observation interval, etc. Also input design is independent of the model set. Nevertheless, the optimal model set and the optimal input, or the minimum representation and inherent errors obtained from these two independent procedures, are related as they are determined by the a priori set. This will be shown in Section 2.6 and Chapter 3. When an optimal model set is chosen, the input design can be done for the optimal model set with the unmodeled dynamics in consideration. This will be discussed in detail in Chapter 4.

To prove the proposition we will need the following lemma which can be shown using similar devices as in [43] (see Appendix A).

**Lemma 2.1** *Under the Assumption 1,*

$$\inf_{k_{est}\in \mathbf{L}} \sup_{k\in \mathbf{S}(0)} \|k - k_{est}\|_{\mathbf{L}} \geq \|\mathbf{S}(0)\|_{\mathbf{L}} \tag{2.15}$$

$$\sup_{k_{true}\in \mathbf{S}_a} \sup_{v_{true}\in \mathbf{V}} \sup_{k_1,k_2\in \mathbf{S}(y)} \|k_1 - k_2\|_{\mathbf{L}} \leq 2\|\mathbf{S}(0)\|_{\mathbf{L}}. \tag{2.16}$$

**Proof of Proposition 2.1** Since $0 \in \mathbf{S}_a$ and $0 \in \mathbf{V}$, by setting $k_{true} = 0$ and $v_{true} = 0$, we get

$$\begin{aligned} e^T(u, \mathbf{X}_n) &\geq \inf_{k_{est}\in \mathbf{X}_n} \sup_{k\in \mathbf{S}(0)} \|k - k_{est}\|_{\mathbf{L}}, && (2.17)\\ &\geq \inf_{k_{est}\in \mathbf{L}} \sup_{k\in \mathbf{S}(0)} \|k - k_{est}\|_{\mathbf{L}} && (2.18)\\ &\geq \delta^T(u). \quad \text{(by Lemma 2.1)} && (2.19) \end{aligned}$$

On the other hand, since $k_{true} \in \mathbf{S}(y)$, by (2.9),

$$e^T(u, \mathbf{X}_n) \geq \sup_{k_{true}\in \mathbf{S}_a} \inf_{k_{est}\in \mathbf{X}_n} \|k_{true} - k_{est}\|_{\mathbf{L}} = dist(\mathbf{S}_a, \mathbf{X}_n). \tag{2.20}$$

Combining (2.19) and (2.20), we get the lower bound in (2.14).

To show the upper bound, assume $\hat{k} \in \mathbf{S}(y)$. Since $\hat{k} \in \mathbf{S}_a$, by definition of $dist(\mathbf{S}_a, \mathbf{X}_n)$, $\forall \epsilon > 0$, $\exists \hat{k}_{est} \in \mathbf{X}_n$, such that

$$\|\hat{k} - \hat{k}_{est}\|_{\mathbf{L}} \leq dist(\mathbf{S}_a, \mathbf{X}_n) + \epsilon. \tag{2.21}$$

It follows that

$$\begin{aligned} \inf_{k_{est}\in \mathbf{X}_n} \sup_{k\in \mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}} &\leq \sup_{k\in \mathbf{S}(y)} \|k - \hat{k}_{est}\|_{\mathbf{L}},\\ &\leq \sup_{k\in \mathbf{S}(y)} \left(\|k - \hat{k}\|_{\mathbf{L}} + \|\hat{k} - \hat{k}_{est}\|_{\mathbf{L}}\right),\\ &\leq \sup_{k,\hat{k}\in \mathbf{S}(y)} \|k - \hat{k}\|_{\mathbf{L}} + dist(\mathbf{S}_a, \mathbf{X}_n) + \epsilon. \end{aligned}$$

Since the above inequalities hold for all $\epsilon > 0$, we have

$$\inf_{k_{est} \in \mathbf{X}_n} \sup_{k \in \mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}} \leq \sup_{k, \hat{k} \in \mathbf{S}(y)} \|k - \hat{k}\|_{\mathbf{L}} + dist(\mathbf{S}_a, \mathbf{X}_n). \tag{2.22}$$

Therefore, by (2.9),

$$e^T(u, \mathbf{X}_n) \leq \sup_{k_{true} \in \mathbf{S}_a} \sup_{v_{true} \in \mathbf{V}} \sup_{k, \hat{k} \in \mathbf{S}(y)} \|k - \hat{k}\|_{\mathbf{L}} + dist(\mathbf{S}_a, \mathbf{X}_n). \tag{2.23}$$

By Lemma 2.1,

$$\sup_{k_{true} \in \mathbf{S}_a} \sup_{v_{true} \in \mathbf{V}} \sup_{k, \hat{k} \in \mathbf{S}(y)} \|k - \hat{k}\|_{\mathbf{L}} \leq 2\delta^T(u). \tag{2.24}$$

It is easy to verify that

$$2\delta^T(u) + dist(\mathbf{S}_a, \mathbf{X}_n) \leq 3 \max\left\{\delta^T(u), dist(\mathbf{S}_a, \mathbf{X}_n)\right\}. \tag{2.25}$$

□

Similarly, the shift-invariant worst-case identification error $\bar{e}^T(u, \mathbf{X}_n)$ can be expressed in terms of *shift-invariant inherent error*

$$\bar{\delta}^T(u) := \sup_{t_0 \in \mathbb{Z}} \delta^T(u). \tag{2.26}$$

and the representation error.

**Corollary 2.1** *Under Assumption 1,*

$$\max\left\{\bar{\delta}^T(u), dist(\mathbf{S}_a, \mathbf{X}_n)\right\} \leq \bar{e}^T(u, \mathbf{X}_n) \leq 3 \max\left\{\bar{\delta}^T(u), dist(\mathbf{S}_a, \mathbf{X}_n)\right\} \tag{2.27}$$

**Proof** Take the supremum of the quantities in (2.14) over all $t_0 \in \mathbb{Z}$.

□

Between the two sources of the inherent error, namely, lack of data and measurement noise, it is the former that usually puts the more severe constraint on fast identification. The measurement noise can be overcome by increasing the power of the input, a measure feasible on a short-time interval. On the other hand collection of more data can only be done by prolonging the observation interval, i.e., slowing down the identification. To isolate the effects of lack of data, we will concentrate on the optimal input design problem in the noise-free case. In the case where the observations are corrupted by bounded but unknown noise, bounds on the minimum inherent error have been recently obtained for certain data sets by Poolla and Tikku [40] and Kacewicz and Milanese [20]. These bounds are not accurate in the noise free case. Therefore, better bounds can be derived for the noisy case by combining the bounds in [40, 20] and the results in the thesis.

In the noise free case, the inherent error and the shift-invariant inherent error will be denoted by $\delta_0^T$ and $\bar{\delta}_0^T$ respectively, i.e.,

$$\begin{aligned} \delta_0^T(u) &:= \sup\left\{\|k\|_{\mathbf{L}} : \; k \in \mathbf{S}_a \; and \; \Phi_u(k)(t) = 0 \; \forall t \in [t_0, t_0 + T)\right\} && (2.28) \\ &= \sup\left\{\|k\|_{\mathbf{L}} : \; k \in \mathbf{S}_a \bigcap Null\left(\mathbf{P}_{[t_0, t_0+T)}\Phi_u\right)\right\}. && (2.29) \end{aligned}$$

They are equivalent to the limiting case inherent errors when the signal to noise ratio tends to infinity.

## 2.3 Gel'fand n-Width and Time n-Width

In the noise-free case, the inherent error optimized over all inputs depends on the length of the interval, and corresponds to a notion of n-width which we introduce here as follows.

**Definition 2.2** *For any $n \in \mathbb{Z}_+$ and arbitrary $t_0 \in \mathbb{Z}$, the time n-width is defined as*

$$\theta^n(\mathbf{S}_a, \mathbf{L}) := \begin{cases} \inf_{u \in \mathbf{U}} \delta_0^n(u) & n > 0 \\ \|\mathbf{S}_a\|_{\mathbf{L}}, & n = 0 \end{cases} \qquad (2.30)$$

*An optimal input is the one for which the infimum in (2.30) is attained, i.e.,* $\delta_0^n(u_{opt}) = \theta^n(\mathbf{S}_a, \mathbf{L})$.

(Since the input is optimizes over a subspace, its magnitude may be unbounded. $\theta^n$ is equivalent to the optimal inherent error with bounded noise.)

In the definition of *time n-width* $\theta^n$, the output observation interval $[t_0, t_0 + n)$ is viewed as being fixed, and the input as being optimized for that interval. The input is allowed to start prior to the observation interval for the following three reasons

(i) In on-line identification of a time-invariant system, the input past prior to the start of the identification is usually unknown. In this case, the output can be split into two parts. The first part is the free response of the system caused by the initial condition at $t_0$. The second is the output produced by the input after $t_0$. Since the initial condition is not known, the free response affects the observation as a disturbance, which can be quite large at the beginning of the identification. However, if the impulse response of the system decays with time, this disturbance eventually becomes small. In this case, there is the freedom to design the input prior to the start of the observations of (the second part of) the output.

(ii) In adaptive control, there is frequently the option of adding deterministic components, such as almost periodic functions, to the input to facilitate identification. Observations of such components on a time interval $[t_0, t_0 + T)$ completely determine their past prior to $t_0$, which can be considered as "known" for the purpose of the experiment. The question arises as to whether an appropriate choice of such a prior excitation can improve identification speed. To resolve this question, we allow the inputs to start prior to the observation interval.

(iii) Inputs which start prior to the observations will play a role in "moving window" adaptive control of time varying systems (see Chapter 8 for details).

Actually, in on line identification the input is seldom completely free to be optimized. The point of finding optimal inputs for fast identification is, rather, to provide lower bounds and an ideal against which actual input ensembles can be compared, and towards which they can eventually be modified, e.g., by the introduction of a dither signal.

The time n-width characterizes the time complexity of the data acquisition process. It is the best achievable inherent error with $n$ consecutive output observations. The inverse of the time n-width function gives the least time needed to reduce the inherent error to any specified level.

The optimal input for $\theta^n$ typically loses its optimality when shifted in relation to the observation interval. When the observation interval is not fixed in relation to the input, $\theta^n$ gives a lower bound which may be unattainable. In particular, in the adaptive control of slowly time-varying systems, the identified model is periodically updated on the basis of measurements from the recent past, and the model is then used to update the feedback law as in [58]. The observation interval lies in a "moving window" of constant length which advances in relation to the input, and a single input must therefore be effective for many intervals. For such cases, we introduce the second n-width, $\bar{\theta}^n$, which is the optimized shift-invariant inherent error. It provides a benchmark for the comparison of suboptimal input ensembles, whether free or fixed.

**Definition 2.3** *The shift invariant time n-width is defined as*

$$\bar{\theta}^n\left(\mathbf{S}_a, \mathbf{L}\right) := \begin{cases} \inf_{u \in \mathbf{U}} \bar{\delta}_0^n(u) & n > 0 \\ \|\mathbf{S}_a\|_{\mathbf{L}}, & n = 0 \end{cases} \tag{2.31}$$

Obviously, $\theta^n\left(\mathbf{S}_a, \mathbf{L}\right) \leq \bar{\theta}^n\left(\mathbf{S}_a, \mathbf{L}\right)$.

$\theta^n$ is the best achievable inherent error on an interval of length $n$ when the location of the interval is not fixed. It describes the global time-complexity of the data acquisition process.

The time n-width is bounded below by the Gel'fand n-width under mild conditions (see Chapter 3 for details). In many cases these two n-widths coalesce.

**Definition 2.4** *Let* $\mathbf{L}$ *be a normed linear space and* $\mathbf{S}_a$ *a subset of* $\mathbf{L}$. *The Gel'fand n-width of* $\mathbf{S}_a$ *in* $\mathbf{L}$ *is given by*

$$d^n(\mathbf{S}_a, \mathbf{L}) := \inf_{\mathbf{L}^n} \sup_{k \in \mathbf{S}_a \bigcap \mathbf{L}^n} \|k\|_{\mathbf{L}}, \tag{2.32}$$

*where the infimum is taken over all subspaces* $\mathbf{L}^n$ *of* $\mathbf{L}$ *of codimension* $n$. *A subspace is said to be of codimension* $n$ *if there exist* $n$ *independent bounded linear functionals* $f_1, \ldots, f_n$ *such that* $\mathbf{L}^n = \{k \in \mathbf{L} : f_i(k) = 0, \ i = 1, \ldots, n\}$. *If* $\mathbf{L}^n$ *is a subspace of codimension at most* $n$ *for which* $d^n(\mathbf{S}_a, \mathbf{L}) = \sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}_a \bigcap \mathbf{L}^n\}$, *then* $\mathbf{L}^n$ *is called an optimal subspace for the Gel'fand n-width* $d^n(\mathbf{S}_a, \mathbf{L})$.

The Gel'fand n-width can be seen as the optimized inherent error when identification is based on $n$ arbitrary linear measurements, whereas in the case of the n-width $\theta^n$ these measurements are restricted to be $n$ consecutive output values. The Gel'fand n-width characterizes the experimental complexity of an identification problem. The inverse of the Gel'fand n-width gives the least number of measurements needed to reduce the uncertainty to a predetermined value.

The properties of the time and Gel'fand n-widths and the relation between the two will be delineated in Chapter 3. Estimates of these n-widths for a class of a priori data sets will be given there.

## 2.4 Optimal Affine Representation and Kolmogorov n-Width

Generally speaking, model set optimization over all n-parameter models is a difficult problem. In this thesis, we restrict ourselves to affine models, i.e., finite dimensional subspaces of $\mathbf{L}$. Affine models, particularly Laguerre models, have been used in both identification and adaptive control. See, for example the papers by Belanger et al [60] and Gunnarsson and Wahlberg [16].

By the definition of representation error (2.13), the minimum representation error of $\mathbf{S}_a$ by a $n$-dimensional subspace is

$$d_n\left(\mathbf{S}_a, \mathbf{L}\right) := \inf_{\mathbf{X}_n \subset \mathbf{L}} dist(\mathbf{S}_a, \mathbf{X}_n). \tag{2.33}$$

This is exactly the Kolmogorov $n$-width of the a priori uncertainty set $\mathbf{S}_a$.

**Definition 2.5** *The n-width, in the sense of Kolmogorov, of* $\mathbf{S}_a$ *in* $\mathbf{L}$ *is given by*

$$d_n(\mathbf{S}_a, \mathbf{L}) := \inf_{\mathbf{X}_n} \sup_{k \in \mathbf{S}_a} \inf_{g \in \mathbf{X}_n} \|k - g\|_{\mathbf{L}}, \tag{2.34}$$

*where the infimum is taken over all n-dimensional subspaces of* $\mathbf{L}$*. If*

$$d_n(\mathbf{S}_a, \mathbf{L}) := \sup_{k \in \mathbf{S}_a} \inf_{g \in \mathbf{X}_n} \|k - g\|_{\mathbf{L}}$$

*for some subspace* $\mathbf{X}_n$ *of dimension at most n, then* $\mathbf{X}_n$ *is said to be an optimal subspace for* $d_n(\mathbf{S}_a, \mathbf{L})$.

The optimal subspace gives the optimal $n$-dimensional affine model for the uncertainty set $\mathbf{S}_a$.

The Kolmogorov n-width characterizes the representation complexity of an identification problem. The inverse function of $d_n$ was called the *metric dimension* function by Zames [52] and viewed as an appropriate measure of metric complexity of uncertainty sets in feedback systems. It is the dimension of the smallest subspace whose elements are capable of approximating arbitrary points of the a priori data set $\mathbf{S}_a$ to a specified tolerance.

Each one of the three notions of n-width, i.e., Kolmogorov, Gel'fand, and time n-widths, describes the complexity of a distinct aspect of an identification problem. None of them describes the complexity of an identification problem completely. However, it will be shown in Chapter 3 that in many special cases, they coincide and therefore can used interchangeably.

## 2.5 The Optimal Worst-Case Identification Error

Using Proposition 2.1, we can get upper and lower bounds of the optimal worst-case identification error in terms of time n-width and Kolmogorov n-width, with the bounds different from each other only by a factor of three.

**Proposition 2.2** *Under Assumption 1, the optimal noise-free worst-case identification error has the following lower and upper bounds,*

$$\max\left\{\theta^T(\mathbf{S}_a,\mathbf{L}), d_n(\mathbf{S}_a,\mathbf{L})\right\} \leq \inf_{u\in\mathbf{U}} \inf_{\mathbf{X}_n\subset\mathbf{L}} e^T(u,\mathbf{X}_n) \leq 3\max\left\{\theta^T(\mathbf{S}_a,\mathbf{L}), d_n(\mathbf{S}_a,\mathbf{L})\right\} \tag{2.35}$$

**Proof** Take the infimum of the quantities in (2.14) over all $u \in \mathbf{U}$ and $\mathbf{X}_n$ in $\mathbf{L}$.

□

If $d^T(\mathbf{S}_a,\mathbf{L}) \leq \theta^T(\mathbf{S}_a,\mathbf{L})$, then the optimal worst-case identification error is bounded below by $\max\left\{d^T(\mathbf{S}_a,\mathbf{L}), d_n(\mathbf{S}_a,\mathbf{L})\right\}$. It will be shown in Chapter 3 that if the a priori identification error set has a certain property of monotone decrease, then the maximum of the two n-widths is also an upper bound, i.e.,

$$\inf_{u\in\mathbf{U}}\ \inf_{\mathbf{X}_n\subset\mathbf{L}} e^T(u,\mathbf{X}_n) = \max\left\{d^T(\mathbf{S}_a,\mathbf{L}), d_n(\mathbf{S}_a,\mathbf{L})\right\}. \tag{2.36}$$

Similarly, the optimal shift-invariant worst-case identification error can be expressed in terms of the shift-invariant time n-width and Kolmogorov n-width as follows

$$\max\left\{\bar{\theta}^T(\mathbf{S}_a,\mathbf{L}), d_n(\mathbf{S}_a,\mathbf{L})\right\} \le \inf_{u\in\mathbf{U}}\ \inf_{\mathbf{X}_n\subset\mathbf{L}} \bar{e}^T(u,\mathbf{X}_n) \le 3\max\left\{\bar{\theta}^T(\mathbf{S}_a,\mathbf{L}), d_n(\mathbf{S}_a,\mathbf{L})\right\} \tag{2.37}$$

Proposition 2.2 establishes the relation between identification speed and metric complexity of the a priori data set measured in the Gel'fand, Kolmogorov, and time n-widths. It implies that the length of time needed to identify a system to a given tolerance is proportional to the metric complexity of the a priori data set. These n-widths have been computed by the author and others [57, 48, 29, 38] for certain special cases of the a priori data set using various ad hoc methods. These results are summarized in the following examples. In Chapter 3 they will be derived from a single monotonicity principle, which will lead to exact estimates for some new data sets.

## 2.6 Several Examples

**Example 2.1** Let $\mathbf{U} = \mathbf{Y} = \ell^\infty$, $n > 0$, $C > 0$, and $0 < r < 1$. It can be shown that

(i) if

$$\mathbf{S}_{a_1} := \left\{k \in \ell^1[0,\infty): \ \ |k(\tau)| \le C r^\tau, \ \forall \tau \in \mathbb{Z}_+\right\}, \tag{2.38}$$

then

$$d_n\left(\mathbf{S}_{a_1},\ell^1\right) = d^n\left(\mathbf{S}_{a_1},\ell^1\right) = \theta^n\left(\mathbf{S}_{a_1},\ell^1\right) = \frac{C}{1-r} r^n; \tag{2.39}$$

(ii) if

$$\mathbf{S}_{a_2} := \left\{ k \in \ell^1 : \ \sum_{\tau=0}^{\infty} |k(\tau)| r^{-\tau} \leq C \right\}, \tag{2.40}$$

then

$$d_n\left(\mathbf{S}_{a_2}, \ell^1\right) = d^n\left(\mathbf{S}_{a_2}, \ell^1\right) = \theta^n\left(\mathbf{S}_{a_2}, \ell^1\right) = Cr^n; \tag{2.41}$$

In (i) and (ii) $\mathbf{L} = \ell^1[0,\infty)$. In the next two cases $\mathbf{L}$ is the Wiener algebra, i.e., the $\mathbf{H}^\infty$-normed algebra of functions in $\mathbf{H}^\infty(D)$ with Fourier coefficients (restricted) in $\ell^1[0,\infty)$.

(iii) if

$$\mathbf{S}_{a_3} = \left\{ H(z) \in \mathbf{H}^\infty\left(D_{r^{-1}}\right) : \ \|H\|_{\infty, r^{-1}} \leq C \right\}, \tag{2.42}$$

then

$$d_n\left(\mathbf{S}_{a_3}, \mathbf{L}\right) = d^n\left(\mathbf{S}_{a_3}, \mathbf{L}\right) = \theta^n\left(\mathbf{S}_{a_3}, \mathbf{L}\right) = Cr^n; \tag{2.43}$$

(iv) if

$$\mathbf{S}_{a_4} = \left\{ H(z) \in \mathbf{H}^\infty(D) : \ \|H'\|_\infty \leq C \right\}, \tag{2.44}$$

where $H'(z)$ denotes the derivative of $H$, then

$$d_n\left(\mathbf{S}_{a_4}, \mathbf{L}\right) = d^n\left(\mathbf{S}_{a_4}, \mathbf{L}\right) = \theta^n\left(\mathbf{S}_{a_4}, \mathbf{L}\right) = C/n. \tag{2.45}$$

In each of these examples, the optimal affine model is the FIR model $\mathbf{L}_n := sp\{1, z, \ldots, z^{n-1}\}$, the[1] optimal subspace of codimension $n$ for the Gel'fand n-widths is $\mathbf{L}^n = \{k \in \mathbf{L} : \ k(\tau) = 0 \ \forall \tau \in [0, n)\}$, and the optimal input is a unit impulse applied at the start of the observation interval. Moreover, this FIR model and the impulse input form an optimal model-set-input pair for minimizing the worst-case identification error $e^T(u, \mathbf{X}_n)$, and for each $\mathbf{S}_{a_i}$, $i = 1, \ldots, 4$.

$$\inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \in \mathbf{L}} e^T(u, \mathbf{X}_n) = \max\left\{ d^T(\mathbf{S}_{a_i}, \mathbf{L}), d_n(\mathbf{S}_{a_i}, \mathbf{L}) \right\}. \tag{2.46}$$

[1] By the usual abuse of notation, $z^i$ denotes the $i$-th power function in $\mathbf{H}^\infty$.

**Example 2.2** Under the assumptions of Example 2.1,

$$\theta^T(\mathbf{S}_{a_i},\mathbf{L}) \le \theta^T(\mathbf{S}_{a_i},\mathbf{L}) \le \alpha_i \theta^T(\mathbf{S}_{a_i},\mathbf{L}), \tag{2.47}$$

where $\alpha_1 = \alpha_2 = \alpha_3 = 2$ and $\alpha_4 = \pi$. For each $\mathbf{S}_{a_i}$, $i = 1, \ldots, 4$,

$$\begin{aligned} \max\left\{d^T(\mathbf{S}_{a_i},\mathbf{L}), d_n(\mathbf{S}_{a_i},\mathbf{L})\right\} &\le \inf_{u\in\mathbf{U}} \inf_{\mathbf{X}_n\in\mathbf{L}} \bar{e}^T(u,\mathbf{X}_n) & (2.48)\\ &\le 3\max\left\{\alpha_i d^T(\mathbf{S}_{a_i},\mathbf{L}), d_n(\mathbf{S}_{a_i},\mathbf{L})\right\}. & (2.49)\end{aligned}$$

In these examples, there is a loss of accuracy whenever there is no freedom to position the observation interval advantageously, but the loss never exceeds a factor of $\pi$.

# Chapter 3

# Estimation of Time, Gel'fand, and Kolmogorov n-Widths

As shown in Chapter 2, the resolution of the fast identification problem involves the estimation of the various n-widths of the a priori data set. All of the cases in Example 2.1 have certain properties in common, e.g., that the optimal input is an impulse at the start of the observation interval, that the n-widths equal the norm of a truncated impulse response, and that the impulse responses in the a priori data set drop off with time. It seems natural, therefore, to seek a common principle from which these n-widths could be derived, which is what we propose to do next. We will obtain a principle based on a property of monotone decrease of these data sets. For such data sets, the optimal worst-case estimation error equals the maximum of the n-widths in Gel'fand and Kolmogorov senses, the optimal input is an impulse at the start of the observation interval, and the optimal affine model is an FIR model. The claims as to the shift-invariant time n-width in Example 2.2 will also be proved.

The material in this chapter have been published in [24, 58, 23].

## 3.1 Generalization of Time n-Width

The n-width $\theta^n$ can be extended to more general classes of operators as follows. Let $\mathbf{E}$ be a normed linear space of the functions $f : \mathbb{Z} \to \mathbb{R}$ which is invariant under the bilateral shift $\mathbf{T} : \mathbf{E} \to \mathbf{E}$, $(\mathbf{T}f)(t) := f(t-1)$, and invariant under the time-reversal involution $f^*(t) := f(-t)$. Let $\mathbf{L}$ be any subspace of $(\mathbf{E})^*$, the dual space of $\mathbf{E}$, i.e., $\mathbf{L}$ consists of bounded linear functionals on $\mathbf{E}$. Suppose furthermore that $\mathbf{L}$ is a normed space of functions from $\mathbb{Z}_+$ to $\mathbb{R}$. If $\mathbf{U}$ is a subspace of $\mathbf{E}$, then for $\mathbf{S}_a \subset \mathbf{L}$ our previous definitions of the n-widths $\theta^n(\mathbf{S}_a, \mathbf{L})$ and $\bar{\theta}^n(\mathbf{S}_a, \mathbf{L})$ remain valid. It should be noted that the norm on $\mathbf{L}$ may be different from the norm on the dual of $\mathbf{E}$, and $\mathbf{U}$ can be any subspace of $\mathbf{E}$, e.g., $\mathbf{E} = \ell^2(-\infty,\infty)$, $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{H}^\infty}$, and $\mathbf{U} = sp\{u_1, u_2, \ldots, u_n\} \subset \ell^2(-\infty,\infty)$.

The next proposition gives a shift invariant property of $\theta^n$.

**Proposition 3.1** *If $\mathbf{U}$ is invariant under the bilateral shift and time-reversal involution, (i.e., $u \in \mathbf{U} \Rightarrow \mathbf{T}(u) \in \mathbf{U}$ and $(u)^* \in \mathbf{U}$) then $\theta^n$ is independent of the location of the observation interval, $t_0$.*

**Proof** To express the dependency of $\theta^n$ and $\delta_0^n$ on $t_0$ explicitly, we denote them by $\theta^n(t_0)$ and $\delta_0^n(u, t_0)$ respectively. It will be shown that $\theta^n(t_0) = \theta^n(0)\ \forall t_0 \in \mathbb{Z}$.

By definition of $\theta^n(t_0)$, $\forall \epsilon > 0$, $\exists u \in \mathbf{U}$ such that

$$\delta^n(u, t_0) \leq \theta^n(t_0) + \epsilon. \tag{3.1}$$

Since $\delta^n(u, t_0) = \|\mathbf{S}_a \bigcap Null\left(\mathbf{P}_{[t_0, t_0+n)}\mathbf{\Phi}_u\right)\|_{\mathbf{L}}$, and

$$\begin{aligned}
Null\left(\mathbf{P}_{[t_0,t_0+n)}\mathbf{\Phi}_u\right) &= \left\{k \in \mathbf{L} : \sum_{\tau=0}^{\infty} k(\tau)u(t-\tau) = 0\ \forall t \in [t_0, t_0+n)\right\}, \\
&= \left\{k \in \mathbf{L} : \sum_{\tau=0}^{\infty} k(\tau)u(t_0+t-\tau) = 0\ \forall t \in [0, n)\right\}, \\
&= Null\left(\mathbf{P}_{[0,n)}\mathbf{\Phi}_{\mathbf{T}^{-t_0}(u)}\right),
\end{aligned}$$

we have

$$\delta^n(u, t_0) = \delta^n(\mathbf{T}^{-t_0}(u), 0). \tag{3.2}$$

Since $\mathbf{U}$ is invariant under shifting and time reversal, $u \in \mathbf{U} \Rightarrow \mathbf{T}^{-t_0}(u) \in \mathbf{U}$. Therefore,

$$\begin{aligned} \theta^n(0) &\leq \delta^n(\mathbf{T}^{-t_0}(u), 0), \\ &= \delta^n(u, t_0) \\ &\leq \theta^n(t_0) + \epsilon \quad \text{(by (3.1))}. \end{aligned}$$

This implies that $\theta^n(0) \leq \theta^n(t_0)$ as $\epsilon$ is arbitrary. The proof is completed by showing $\theta^n(0) \geq \theta^n(t_0)$, using a similar argument.

□

It should be noted that, although $\theta^n$ is shift invariant under the conditions of the Proposition 3.1, it can not replace $\bar{\theta}^n$ as a measure of the time complexity when the time location of the observation is not fixed. The invariance of $\theta^n$ only means that if an input is optimal for one fixed interval, then when the interval is shifted, the optimal input for the shifted interval is the shifted input, which will stay in the set $\mathbf{U}$ of admissible inputs if $\mathbf{U}$ is invariant under shifting and time reversal.

## 3.2 Relation Between Time n-Width and Gel'fand n-Width

The n-width $\theta^n$ is related to the Gel'fand n-width, a standard notion in metric complexity theory, and with some restrictions $\theta^n$ is bounded below by the Gel'fand n-width as follows.

**Proposition 3.2** *If there exists $\lambda > 0$ such that $\|k\|_{\mathbf{L}} \geq \lambda \|k\|_{\mathbf{E}^*}$ for all $k \in \mathbf{S}_a$, then, for all $\mathbf{U} \subset \mathbf{E}$,*

$$d^n(\mathbf{S}_a, \mathbf{L}) \leq \theta^n(\mathbf{S}_a, \mathbf{L}) \tag{3.3}$$

**Proof** For any $u \in \mathbf{U}$, write $u_i = T^{-i}u^*$, which is in $\mathbf{E}$ under our assumption that $\mathbf{U} \subset \mathbf{E}$ and $\mathbf{E}$ is closed under $T$ and time-reversal $(\cdot)^*$. It will be shown later that under the hypotheses of the proposition, there exists a subspace $\mathbf{S} \subset \mathbf{L}$ containing the a priori uncertainty set $\mathbf{S}_a$ and with the property that, for each $u \in \mathbf{U}$, the sum $\sum_{t=0}^{\infty} k(t)u_i(t)$ defines a linear functional bounded in the $\mathbf{L}$ norm on $\mathbf{S}$. Now, let $L^n(u)$ be the space consisting of those $k \in \mathbf{S}$ which lie in the intersection of the null spaces of the functionals determined by the $u_i$, $i = 0, \ldots, n-1$. $L^n(u)$ is a subspace of codimension $n$ in $\mathbf{S}$. As $d^n$ is by definition an infimum over all spaces of codimension $n$,

$$\left\| \mathbf{S}_a \bigcap L^n(u) \right\|_{\mathbf{L}} \geq d^n(\mathbf{S}_a, \mathbf{S}), \tag{3.4}$$

Since (3.4) holds for all $u \in \mathbf{U}$, the infimum $\theta^n$ of the left side of (3.4) over $u$, satisfies

$$\theta^n(\mathbf{S}_a, \mathbf{L}) \geq d^n(\mathbf{S}_a, \mathbf{S}). \tag{3.5}$$

Now, using the fact that every bounded linear functional on $\mathbf{S}$ can be extended to a bounded linear functional on $\mathbf{L}$ with preservation of norm, (by the Hahn-Banach Theorem,) it is not hard to show that $d^n(\mathbf{S}_a, \mathbf{S}) = d^n(\mathbf{S}_a, \mathbf{L})$ [38]. The proposition follows.

It remains to show the existence of such a subspace $\mathbf{S}$. Put

$$\mathbf{S} := \{k \in \mathbf{L} : \ ck \in \mathbf{S}_a, \ \mathit{for\ some}\ c \in \mathbb{R}\}. \tag{3.6}$$

As $\mathbf{S}_a$ is a convex set which contains the origin, $\mathbf{S}$ is a subspace. Because $\mathbf{S} \subset \mathbf{L} \subset \mathbf{E}^*$, $u_i \in \mathbf{E}$ defines a linear functional on $\mathbf{S}$ bounded in the $\mathbf{E}^*$ norm. Since $\|k\|_{\mathbf{L}} \geq \lambda \|k\|_{\mathbf{E}^*}$ for all $k \in \mathbf{S}$ by hypothesis,

$$\sup_{k \in \mathbf{S}} \frac{|u_i(k)|}{\|k\|_{\mathbf{L}}} \leq \sup_{k \in \mathbf{S}} \frac{|u_i(k)|}{\lambda \|k\|_{\mathbf{E}^*}}. \tag{3.7}$$

Therefore, the functional on $\mathbf{S}$ defined by $u$ is also bounded in the $\mathbf{L}$ norm, and $\mathbf{S}$ has the properties claimed.

□

The following examples indicate that the condition in Proposition 3.2 is mild.

**Example 3.1** (i) If $\mathbf{E} = \ell^\infty(-\infty,\infty)$ and $\mathbf{L} = \ell^1[0,\infty)$, then the condition in Proposition 3.2 is satisfied for all convex and symmetric subsets of $\mathbf{L}$, as $\|\cdot\|_{\ell^1} \geq \|\cdot\|_{\ell^\infty}$.

(ii). If $\mathbf{S}_a$ is contained in a finite dimensional subspace of $\mathbf{L}$, then the condition in Proposition 3.2 is satisfied for all $\mathbf{E}$ and $\mathbf{L}$, as all the norms are equivalent on a finite dimensional space. One typical example of this is the case when $\mathbf{U} = \mathbf{E} = \ell^\infty(-\infty,\infty)$, $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{H}^\infty}$, and $\mathbf{S}_a = sp\{1, z, \ldots, z^n\}$.

## 3.3 A Set of Monotone Decreasing Systems

All the data sets introduced in Example 2.1 are monotone decreasing in a certain sense which will now be made more precise.

By the norm $\|\mathbf{S}\|_{\mathbf{L}}$ of any sebset $\mathbf{S}$ of $\mathbf{L}$ we mean $\|\mathbf{S}\|_{\mathbf{L}} := \sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}\}$. A subset $\mathbf{S}$ of $\mathbf{L}$ will be called *monotone decreasing* if given any fixed interval $[t_0, t_1)$, the norm of $\mathbf{S}$ intersected with any subspace of functions of $\mathbf{L}$ having support on a (variable) subinterval $[t_0', t_0' + i)$ of $[t_0, t_1)$ is monotone decreasing as $t_0'$ increases. A somewhat more general property than monotonicity of $\mathbf{S}$ requires the previous statement to be true only for subintervals of length $i \leq q$, in which case $\mathbf{S}$ will be called $q$-monotone decreasing. We will now define these notions of monotonicity more formally after introducing some notation.

For sets $\mathbf{S} \subset \mathbf{L}$, we shall wish to consider subsets of functions with support restricted to an interval, and introduce the

**Notation** *For any subset* $\mathbf{S}$ *of* $\mathbf{L}$, $\mathbf{S}|_{[t_1,t_2)}$ *denotes the subset of functions of* $\mathbf{S}$ *with support in the interval* $[t_1, t_2)$ *of* $\mathbb{Z}$, *i.e.,* $\mathbf{S}|_{[t_1,t_2)} := \mathbf{S} \bigcap \mathbf{P}_{[t_1,t_2)}\mathbf{S}$. *For a set of the form* $\mathbf{S}|_{[0,m)}$, *a p-section is its intersection with any p-dimensional subspace of* $\mathbf{L}|_{[0,m)}$. *A tail p-section of* $\mathbf{S}|_{[0,m)}$ *is the intersection of* $\mathbf{S}$ *with the*[1] *span* $sp\{z^{m-p}, \ldots, z^{m-1}\}$.

**Definition 3.1** $\mathbf{S}_a$ *will be called q-monotone decreasing,* $1 \leq q \leq \infty$, *if every p-section* $X_p$ *in* $\mathbf{S}_a|_{[0,m)}$ *satisfies*

$$\|X_p\|_{\mathbf{L}} \geq \|\mathbf{S}_a|_{[m-p,m)}\|_{\mathbf{L}}$$

*whenever* $1 \leq p < m < \infty$, $p \leq q \leq \infty$. *(Here "*$\mathbf{S}_1$ *is smaller than* $\mathbf{S}_2$*" means that* $\|\mathbf{S}_1\|_{\mathbf{L}} \leq \|\mathbf{S}_2\|_{\mathbf{L}}$.*)*

In other words, in any subset of the form $\mathbf{S}_a|_{[0,m)}$, the smallest $p$-section is the tail $p$-section and this is true for all $p$ up to some $q$.

## 3.4 Estimation of Gel'fand n-Width

In this section, we estimate the Gel'fand n-width of a $q$-monotone decreasing data set and give several corollaries which cover the results given in Example 2.1.

**Theorem 3.1** *If the a priori set* $\mathbf{S}_a$ *is q-monotone decreasing,* $(1 \leq q \leq \infty,)$ *then the Gel'fand n-width* $d^n$ *has bounds*

$$\|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} \leq d^n(\mathbf{S}_a, \mathbf{L}) \leq \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}, \quad (p \leq q,\ p < \infty.) \tag{3.8}$$

*Moreover, if*

$$\lim_{p \to q} \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}, \tag{3.9}$$

*then* $d^n(\mathbf{S}_a, \mathbf{L}) = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}$, *and the subspace* $L^n_{opt} = \{k : k(i) = 0,\ i = 0, 1, \ldots, n-1\}$ *is optimal for* $d^n$.

[1] By the usual abuse of notation, $z^i$ denotes the $i$-th power function in $\mathbf{H}^\infty$.

**Proof** To show the lower bound in (3.8), fix $p \leq q$, $p < \infty$. By definition of $d^n$,

$$d^n(\mathbf{S}_a, \mathbf{L}) = \inf_{L^n} \|\mathbf{S}_a \bigcap L^n\|_{\mathbf{L}} \tag{3.10}$$

$$\geq \inf_{L^n} \|\mathbf{S}_a|_{[0,n+p)} \bigcap L^n|_{[0,n+p)}\|_{\mathbf{L}} \tag{3.11}$$

It will be shown that for any $L^n$, $L^n|_{[0,n+p)}$ is a subspace of $\mathbf{L}|_{[0,n+p)}$ with dimension greater or equal to $p$. Therefore, by definition of $p$-section, the boundary of $\mathbf{S}_a|_{[0,n+p)} \bigcap L^n|_{[0,n+p)}$ is either a $p$-section or contains a $p$-section. It will follow, by the monotonicity of $\mathbf{S}_a$, that the last term in (3.11) is not less than $\|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}}$, giving the lower bound of (3.8) for all finite $p$, $p \leq q$.

To show that $\dim\left(L^n|_{[0,n+p)}\right) \geq p$, we notice that

$$L^n|_{[0,n+p)} = \left\{ k \in sp\left\{1, z, \ldots, z^{n+p-1}\right\} : \sum_{t=0}^{n+p-1} k(t) f_i(t) = 0, \ i = 1, \ldots, n \right\}, \tag{3.12}$$

where $f_i$'s are the functionals defining $L^n$. Put

$$F = \begin{pmatrix} f_1(0) & \ldots & f_1(n+p-1) \\ \vdots & \ldots & \vdots \\ f_n(0) & \ldots & f_n(n+p-1) \end{pmatrix}. \tag{3.13}$$

By (3.12), we have

$$L^n|_{[0,n+p)} = \left\{ k \in \mathbf{L} : \ k = \sum_{t=0}^{n+p-1} k(t) z^t, \ (k(0), \ldots, k(n+p-1))^T \in Null(F) \right\}. \tag{3.14}$$

It follows that $\dim\left(L^n|_{[0,n+p)}\right) = \dim(Null(F)) \geq p$.

The upper bound is achieved by taking $L^n = L^n_{opt}$. When the upper bound equals the lower bound in (3.8), the optimality of $L^n_{opt}$ follows from the definition of $d^n$.

□

The estimates of Gel'fand n-width described in Example 2.1 are established by the following corollaries.

**Corollary 3.1** *Let $f \in \ell^1[0,\infty)$ be a monotone decreasing positive function. If*

$$\mathbf{S}_a = \left\{k \in \ell^1[0,\infty) : \quad |k(\tau)| \le f(\tau), \ \forall \tau \in \mathbb{Z}_+\right\}, \tag{3.15}$$

*then*

$$d^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = \|\mathbf{P}_{[n,\infty)}(f)\|_{\ell^1}, \tag{3.16}$$

*and the subspace $L^n_{opt} = \{k : \ k(i) = 0, \ i = 0, 1, \ldots, n-1\}$ is optimal for $d^n$.*

**Proof** It will be shown that $\mathbf{S}_a$ is $\infty$-monotone decreasing. For this it is enough to show that for any positive integers $p < m < \infty$, which will be held fixed in the proof, if $\mathbf{M}$ is a $p$-dimensional subspace of $sp\{1, z, \ldots, z^{m-1}\}$, then the $p$-section $\mathbf{A} := \mathbf{M} \bigcap \mathbf{S}_a|_{[0,m)}$ of the set $\mathbf{S}_a|_{[0,m)}$ is not smaller than the tail $p$-section $\mathbf{S}_a|_{[m-p,m)}$, i.e., there is a function $k \in \mathbf{A}$ such that

$$\|k\|_{\ell^1} \ge \|\mathbf{S}_a|_{[m-p,m)}\|_{\ell^1} = \|\mathbf{P}_{[m-p,m)}(f)\|_{\ell^1}, \tag{3.17}$$

where $\|\cdot\|_{\ell^1}$ denotes the $\ell^1$ norm, and the last identity holds because $\mathbf{S}_a$ is closed under truncation $\mathbf{P}_{[m-p,m)}$. It will be shown that in fact there exists $k \in \mathbf{A}$ which touches the boundary of $\mathbf{A}$ at $p$ points at least, i.e., there exist $\tau_1, \tau_2, \ldots, \tau_p$ in the interval $[0, m)$ at which $|k(\tau_i)| = |f(\tau_i)|$, and $|k(\tau)| \le |f(\tau)|$ elsewhere in the interval $[0, m)$. Such a $k$ clearly has the requisite property (3.17) because $f$ is monotone decreasing. Its existence will be established by induction on the number of points touching the boundary.

Let $k_1$ be a non-zero vector in $\mathbf{M}$. Since $f$ is positive, there exists a constant $a \in \mathbb{R}$ such that $ak_1 \in \mathbf{A}$ and $ak_1$ touches the boundary of $\mathbf{A}$ at 1 point at least, say at $\tau_1 \in [0, m)$. Suppose, next, that for some integer $i$, $1 \le i < p$, there exists $k_i \in \mathbf{A}$ which touches that boundary of $\mathbf{A}$ at (least at) $i$ points, $\tau_1, \tau_2, \ldots, \tau_i$ in $[0, m)$. Let us show that there exist $k_{i+1} \in \mathbf{A}$ which touches the boundary at (least at) $i+1$ points, $\tau_1, \tau_2, \ldots, \tau_{i+1}$.

Let $\{v_1, \ldots, v_p\}$ be a basis of $\mathbf{M}$, and $\mathbf{V}_i : \mathbf{M} \to sp\{z^{\tau_1}, \ldots, z^{\tau_i}\}$ be the transformation whose matrix representation relative to this basis is

$$\begin{pmatrix} v_1(\tau_1) & \ldots & v_p(\tau_1) \\ \vdots & \ldots & \vdots \\ v_1(\tau_i) & \ldots & v_p(\tau_i) \end{pmatrix}. \tag{3.18}$$

For $i < p$, $Null(\mathbf{V}_i) \neq 0$. Let $\Delta k_i \neq 0$ be any function in $Null(\mathbf{V}_i)$. Then $\Delta k_i \in \mathbf{M}$ and $k_i + a\Delta k_i \in \mathbf{M}$ for all $a \in \mathbb{R}$. Since by this construction $\Delta k_i(\tau_j) = 0$, $j = 1, 2, \ldots, i$, $k_i + a\Delta k_i$ will stay on the boundary of $\mathbf{A}$ at $\tau_1, \tau_2, \ldots, \tau_i$ for all $a \in \mathbb{R}$. Since $\Delta k_i \neq 0$, $k_i + a\Delta k_i =: k_{i+1}$ must touch the boundary of $\mathbf{A}$ at some point $\tau_{i+1}$ in $[0, m)$ for some value $a \in \mathbb{R}$, i.e., $|k(\tau_{i+1})| = |f(\tau_{i+1})|$, and $\tau_{i+1} \neq \tau_1, \tau_2, \ldots, \tau_i$. Thus $k_{i+1}$ touches the boundary of $\mathbf{A}$ at (least at) $i+1$ points, and has the requisite properties.

Since

$$\lim_{p\to\infty} \|\mathbf{S}_a|_{[n,n+p)}\|_{\ell^1} = \lim_{p\to\infty} \|\mathbf{P}_{[n,n+p)}(f)\|_{\ell^1} = \|\mathbf{P}_{[n,\infty)}(f)\|_{\ell^1},$$

the identity (3.9) applies. Therefore, the theorem implies that $d^n = \|\mathbf{P}_{[n,\infty)}(f)\|_{\ell^1}$, and the subspace $L^n_{opt} = \{k : \ k(i) = 0, \ i = 0, 1, \ldots, n-1\}$ is optimal.

□

**Corollary 3.2** *Let $f \in \ell^1[0,\infty)$ be a monotone decreasing positive function. If*

$$\mathbf{S}_a = \left\{ k \in \ell^1[0,\infty) : \ \sum_{\tau=0}^{\infty} |k(\tau)| f^{-1}(\tau) \leq 1 \right\}, \tag{3.19}$$

*then*

$$d^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = f(n), \tag{3.20}$$

*and the subspace $L^n_{opt} = \{k : \ k(i) = 0, \ i = 0, 1, \ldots, n-1\}$ is optimal for $d^n$.*

**Proof** We will prove that $\mathbf{S}_a$ is 1-monotone decreasing. For this it is enough to show that every boundary function in $\mathbf{S}_a|_{[0,m]}$ is not less than the tail function $k_m$ defined by

$$k_m(i) = \begin{cases} 0 & i \neq m \\ f(m) & i = m \end{cases} \tag{3.21}$$

Let $k \in \mathbf{S}_a|_{[0,m]}$ be a boundary function, i.e., $\sum_{i=0}^{m} |k(i)| f^{-1}(i) = 1$. Then, as $f$ is monotone decreasing,

$$\|k\|_{\ell^1} \geq f(m) \sum_{i=0}^{m} |k(i)| f^{-1}(i) = f(m) = \|k_m\|.$$

Thus $\mathbf{S}_a$ satisfies the monotonicity hypothesis in Theorem 3.1.

To show that the identity (3.9) holds, i.e., $\|\mathbf{S}_a|_{[n,n+1)}\|_{\ell^1} = \|\mathbf{S}_a|_{[n,\infty)}\|_{\ell^1}$, it is enough to show that $\|\mathbf{S}_a|_{[n,n+1)}\|_{\ell^1} \geq \|\mathbf{S}_a|_{[n,\infty)}\|_{\ell^1}$. Let $k \in \mathbf{S}_a|_{[n,\infty)}$. As $f$ is decreasing and $\sum_{i=n}^{\infty} |k(i)| f^{-1}(i) \leq 1$, we get

$$\|k\|_{\ell^1} \leq f(n) \sum_{i=n}^{\infty} |k(i)| f^{-1}(i) \leq f(n).$$

Hence, $\|\mathbf{S}_a|_{[n,\infty)}\|_{\ell^1} \leq f(n) = \|k_n\|_{\ell^1} \leq \|\mathbf{S}_a|_{[n,n+1)}\|_{\ell^1}$, as $k_n \in \mathbf{S}_a|_{[n,n+1)}$.

Therefore, $d^n = \|\mathbf{S}_a|_{[n,n+1)}\|_{\ell^1} = f(n)$ and the subspace $L^n_{opt} = \{k : k(i) = 0,\ i = 0, 1, \ldots, n-1\}$ is optimal.

□

A slightly different version of the next proposition is a standard theorem in [38, Theorem 2.1, p.250] which can be deduced from our common principle.

**Corollary 3.3** *Let* $\mathbf{L}$ *be the Wiener algebra, i.e., the* $\mathbf{H}^\infty$*-normed algebra of functions in* $\mathbf{H}^\infty(D)$ *with Fourier coefficients in* $\ell^1[0,\infty)$. *Let* $0 < r \leq 1$, $l \geq 0$ *and* $l \neq 0$ *if* $r = 1$. *If*

$$\mathbf{S}(r,l) = \left\{ K : \ K^{(l)} \in \mathbf{H}^\infty(D_{r^{-1}}),\ \|K^{(l)}\|_{\infty,r^{-1}} \leq C \right\}, \tag{3.22}$$

*where $K^{(l)}$ denotes the $l$-th derivative of $K$, then*

$$d^n\left(\mathbf{S}(r,l),\mathbf{L}\right) = \begin{cases} \|\mathbf{S}(r,l)\|_{\mathbf{L}} & n < l, \\ \frac{(n-l)!}{n!}Cr^{n-l} & n \geq l. \end{cases} \tag{3.23}$$

*and the subspace $L^n_{opt} = \{k:\ k(i) = 0,\ i = 0, 1, \ldots, n-1\}$ is optimal.*

**Proof** The proof is trivial when $n < l$. We prove this corollary for $n \geq l$. $\mathbf{S}(r,l)$ is l-monotone decreasing, i.e., every boundary vector in $\mathbf{S}(r,l)|_{[0,n]}$ is not less than the tail vector $k_n$ defined by

$$k_n(i) = \begin{cases} 0 & i \neq n \\ \frac{(n-l)!}{n!}Cr^{n-l} & i = n \end{cases} \tag{3.24}$$

This fact follows from a theorem in Pinkus [38, Theorem 2.1, p.250] which implies that $\mathbf{S}(r,l)|_{[0,n]}$ contains a $n+1$-dimensional $\mathbf{H}^\infty$ ball of radius $\frac{(n-l)!}{n!}Cr^{n-l}$. That radius equals the $\mathbf{H}^\infty$ norm of $k_n$, which in turn equals $\|\mathbf{S}(r,l)|_{[n,\infty)}\|_{\mathbf{L}}$. Since $k_n$ as defined in (3.24) is in $\mathbf{S}(r,l)|_{[n,n+1)}$, $\|\mathbf{S}(r,l)|_{[n,\infty)}\|_{\mathbf{L}} = \|k_n\|_{\mathbf{L}} = \|\mathbf{S}(r,l)|_{[n,n+1)}\|_{\mathbf{L}}$. It follows that identity (3.9) holds. Therefore, $d^n = \frac{(n-l)!}{n!}Cr^{n-l}$ and the subspace $L^n_{opt} = \{k:\ k(i) = 0,\ i = 0, 1, \ldots, n-1\}$ is optimal.

□

## 3.5 Estimation of the Time n-Width

Invoking Proposition 3.2 and Theorem 3.1, we obtain an estimate of time n-width of a $q$-monotone decreasing data set.

**Theorem 3.2** *Let $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty,\infty)$. If the a priori set $\mathbf{S}_a$ is $q$-monotone decreasing, then the n-width $\theta^n$ has bounds*

$$\|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} \leq \theta^n\left(\mathbf{S}_a,\mathbf{L}\right) \leq \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}, \quad (p \leq q,\ p < \infty). \tag{3.25}$$

*Moreover, if*

$$\lim_{p \to q} \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}, \tag{3.26}$$

*then* $\theta^n(\mathbf{S}_a, \mathbf{L}) = d^n(\mathbf{S}_a, \mathbf{L}) = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}$, *and an impulse at the start of the observation interval is optimal for* $\theta^n$.

**Proof** The theorem can be proved by invoking Theorem 3.1 and Proposition 3.2. The upper bound in (3.25) is achieved by taking $u$ to be an impulse at the start of the observation interval. To show the lower bound in (3.25), we notice that for all $p \leq q$, $p < \infty$, $\theta^n(\mathbf{S}_a, \mathbf{L}) \geq \theta^n\left(\mathbf{S}_a|_{[0,n+p)}, \mathbf{L}\right)$. Since $\mathbf{S}_a|_{[0,n+p)}$ is contained in a finite dimensional subspace, and in such a case the hypotheses of Proposition 3.2 are automatically fulfilled, $\theta^n(\mathbf{S}_a, \mathbf{L}) \geq d^n\left(\mathbf{S}_a|_{[0,n+p)}, \mathbf{L}\right)$, by that proposition. Since $\mathbf{S}_a$ is monotone decreasing by hypothesis, $\mathbf{S}_a|_{[0,n+p)}$ is also monotone decreasing. Hence by Theorem 3.1 $\theta^n(\mathbf{S}_a, \mathbf{L}) \geq \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}}$. Noticing that the above inequality holds for all finite $p \leq q$, we get the lower bound in (3.25). When the upper bound coincides with lower bound, by Theorem 3.1 $\theta^n = d^n = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}$, and the optimality of the impulse follows from definition of $\theta^n$.

□

The estimates of time n-width described in Example 2.1 are established by the following corollaries to Theorem 3.2, which follow immediately from the corollaries to Theorem 3.1.

**Corollary 3.4** *Let* $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty, \infty)$. *If* $\mathbf{S}_a$ *is as defined in Corollary 3.1, then*

$$\theta^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = d^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = \|\mathbf{P}_{[n,\infty)}(f)\|_{\ell^1}, \tag{3.27}$$

*and the optimal input is an impulse at the start of the observation interval.*

**Corollary 3.5** *Let* $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty, \infty)$. *If* $\mathbf{S}_a$ *is as defined in Corollary 3.2, then*

$$\theta^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = d^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = f(n), \tag{3.28}$$

*and the optimal input is an impulse at the start of the observation interval.*

**Corollary 3.6** *Let* $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty, \infty)$ *and* $\mathbf{L}$ *be the Wiener algebra. If* $\mathbf{S}_a$ *is as defined in Corollary 3.3, then*

$$\theta^n\left(\mathbf{S}(r,l), \mathbf{L}\right) = d^n\left(\mathbf{S}(r,l), \mathbf{L}\right) = \begin{cases} \|\mathbf{S}(r,l)\|_{\mathbf{L}} & n < l, \\ \frac{(n-l)!}{n!} C r^{n-l} & n \geq l, \end{cases} \tag{3.29}$$

*and the optimal input is an impulse at the start of the observation interval.*

## 3.6 Estimation of Kolmogorov n-Width

For estimation of the Kolmogorov n-width of a $q$-monotone decreasing data set, we will need an additional assumption.

**Assumption 2** $\mathbf{S}_a$ *is closed under the truncations* $\mathbf{P}_{[n,\infty)}$*, or more generally*[2]*, there is a causal Cesaro operation* $\mathcal{C}_n$ *which maps* $\mathbf{S}_a$ *into* $\mathcal{C}_n(\mathbf{S}_a)$ *and satisfies*

$$\left\| \left(\mathbf{I} - \mathbf{P}_{[0,n)}\mathcal{C}_n\right) \mathbf{S}_a \right\|_{\mathbf{L}} \leq \left\| \mathbf{S}_a|_{[n,\infty)} \right\|_{\mathbf{L}}. \tag{3.30}$$

Here a "Cesaro operation" is any map satisfying (3.30), and is so called because it will typically be obtained via a Cesaro summation; i.e., the first $n$ samples of each impulse response in $\mathbf{S}$ will be multiplied by a weighting function which deceases with time.

**Theorem 3.3** *Under the Assumption 2, if the a priori set* $\mathbf{S}_a$ *is* $q$*-monotone decreasing, then the Kolmogorov n-width* $d_n$ *has bounds*

$$\|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} \leq d_n\left(\mathbf{S}_a, \mathbf{L}\right) \leq \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}, \quad (p \leq q,\ p < \infty). \tag{3.31}$$

*Moreover, if*

$$\lim_{p \to q} \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}, \tag{3.32}$$

[2]If $\mathbf{S}_a$ is closed under truncations then (3.30) is satisfied with $\mathcal{C}_n$ equal to the identity

*then* $d_n(\mathbf{S}_a, \mathbf{L}) = \theta^n(\mathbf{S}_a, \mathbf{L}) = d^n(\mathbf{S}_a, \mathbf{L}) = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}$, *where* $\mathbf{U} = \ell^\infty(-\infty, \infty)$ *in* $\theta^n$, *and the span* $\mathbf{L}_n := sp\{1, z, \ldots, z^{n-1}\}$ *is an optimal subspace.*

In the proof of the theroem, we will use the following standard result in duality theory [27].

**Lemma 3.1** *Let* $\mathbf{L}$ *be a normed space,* $\mathbf{X}$ *be a subspace and* $\mathbf{X}^\perp$ *be the annihilator of* $\mathbf{X}$. *Then for each* $k \in \mathbf{L}$

$$\inf_{x\in\mathbf{X}} \|k - x\| = \max_{\psi\in\mathbf{X}^\perp, \|\psi\|\le 1} \psi(k), \tag{3.33}$$

*where "max" indicates the supremum is attained.*

**Proof of Theorem 3.3** To show the lower bound in (3.31), fix $p \le q$, $p < \infty$. By definition of $d_n$,

$$\begin{aligned} d_n(\mathbf{S}_a, \mathbf{L}) &\ge d_n\left(\mathbf{S}_a|_{[0,n+p)}, \mathbf{L}\right) & (3.34)\\ &= \inf_{\mathbf{X}_n} \sup_{k\in\mathbf{S}_a|_{[0,n+p)}} \inf_{x\in\mathbf{X}_n} \|k - x\|_{\mathbf{L}} & (3.35)\end{aligned}$$

By Lemma 3.1, the last infimum can be replaced by a maximum as following:

$$\inf_{x\in\mathbf{X}_n} \|k - x\|_{\mathbf{L}} = \max_{\psi\in\mathbf{X}_n^\perp, \|\psi\|\le 1} \psi(k). \tag{3.36}$$

Therefore, we have

$$d_n(\mathbf{S}_a, \mathbf{L}) \ge \inf_{\mathbf{X}_n} \sup_{k\in\mathbf{S}_a|_{[0,n+p)}} \max_{\psi\in\mathbf{X}_n^\perp, \|\psi\|\le 1} \psi(k). \tag{3.37}$$

Assume $x_1, x_2, \ldots, x_n$ form a basis of $\mathbf{X}_n$, i.e., $\mathbf{X}_n = sp\{x_1, x_2, \ldots, x_n\}$. Each $x_i$ is a sequence, $x_i = \{x_i(0), x_i(1), \ldots\}$. Put

$$X = \begin{pmatrix} x_1(0) & \ldots & x_1(n+p-1) \\ \vdots & \ldots & \vdots \\ x_n(0) & \ldots & x_n(n+p-1) \end{pmatrix}. \tag{3.38}$$

We have

$$\mathbf{X}_n^{\perp}|_{[0,n+p)} = \left\{ k \in \mathbf{L} : \ k = \sum_{t=0}^{n+p-1} k(t) z^t, \ (k(0), \ldots, k(n+p-1))^T \in Null(X) \right\}. \tag{3.39}$$

Obviously, $\mathbf{X}_n^{\perp}|_{[0,n+p)}$ is a subspace of $\mathbf{X}_n^{\perp}$, and

$$\dim\left(\mathbf{X}_n^{\perp}|_{[0,n+p)}\right) = \dim(Null(X)) \geq p. \tag{3.40}$$

By the definition of $p$-section, the boundary of $\mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}$ is either a $p$-section or contains a $p$-section. Therefore the monotonicity of the a priori uncertainty set implies that

$$\|\mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}\|_{\mathbf{L}} \geq \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}}, \quad \forall \mathbf{X}_n. \tag{3.41}$$

It will be shown that

$$\sup_{k \in \mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}} \ \max_{\psi \in \mathbf{X}_n^{\perp}, \|\psi\| \leq 1} \psi(k) \geq \|\mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}\|_{\mathbf{L}}. \tag{3.42}$$

Therefore,

$$\begin{aligned} d_n(\mathbf{S}_a, \mathbf{L}) &\geq \inf_{\mathbf{X}_n} \sup_{k \in \mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}} \ \max_{\psi \in \mathbf{X}_n^{\perp}, \|\psi\| \leq 1} \psi(k) \quad (\text{by } (3.37)), \\ &\geq \inf_{\mathbf{X}_n} \|\mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}\|_{\mathbf{L}} \quad (\text{by } (3.42)), \\ &\geq \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}}, \quad (\text{by } (3.41)), \end{aligned}$$

which is the desired lower bound in (3.31).

To show (3.42), we notice that for each $k \in \mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}$, $\psi_k := \frac{\|k\|_{\mathbf{L}}}{\|k\|_2^2} \cdot k$ is a functional defined on $sp\{x_1, x_2, \ldots, k\}$, and its norm equals one. By the Hahn-Banach Theorem, it can be extended to $\mathbf{L}$ and the extended functional also has unit norm. If we denote the extended functional still by $\psi_k$, then $\psi_k \in \mathbf{X}_n^{\perp}$ and $\|\psi_k\| = 1$. Therefore,

$$\sup_{k \in \mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}} \ \max_{\psi \in \mathbf{X}_n^{\perp}, \|\psi\| \leq 1} \psi(k) \ \geq \ \sup_{k \in \mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^{\perp}|_{[0,n+p)}} \psi_k(k), \tag{3.43}$$

$$= \sup_{k \in \mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^\perp|_{[0,n+p)}} \|k\|_{\mathbf{L}}. \quad (3.44)$$

$$= \|\mathbf{S}_a|_{[0,n+p)} \bigcap \mathbf{X}_n^\perp|_{[0,n+p)}\|_{\mathbf{L}}. \quad (3.45)$$

By definition, $d_n$ has the following upper bound

$$d_n(\mathbf{S}_a, \mathbf{L}) \le \sup_{k \in \mathbf{S}_a} \inf_{x \in \mathbf{L}_n} \|k - x\|_{\mathbf{L}}. \quad (3.46)$$

Set $x = \mathbf{P}_{[0,n)}\mathcal{C}_n(k)$, where $\mathcal{C}_n$ is a Cesaro operation satisfying (3.30) in Assumption 2. We get

$$d_n(\mathbf{S}_a, \mathbf{L}) \le \sup_{k \in \mathbf{S}_a} \|k - \mathbf{P}_{[0,n)}\mathcal{C}_n(k)\|_{\mathbf{L}}, \quad (3.47)$$

$$\le \|\mathbf{S}_a|_{[0,\infty)}\|_{\mathbf{L}}. \text{ (by Assumption 2)} \quad (3.48)$$

When the upper bound coincides with the lower bound, by Theorem 3.1 and Theorem 3.2, $d_n(\mathbf{S}_a, \mathbf{L}) = \theta^n(\mathbf{S}_a, \mathbf{L}) = d^n(\mathbf{S}_a, \mathbf{L}) = \|\mathbf{S}_a|_{[0,\infty)}\|_{\mathbf{L}}$, the optimality of the subspace $\mathbf{L}_n$ follows from the definition.

□

The estimates of Kolmogorov n-width described in Example 2.1 are established by the following corollaries to Theorem 3.3

**Corollary 3.7** *If $\mathbf{S}_a$ is as defined in Corollary 3.1, then*

$$d_n\left(\dot{\mathbf{S}}_a, \ell^1[0,\infty)\right) = \theta^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = d^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = \|\mathbf{P}_{[n,\infty)}(f)\|_{\ell^1}, \quad (3.49)$$

*where $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty,\infty)$ in $\theta^n$, and the span $\mathbf{L}_n := sp\{1, z, \ldots, z^{n-1}\}$ is an optimal subspace.*

**Proof** To apply Theorem 3.3, it is enough to show that $\mathbf{S}_a$ satisfies Assumption 2, and the rest will follow from the proof of Corollary 3.1. Since $\mathbf{S}_a$ is closed under truncation, Assumption 2 is satisfied.

□

**Corollary 3.8** *If* $\mathbf{S}_a$ *is as defined in Corollary 3.2, then*

$$d_n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = \theta^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = d^n\left(\mathbf{S}_a, \ell^1[0,\infty)\right) = f(n), \tag{3.50}$$

*where* $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty,\infty)$ *in* $\theta^n$, *and the span* $\mathbf{L}_n := sp\{1, z, \ldots, z^{n-1}\}$ *is an optimal subspace.*

**Proof** Since $\mathbf{S}_a$ is closed under truncation, Assumption 2 is satisfied. The rest of the proof follows from the proof of Corollary 3.2.

□

Example 2.1, parts (i) and (ii) follow from the preceding corollaries when $f(t) = Cr^t$.

**Corollary 3.9** *Let* $\mathbf{L}$ *be the Wiener algebra. If* $\mathbf{S}_a$ *is as defined in Corollary 3.3, then*

$$d_n(\mathbf{S}(r,l), \mathbf{L}) = \theta^n(\mathbf{S}(r,l), \mathbf{L}) = d^n(\mathbf{S}(r,l), \mathbf{L}) = \begin{cases} \|\mathbf{S}(r,l)\|_{\mathbf{L}} & n < l, \\ \frac{(n-l)!}{n!} C r^{n-l} & n \geq l, \end{cases} \tag{3.51}$$

*where* $\mathbf{E} = \mathbf{U} = \ell^\infty(-\infty,\infty)$ *in* $\theta^n$, *and the span* $\mathbf{L}_n := sp\{1, z, \ldots, z^{n-1}\}$ *is an optimal subspace.*

**Proof** Theorem 2.1 in [38, p.250] implies that there exists a mapping $\mathcal{C}_n$ on $\mathbf{L}$ such that $\forall k \in \mathbf{S}(r,l)$,

$$\left\|\left(\mathbf{I} - \mathbf{P}_{[0,n)}\mathcal{C}_n\right)(k)\right\|_{\mathbf{L}} = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}}.$$

Hence, Assumption 2 is satisfied. The rest of the proof follows from the proof of Corollary 3.3.

□

Example 2.1, parts (iii) and (iv) are special cases of Corollary 3.6 for the sets $\mathbf{S}(r,0)$ and $\mathbf{S}(1,1)$.

Under the conditions of Theorem 3.3, the FIR model is optimal for reducing modeling error. More than that, the FIR model taken together with an impulse input at the start of the observation interval, form an optimal (model set and input) pair for reducing the worst-case identification error.

## 3.7 The Optimal Worst-Case Identification Error

**Theorem 3.4** *Let* $\mathbf{U} = \ell^{\infty}(-\infty, \infty)$. *Under the Assumption 1 and Assumption 2, if the a priori data set is q-monotone decreasing and identity*

$$\lim_{p \to q} \|\mathbf{S}_a|_{[n,n+p)}\|_{\mathbf{L}} = \|\mathbf{S}_a|_{[n,\infty)}\|_{\mathbf{L}} \tag{3.52}$$

*holds, then the optimal noise free worst-case identification error*

$$\begin{aligned} \inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \subset \mathbf{L}} e^T(u, \mathbf{X}_n) &= \max\left\{d^T(\mathbf{S}_a, \mathbf{L}), d_n(\mathbf{S}_a, \mathbf{L})\right\} &(3.53)\\ &= \|\mathbf{S}_a|_{[m,\infty)}\|_{\mathbf{L}}, &(3.54)\end{aligned}$$

*where* $m = \min\{n, T\}$. *The optimal input is an impulse at the start of the observation interval and the optimal affine model set is the FIR model* $\mathbf{L}_n = sp\{1, z, \ldots, z^{n-1}\}$.

**Proof** Theorem 3.3 implies (3.54). By Proposition 2.2 and Theorem 3.2, we have

$$\inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \subset \mathbf{L}} e^T(u, \mathbf{X}_n) \geq \max\left\{d^T(\mathbf{S}_a, \mathbf{L}), d_n(\mathbf{S}_a, \mathbf{L})\right\}. \tag{3.55}$$

It is left to show that

$$\inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \subset \mathbf{L}} e^T(u, \mathbf{X}_n) \leq \max\left\{d^T(\mathbf{S}_a, \mathbf{L}), d_n(\mathbf{S}_a, \mathbf{L})\right\}. \tag{3.56}$$

Let $u$ be an impulse at $t_0$, the start of the observation interval and $\mathbf{X}_n = \mathbf{L}_n$. By definition of $e^T(u, \mathbf{X}_n)$,

$$\inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \subset \mathbf{L}} e^T(u, \mathbf{X}_n) \leq \sup_{k_{true} \in \mathbf{S}_a} \inf_{k_{est} \in \mathbf{L}_n} \sup_{k \in \mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}}. \tag{3.57}$$

Since the input is an impulse at the start of the observation interval, the a posteriori uncertainty set $\mathbf{S}(y)$ has the form

$$\mathbf{S}(y) = \left\{k \in \mathbf{S}_a : \ \mathbf{P}_{[0,T)}(k) = \mathbf{P}_{[0,T)}(k_{true})\right\} \qquad (3.58)$$

Let $\mathcal{C}_m$ be a Cesaro operation satisfying (3.30) in Assumption 2, and $k_{est} = \mathbf{P}_{[0,m)}\mathcal{C}_m$ $(k_{true})$, which is in $\mathbf{L}_n$. By the causality of $\mathcal{C}_m$, $k_{est}$ is equal to $\mathbf{P}_{[0,m)}\mathcal{C}_m\mathbf{P}_{[0,m)}\,(k_{true})$. Therefore, for all $k \in \mathbf{S}\,(y)$,

$$\begin{aligned}
\|k - k_{est}\|_{\mathbf{L}} &= \|k - \mathbf{P}_{[0,m)}\mathcal{C}_m\mathbf{P}_{[0,m)}\,(k_{true})\,\|_{\mathbf{L}}, \quad \text{(by the causality of } \mathcal{C}_m.) \\
&= \|k - \mathbf{P}_{[0,m)}\mathcal{C}_m\mathbf{P}_{[0,m)}(k)\|_{\mathbf{L}} \\
&\qquad \text{(because } \mathbf{P}_{[0,m)}(\mathrm{k}) = \mathbf{P}_{[0,m)}\,(\mathrm{k}_{true}) \ \text{ for } \mathrm{k} \in \mathbf{S}\,(\mathrm{y})) \\
&= \|k - \mathbf{P}_{[0,m)}\mathcal{C}_m(k)\|_{\mathbf{L}} \quad \text{(by the causality of } \mathcal{C}_m.) \\
&\leq \left\|\mathbf{S}_a|_{[m,\infty)}\right\|_{\mathbf{L}} \quad \text{(by (3.30) in Assumption 2).}
\end{aligned}$$

Since the above inequalities hold for all $k_{true} \in \mathbf{S}_a$ and all $k \in \mathbf{S}(y)$, noticing (3.57), we get the desired upper bound.

□

The optimal worst-case identification errors for the data sets in Example 2.1 are derived from the following Corollaries to Theorem 3.4. The fact that all the sets in these corollaries satisfy the conditions in Theorem 3.4 follows from the corollaries to Theorem 3.1 and Theorem 3.3.

**Corollary 3.10** *Let $\mathbf{U} = \ell^\infty(-\infty,\infty)$. If $\mathbf{S}_a$ is defined as in Corollary 3.1, then*

$$\inf_{u\in\mathbf{U}} \ \inf_{\mathbf{X}_n\subset\ell^1} e^T(u,\mathbf{X}_n) = \max\left\{d^T\left(\mathbf{S}_a,\ell^1\right), d_n\left(\mathbf{S}_a,\ell^1\right)\right\} = \|\mathbf{P}_{[m,\infty)}(f)\|_{\ell^1}, \qquad (3.59)$$

*where $m = \min\{n,T\}$. The optimal input is an impulse at the start of the observation interval and the optimal affine model set is the FIR model $\mathbf{L}_n = sp\,\{1, z, \ldots, z^{n-1}\}$.*

**Corollary 3.11** *Let* $\mathbf{U} = \ell^{\infty}(-\infty, \infty)$. *If* $\mathbf{S}_a$ *is defined as in Corollary 3.2, then*

$$\inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \subset \ell^1} e^T(u, \mathbf{X}_n) = \max\left\{d^T\left(\mathbf{S}_a, \ell^1\right), d_n\left(\mathbf{S}_a, \ell^1\right)\right\} = f(m), \tag{3.60}$$

*where* $m = \min\{n, T\}$. *The optimal input is an impulse at the start of the observation interval and the optimal affine model set is the FIR model* $\mathbf{L}_n = sp\{1, z, \ldots, z^{n-1}\}$.

Example 2.1, parts (i) and (ii) follow from the preceding corollaries when $f(t) = Cr^t$.

**Corollary 3.12** *Let* $\mathbf{U} = \ell^{\infty}(-\infty, \infty)$ *and* $\mathbf{L}$ *be the Weiner algebra. Let* $0 < r \leq 1$ *and* $p \neq 0$ *if* $r = 1$. *If* $\mathbf{S}_a$ *is defined as in Corollary 3.3, then*

$$\inf_{u \in \mathbf{U}} \inf_{\mathbf{X}_n \subset \mathbf{L}} e^T(u, \mathbf{X}_n) = \max\left\{d^T(\mathbf{S}_a, \mathbf{L}), d_n(\mathbf{S}_a, \mathbf{L})\right\} = \begin{cases} C & m < p, \\ \frac{(m-p)!}{m!} C r^{m-p} & m \geq p, \end{cases} \tag{3.61}$$

*where* $m = \min\{n, T\}$. *The optimal input is an impulse at the start of the observation interval and the optimal affine model set is the FIR model* $\mathbf{L}_n = sp\{1, z, \ldots, z^{n-1}\}$.

Example 2.1, parts (iii) and (iv) are special cases of Corollary 3.12 for the sets $\mathbf{S}(r, 0)$ and $\mathbf{S}(1, 1)$.

## 3.8 Estimation of Shift Invariant Time n-Width

**Proposition 3.3** *Under the hypotheses of Example 2.2,*

$$\theta^n(\mathbf{S}_{a_i}, \mathbf{L}) \leq \bar{\theta}^n(\mathbf{S}_{a_i}, \mathbf{L}) \leq \alpha \theta^n(\mathbf{S}_{a_i}, \mathbf{L}). \tag{3.62}$$

*where the upper bound is valid for the following* $\alpha$. *For* $\mathbf{S}_{a_1}$, $\mathbf{S}_{a_2}$ *and* $\mathbf{S}_{a_3}$, $\alpha = 2$; *for* $\mathbf{S}_{a_4}$, $\alpha = \pi$.

**Proof** The lower bound follows from the fact that the supremum defining $\bar{\theta}^n$ is over a larger set than for $\theta^n$.

For the upper bound, we consider a sequence of impulses as the input, i.e., $u(t) = 1$ for $t = mn$, $m \in \mathbb{Z}$, and $u(t) = 0$ elsewhere. Since the input is a periodic function of period $n$, the output is similarly a periodic function and is zero on an interval of length $n$ if and only if it is zero everywhere. In this case the location of the observation interval does not affect $\delta_0^T(u)$ and we can arbitrarily set $t_0 = 0$, whereupon

$$\bar{\theta}^n(\mathbf{S}_{a_i}, \mathbf{L}) \leq \sup\left\{\|k\|_{\mathbf{L}} :\ k \in \mathbf{S}_{a_i},\ (Ku)(t) = 0,\ 0 \leq t < n\right\}. \qquad (3.63)$$

i) In the cases of $\mathbf{S}_{a_1}$ and $\mathbf{S}_{a_2}$, $(Ku)(t) = 0,\ 0 \leq t < n$, implies that

$$\sum_{\tau=-t}^{\infty} k(t+\tau)u(-\tau) = 0,\ 0 \leq t < n$$

Since $u$ is a sequence of impulses, we have

$$\sum_{m=0}^{\infty} k(t+mn) = 0,\ 0 \leq t < n,$$

i.e. $k(t) = -\sum_{m=1}^{\infty} k(t+mn),\ 0 \leq t < n$. It follows that

$$\begin{aligned} \|k\|_{\ell^1} &= \sum_{t=0}^{n-1} |k(t)| + \sum_{t=n}^{\infty} |k(t)|, \\ &\leq \sum_{t=0}^{n-1} \sum_{m=1}^{\infty} |k(t+mn)| + \|\mathbf{P}_{[n,\infty)}(k)\|_{\ell^1}, \\ &= 2\|\mathbf{P}_{[n,\infty)}(k)\|_{\ell^1}. \end{aligned}$$

By (3.63), we have

$$\bar{\theta}^n\left(\mathbf{S}_{a_i}, \ell^1\right) \leq 2\sup\left\{\|\mathbf{P}_{[n,\infty)}(k)\|_{\ell^1} :\ k \in \mathbf{S}_{a_i}\right\} = 2\theta^n\left(\mathbf{S}_{a_i}, \ell^1\right).$$

ii) In the case of $\mathbf{S}_{a_3}$, considering the discrete Fourier transforms of both the input and the output, we have

$$Y\left(e^{i\omega_j}\right) = K\left(e^{i\omega_j}\right) U\left(e^{i\omega_j}\right),$$

where $K(\omega_j) = \sum_{\tau=0}^{\infty} k(\tau) e^{i\omega_j \tau}$ and $\omega_j = \frac{2\pi j}{n}$, $j = 0, 1, \cdots, n-1$. Therefore by (3.63),

$$\begin{aligned} \theta^n(\mathbf{S}_{a_3}, \mathbf{H}^\infty) &\leq \sup\left\{\|K\|_\infty : K \in \mathbf{S}_{a_3},\ K\left(e^{i\omega_j}\right) U\left(e^{i\omega_j}\right) = 0,\ j = 0, \ldots, n-1\right\}, \\ &= \sup\left\{\|K\|_\infty : K \in \mathbf{S}_{a_3},\ K\left(e^{i\omega_j}\right) = 0,\ j = 0, 1, \ldots, n-1\right\}, \\ &= \sup\left\{\|K\|_\infty : K \in B_n \mathbf{S}_{a_3}\right\}. \end{aligned}$$

where $B_n$ is a Blaschke product in $\mathbf{H}^\infty(D_{r^{-1}})$,

$$B_n(z) = \prod_{j=0}^{n-1} \frac{rz - re^{i\omega_j}}{1 - r^2 e^{i\omega_j} z} = \frac{r^n(z^n - 1)}{1 - r^{2n} z^n}.$$

This implies that $\theta^n(\mathbf{S}_{a_3}, \mathbf{H}^\infty) \leq C\|B_n\|_\infty$. To compute $\|B_n\|_\infty$, we notice that

$$\begin{aligned} |B_n\left(e^{i\omega}\right)|^2 &= \left|\frac{r^n\left(e^{in\omega} - 1\right)}{1 - r^{2n} e^{in\omega}}\right|^2, \\ &= 2r^{2n} \frac{1 - cos(n\omega)}{1 + r^4 - 2r^2 cos(n\omega)}, \\ &\leq 4r^{2n}. \end{aligned}$$

Therefore,

$$\bar{\theta}^n(\mathbf{S}_{a_3}, \mathbf{H}^\infty) \leq 2Cr^n = 2\theta^n(\mathbf{S}_{a_3}, \mathbf{H}^\infty).$$

iii) In the case of $\mathbf{S}_{a_4}$, if $K \in \mathbf{S}_{a_4}$, then $K$ has bounded derivative in $D$ and Hardy's inequality implies that $K(z) = \sum_{\tau=0}^{\infty} k(\tau) z^\tau$ and $\sum_{\tau=0}^{\infty} |k(\tau)| < \infty$. Hence $K(e^{i\omega})$ is defined for all $\omega$ and $K(e^{i\omega}) = \lim_{r\to 1} K(re^{i\omega})$. It follows that

$$\bar{\theta}^n(\mathbf{S}_{a_4}, \mathbf{H}^\infty) \leq \sup\left\{\|K\|_\infty : \|K'\| \leq C,\ K\left(e^{i\omega_j}\right) = 0,\ j = 0, 1, \cdots, n-1\right\}.$$

Now we prove that $K(e^{i\omega})$ is Lipschitz continuous with Lipschitz constant $\|K'\|_\infty$. Since $K(z) \in \mathbf{H}(D)$, integration on the arc $\{re^{i\omega} : \omega_1 \leq \omega \leq \omega_2\}$ gives us

$$\begin{aligned} |K\left(re^{i\omega_1}\right) - K\left(re^{i\omega_2}\right)| &= \left|\int_{\omega_2}^{\omega_1} \frac{d}{d\omega} K\left(re^{i\omega}\right) d\omega\right|, &(3.64) \\ &= \left|\int_{\omega_2}^{\omega_1} \frac{dK\left(re^{i\omega}\right)}{d\left(re^{i\omega}\right)} \frac{d\left(re^{i\omega}\right)}{d\omega} d\omega\right|, &(3.65) \\ &\leq \int_{\omega_2}^{\omega_1} r|K'\left(re^{i\omega}\right)| d\omega, &(3.66) \\ &\leq r\|K'\|_\infty |\omega_1 - \omega_2|. &(3.67) \end{aligned}$$

Therefore,

$$\begin{aligned} \left|K\left(e^{i\omega_1}\right) - K\left(e^{i\omega_2}\right)\right| &= \lim_{r\to 1}\left|K\left(re^{i\omega_1}\right) - K\left(re^{i\omega_2}\right)\right|, & (3.68)\\ &\le \lim_{r\to 1} r\|K'\|_\infty|\omega_1 - \omega_2|, & (3.69)\\ &= \|K'\|_\infty|\omega_1 - \omega_2|. & (3.70)\end{aligned}$$

Hence, for $K \in \mathbf{S}_{a_4}$ satisfying the interpolations $K\left(e^{i\omega_j}\right) = 0$ for $j = 0, 1, \cdots, n-1$, we have

$$\|K\|_\infty \le \|K'\|_\infty \pi/n,$$

and

$$\bar{\theta}^n\left(\mathbf{S}_{a_4}, \mathbf{H}^\infty\right) \le \pi\frac{C}{n} = \pi\theta^n\left(\mathbf{S}_{a_4}, \mathbf{H}^\infty\right).$$

□

# Chapter 4

# Ensemble Input Design for Fast Identification

Assuming that an optimal model set has been chosen to represent the a priori data to a certain tolerance, a major contribution to the worst-case estimation error is the component we have called "inherent error", $\delta^T(u)$. In practical on-line identification, the input is seldom free to be optimized so as to reduce the the inherent error to its minimal value $\theta^T$; the input can only be modified to the extent of having certain desirable ensemble properties, e.g., flat spectrum, by introduction of a dither signal. In this Chapter, we will study how the input should be so modified.

The unmodeled dynamics affect identification as a multiplicative noise, which cannot be eliminated by increasing the input power or by averaging. A characterization of the identifying capability of an input will be given in the gap metric, in the presence of this multiplicative noise. The characterization makes use of certain "spectrum flatness" properties of the input, when the model set is an FIR (finite impulse response) model. In this case, the worst-case inherent error is bounded both above and below by functions of the spectrum flatness. These bounds become large when the spectrum is far from flat, which implies the necessity of a not-far-from- flat spectrum in

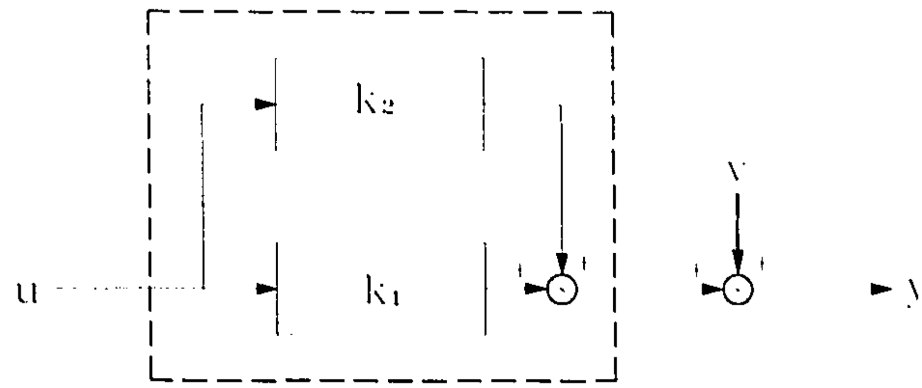

Figure 4.1: Unmodeled dynamics affects identification as multiplicative noise

nonparametric identification. (Portions of the paper are in [23].)

Input design problems have been considered in [11, 12].

## 4.1 Unmodeled Dynamics and the Multiplicative Noise

Let $\mathbf{X}_n$ be an optimal affine model set for $\mathbf{S}_a$, with the representation error $dist(\mathbf{X}_n,\mathbf{S}_a) = \epsilon$. Then each system $k \in \mathbf{S}_a$ can be decomposed into two parts, i.e., $k = k_1 + k_2$, where $k_1 \in \mathbf{X}_n$ and $k_2 \in \mathbf{B}(\epsilon)$, the $\epsilon$-ball in $\mathbf{L}$. We call $k_1$ *modelled dynamics* and $k_2$ *unmodelled dynamics*. If $k_{1,est} \in \mathbf{X}_n$ is an estimate for $k_1$ and chosen as a nominal model of the true system, the error between the nominal $k_{1,est}$ and the true system $k$ is bounded as following

$$\|k_{1,est} - k_1\|_{\mathbf{L}} - \epsilon \leq \|k_{1,est} - k\|_{\mathbf{L}} \leq \|k_{1,est} - k_1\|_{\mathbf{L}} + \epsilon. \tag{4.1}$$

Therefore, an accurate identification can be achieved by representing the uncertainty set accurately by a model set, and then estimating the modelled dynamics $k_1$ accurately.

Given an input $u \in \mathbf{U}$ and the output observations on $[t_0, t_0 + T)$, the modeled dynamics $k_1$ must satisfy the following equations:

$$y(t) = \Phi_u(k_1)(t) + (v(t) + \Phi_u(k_2)(t)) \quad \forall t \in [t_0, t_0 + T) \tag{4.2}$$

where $k_2$ represents the unmodeled dynamics. Since $k_2$ is unknown, the output of the unmodeled dynamics $\Phi_u(k_2)$ remains unknown even if the input $u$ is given. Therefore, the output of the unmodeled dynamics affects the identification of the modeled dynamics as a noise, as shown in Figure 4.1. The magnitude of this noise depends on the magnitude of the $k_2$ and the input. Hence, an accurate representation of the a priori uncertainty set not only reduces the representation error, but also reduces the disturbance caused by the unmodeled dynamics in identification of the modelled ones.

Generally speaking, the noise generated by the unmodeled dynamics behaves differently from the additive noise $v$ in two aspects. First, the magnitude of this noise increases with the magnitude of the input. Therefore, it can not be overcome by increasing the input power. We call this *multiplicative noise.* Second, unlike additive stochastic noise, the multiplicative noise can not be eliminated by averaging. The effects of multiplicative noise can only be reduced by representing the data set accurately and designing the input suitably.

To isolate the effects of multiplicative noise, we first assume that there is no additive noise in the measurements and the a priori uncertainty set is of the form:

$$\mathbf{S}_a = \mathbf{X}_n + \mathbf{B}(\epsilon), \tag{4.3}$$

where $\mathbf{X}_n$ is a $n$-dimensional subspace of $\mathbf{L}$ and $\mathbf{B}(\epsilon)$ is the $\epsilon$-ball of $\mathbf{L}$ centered at the origin. $\mathbf{X}_n$ and $\mathbf{B}(\epsilon)$ represent the modeled and unmodeled dynamics respectively. Later, we will study the cases where measurements are corrupted by additive noise and the a priori uncertainty set is a subset of $\mathbf{X}_n + \mathbf{B}(\epsilon)$ as in Example 2.1.

The effects of the multiplicative noise can be characterized by a map mapping unmodeled dynamics to modeled dynamics. If the norm of this map is large, then small unmodeled dynamics may cause large inherent error.

As we recall, the noise-free inherent error is

$$\begin{aligned} \delta_0^T(u) &= \sup\{\|k\|_{\mathbf{L}} :\ k \in \mathbf{S}_a,\ \mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u(k) = 0\} && (4.4) \\ &= \sup\{\|k\|_{\mathbf{L}} :\ k \in \mathbf{S}_a \bigcap Null(\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u)\}. && (4.5) \end{aligned}$$

For $k \in \mathbf{S}_a \bigcap Null(\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u)$, $k = k_1 + k_2$, where $k_1 \in \mathbf{X}_n$, $k_2 \in \mathbf{B}(\epsilon)$, we have

$$\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u(k_1) = -\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u(k_2). \tag{4.6}$$

Set

$$\begin{aligned} \mathbf{F} &= \mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u(\mathbf{X}_n), \\ \mathbf{S} &= \{k \in \mathbf{L} :\ \mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u(k) \in \mathbf{F}\}. \end{aligned}$$

$\mathbf{S}$ is a linear subspace of $\mathbf{L}$. Let $\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u|_{\mathbf{X}_n}$ and $\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u|_{\mathbf{S}}$ denote the restrictions of $\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u$ on $\mathbf{X}_n$ and $\mathbf{S}$ respectively. Since a $k_2$ satisfying (4.6) is in the subspace $\mathbf{S}$, the relation (4.6) is equivelent to

$$\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u|_{\mathbf{X}_n}(k_1) = -\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u|_{\mathbf{S}}(k_2). \tag{4.7}$$

Equation (4.7) defines a mapping from $\mathbf{S}$ to $\mathbf{X}_n$. We denote it by $\mathbf{M}$.

**Proposition 4.1**

$$\epsilon\,(1 + \|\mathbf{M}\|) \geq \delta_0^T(u) \geq \epsilon\,(\|\mathbf{M}\| - 1)\,, \tag{4.8}$$

*where* $\|\cdot\|$ *is the* $\mathbf{L}$ *induced norm.*

**Proof** Set

$$\mathbf{S}_1 := \{k \in \mathbf{L} :\ k = k_1 + k_2,\ k_1 = \mathbf{M}(k_2),\ k_2 \in \mathbf{S},\ \|k_2\|_{\mathbf{L}} \leq \epsilon\}.$$

By definition of $\delta_0^T(u)$ and (4.7), $\delta_0^T(u) = \|\mathbf{S}_1\|_{\mathbf{L}}$. Since every $k \in \mathbf{S}_1$ has the form $k = (\mathbf{I} + \mathbf{M})(k_2)$ for some $k_2 \in \mathbf{S}$ with $\|k_2\|_{\mathbf{L}} \leq \epsilon$, where $\mathbf{I}$ is the identity mapping, we have $\|\mathbf{S}_1\|_{\mathbf{L}} \leq (1 + \|\mathbf{M}\|)\epsilon$.

On the other hand, by definition of $\|\mathbf{M}\|$, $\forall \eta > 0$, $\exists k_2 \in \mathbf{S}$ such that

$$\|\mathbf{M}\| \leq \|\mathbf{M}(k_2)\|_{\mathbf{L}}/\|k_2\|_{\mathbf{L}} + \eta. \tag{4.9}$$

Set

$$k = \mathbf{M}\left(\frac{k_2}{\|k_2\|_{\mathbf{L}}} \cdot \epsilon\right) + \frac{k_2}{\|k_2\|_{\mathbf{L}}} \cdot \epsilon. \tag{4.10}$$

By definition of $\mathbf{S}_1$, $k \in \mathbf{S}_1$. Since $\delta_0^T(u) = \|\mathbf{S}_1\|_{\mathbf{L}}$, we have

$$\begin{aligned}
\delta_0^T(u) &\geq \|k\|_{\mathbf{L}} \\
&\geq \left\|\frac{\epsilon}{\|k_2\|_{\mathbf{L}}}\mathbf{M}(k_2)\right\|_{\mathbf{L}} - \epsilon \\
&\geq \epsilon(\|\mathbf{M}\| - \eta) - \epsilon \quad \text{(by (4.9))}.
\end{aligned}$$

Since this inequalities hold for all $\eta > 0$, we have $\delta_0^T(u) \geq \epsilon(\|\mathbf{M}\| - 1)$.

□

The Proposition 4.1 shows that it is the norm of the operator $\mathbf{M}$ characterizes the effects of the unmodeled dynamics in identifying the modeled dynamics. It will be shown that when $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_2$, $\|\mathbf{M}\|$ is related to the gap between $\mathbf{X}_n$ and the null space $Null(\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u)$. Especially, when $\mathbf{X}_n$ equals the FIR model, the optimal model for all the a priori sets in Example 2.1, this gap is related to some spectral property of the input. For simplicity, we denote the null space $Null(\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u)$ by $\mathbf{N}_{t_0}$.

## 4.2 A Characterization by Gap

The next theorem shows the relation between the inherent error $\delta_0^T(u)$ and the gap between $\mathbf{X}_n$ and $\mathbf{N}_{t_0}$.

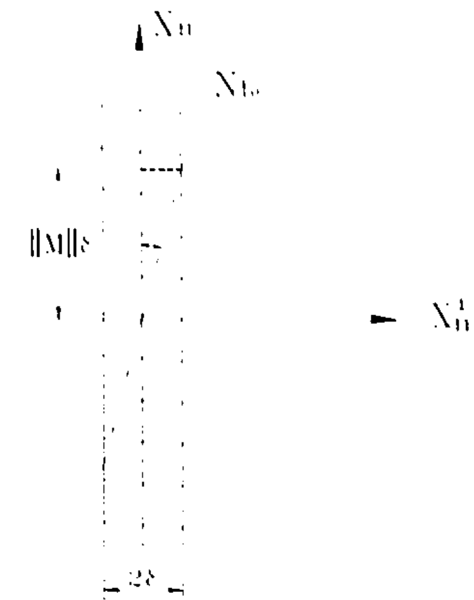

Figure 4.2: Characterization of multiplicative noise by the gap

**Definition 4.1** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be any two closed subspaces of a Hilbert space. The cosine of the angle between them is defined by*

$$\cos(\mathbf{A},\mathbf{B}) := \sup\{|(a,b)| : \|a\| = \|b\| = 1, a \in \mathbf{A}, b \in \mathbf{B}\}, \tag{4.11}$$

*where* $(\cdot,\cdot)$ *denotes the inner product on the Hilbert space. The direct gap between* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined by*

$$\vec{\delta}(\mathbf{A},\mathbf{B}) = \sup_{\|a\|=1, a\in\mathbf{A}} \inf_{b\in\mathbf{B}} \|a-b\|. \tag{4.12}$$

The cosine of the angle and the gap are related as shown in the next proposition, which is extracted from [35].

**Proposition 4.2**

$$\cos(\mathbf{A},\mathbf{B}) = \|\mathbf{P_A}\mathbf{P_B}\| = \vec{\delta}(\mathbf{A},\mathbf{B}^{\perp}), \tag{4.13}$$

*where* $\mathbf{P_A}$ *denotes the projection operator on* $\mathbf{A}$.

**Theorem 4.1** *Let* $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_2$. *If* $\mathbf{N}_{t_0}$ *is a closed subspace of* $\ell^2$ *and* $\mathbf{S}_a$ *is as in (4.3), then*

$$\delta_0^T(u) = \begin{cases} \infty & if \;\; \cos(\mathbf{X}_n, \mathbf{N}_{t_0}) = 1, \\ \epsilon\left(1 - \cos^2(\mathbf{X}_n, \mathbf{N}_{t_0})\right)^{-1/2} & otherwise. \end{cases} \tag{4.14}$$

**Proof** Set

$$\sin(\mathbf{A}, \mathbf{B}) := (1 - \cos^2(\mathbf{A}, \mathbf{B}))^{1/2}. \tag{4.15}$$

Then we have

$$\begin{aligned} \sin^2(\mathbf{X}_n, \mathbf{N}_{t_0}) &= 1 - \sup_{\|f\|=1} \|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}(f)\|^2, \\ &= 1 - \sup_{\|f\|=1, f\in\mathbf{N}_{t_0}} \|\mathbf{P}_{\mathbf{X}_n}(f)\|^2, \\ &= \inf_{\|f\|=1, f\in\mathbf{N}_{t_0}} (1 - \|\mathbf{P}_{\mathbf{X}_n}(f)\|^2), \\ &= \inf_{\|f\|=1, f\in\mathbf{N}_{t_0}} \|\mathbf{P}_{\mathbf{X}_n^\perp}(f)\|^2, \\ &= \inf_{\|f\|=1, f\in\mathbf{N}_{t_0}} \inf_{g\in\mathbf{X}_n} \|f - g\|^2. \end{aligned}$$

Since $\sin(\mathbf{X}_n, \mathbf{N}_{t_0}) = 0$ if and only if $\mathbf{X}_n \bigcap \mathbf{N}_{t_0} \neq \emptyset$, $\delta_0^T(u) = \infty$ if $\cos(\mathbf{S}, \mathbf{N}_{t_0}) = 1$.

Assume $\sin(\mathbf{X}_n, \mathbf{N}_{t_0}) \neq 0$. Since $\alpha \sin(\mathbf{X}_n, \mathbf{N}_{t_0}) = \inf_{\|k\|=\alpha, k\in\mathbf{N}_{t_0}} \|\mathbf{P}_{\mathbf{X}_n^\perp}(k)\|$, $\forall k \in \mathbf{N}_{t_0}$, $\|k\| \sin(\mathbf{X}_n, \mathbf{N}_{t_0}) \leq \|\mathbf{P}_{\mathbf{X}_n^\perp}(k)\|$. Also, by the structure of $\mathbf{S}_a$, $k \in \mathbf{S}_a$ if and only if $\|\mathbf{P}_{\mathbf{X}_n^\perp}(k)\| \leq \epsilon$. It follows that $\forall k \in \mathbf{N}_{t_0} \bigcap \mathbf{S}_a$, $\|k\| \sin(\mathbf{X}_n, \mathbf{N}_{t_0}) \leq \epsilon$. Therefore,

$$\delta_0^T(u) \leq \epsilon(1 - \cos^2(\mathbf{X}_n, \mathbf{N}_{t_0}))^{-1/2}.$$

On the other hand, $\forall \eta > 0$, $\exists k \in \mathbf{N}_{t_0}$ such that $\|k\| \sin(\mathbf{S}, \mathbf{N}_{t_0}) > \|\mathbf{P}_{\mathbf{X}_n^\perp}(k)\| - \eta$. Since $k$ can be chosen so that $\|\mathbf{P}_{\mathbf{X}_n^\perp}(k)\| = \epsilon$, $\forall \eta > 0$, $\exists k \in \mathbf{N}_{t_0} \bigcap \mathbf{S}_a$ such that $\|k\| \sin(\mathbf{X}_n, \mathbf{N}_{t_0}) > \epsilon - \eta$. This implies that

$$\delta_0^T(u) \geq \epsilon(1 - \cos^2(\mathbf{X}_n, \mathbf{N}_{t_0}))^{-1/2}.$$

□

By Proposition 4.2, $\cos(\mathbf{X}_n, \mathbf{N}_{t_0}) = \|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\|$. Exploiting the fact that $\mathbf{X}_n$ and $\mathbf{N}_{t_0}^{\perp}$ are finite dimensional subspaces, we can write these projection operators in matrix forms, which can be used in computing $\|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\|$. In the special case where $\mathbf{X}_n = \mathbf{L}_n$, these matrices are related to the correlation matrix of the input, and its norm is related to the spectrum property of $u$.

Let $\mathbf{X}_n = \overline{sp}\{f_1, f_2, \cdots, f_n\}$, where $\{f_1, \cdots, f_n\}$ is an orthonormal basis in the $\ell^2[0,\infty)$ sense. Let $\{f_i\}_{i=1}^{\infty}$ be an orthonormal basis of $\ell^2[0,\infty)$ obtained by extending $\{f_1, \ldots, f_n\}$. A system $k \in \mathbf{L}$ will be represented by $k = \sum_{i=1}^{\infty} \kappa_i f_i$.

Set $\alpha_i(t) := \mathbf{\Phi}_u(f_i)(t)$. Then the operator $\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u$ has the following matrix representation with respect to the basis $\{f_i\}$,

$$\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u \sim A_{t_0} = \begin{pmatrix} \alpha_1(t_0) & \alpha_2(t_0) & \cdots \\ \alpha_1(t_0+1) & \alpha_2(t_0+1) & \cdots \\ \vdots & \vdots & \vdots \\ \alpha_1(t_0+T-1) & \alpha_2(t_0+T-1) & \cdots \end{pmatrix}. \tag{4.16}$$

We denote the first $n$ columns of $A_{t_0}$ by $B_{t_0}$, i.e.,

$$B_{t_0} = \begin{pmatrix} \alpha_1(t_0) & \alpha_2(t_0) & \cdots & \alpha_n(t_0) \\ \alpha_1(t_0+1) & \alpha_2(t_0+1) & \cdots & \alpha_n(t_0+1) \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1(t_0+T-1) & \alpha_2(t_0+T-1) & \cdots & \alpha_n(t_0+T-1) \end{pmatrix}, \tag{4.17}$$

and the rest of $A_{t_0}$ by $C_{t_0}$. Then $A_{t_0} = [B_{t_0}, C_{t_0}]$.

**Proposition 4.3** *Let $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\ell^2}$ and $\mathbf{S}_a = \mathbf{X}_n + \mathbf{B}(\epsilon)$. If $u \in \ell^2(-\infty,\infty)$ and $A_{t_0}$ is full rank, then $\delta_0^T(u) = \infty$ when $B_{t_0}(n)$ is not full rank, otherwise,*

$$\epsilon\frac{\underline{\sigma}(A_{t_0})}{\underline{\sigma}(B_{t_0})} \leq \delta_0^T(u) \leq \epsilon\frac{\bar{\sigma}(A_{t_0})}{\underline{\sigma}(B_{t_0})}, \tag{4.18}$$

*where $\underline{\sigma}$ and $\bar{\sigma}$ denote the largest and the smallest singular values respectively.*

**Proof** First, we show that the null space $\mathbf{N}_{t_0}$ is closed in $\ell^2[0,\infty)$. Suppose $\{k_i\}$ is a convergent sequence in $\mathbf{N}_{t_0}$. Assume $\hat{k}$ is its limit, i.e., $\lim_{i\to\infty}\|k_i-\hat{k}\|_2=0$. An input $u\in\ell^2[0,\infty)$ defines a continuous linear functional on $\ell^2$ through the summation $\sum_{\tau=0}^{\infty}k(\tau)u(t-\tau)$ for all $t$. Therefore,

$$\sum_{\tau=0}^{\infty}\hat{k}(\tau)u(t-\tau)=\lim_{i\to\infty}\sum_{\tau=0}^{\infty}k_i(\tau)u(t-\tau)=0\ \forall t\in[t_0,t_0+T), \qquad (4.19)$$

which implies that $\hat{k}\in\mathbf{N}_{t_0}$.

By Proposition 4.1,

$$\delta_0^T(u)=\epsilon\left(1-\|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\|^2\right)^{-1/2} \qquad (4.20)$$

Next, we estimate $\|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\|$. The projection operator $\mathbf{P}_{\mathbf{X}_n}$ has matrix representation

$$\mathbf{P}_{\mathbf{X}_n}\ \sim\ \begin{pmatrix} I_{n\times n} & 0\\ 0 & 0\end{pmatrix},$$

with respect to the basis $\{f_i\}$. Since $u\in\ell^2$, the sequences $\{\alpha_i(t)\}_{i=1}^{\infty}$ are in $\ell^2$. Hence, the projection operator $\mathbf{P}_{\mathbf{N}_{t_0}}$ has matrix representation

$$\mathbf{P}_{\mathbf{N}_{t_0}}\ \sim\ I-A_{t_0}^T(A_{t_0}A_{t_0}^T)^{-1}A_{t_0}.$$

Since $A_{t_0}=[B_{t_0},C_{t_0}]$, we have

$$\mathbf{P}_{\mathbf{N}_{t_0}}\ \sim\ I-\begin{pmatrix} B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} & B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}C_{t_0}\\ C_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} & C_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}C_{t_0}\end{pmatrix}.$$

Hence

$$\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\ \sim\ \begin{pmatrix} I_{n\times n}-B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} & -B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}C_{t_0}\\ 0 & 0\end{pmatrix},$$

and

$$\left(\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\right)\left(\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\right)^*$$
$$\sim \begin{pmatrix} \left(I_{n\times n} - B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}\right)^2 + B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}C_{t_0}C_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} & 0 \\ 0 & 0 \end{pmatrix}.$$

Since

$$\begin{aligned}
&\left(I_{n\times n} - B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}\right)^2 + B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}C_{t_0}C_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} \\
&= I_{n\times n} - 2B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} + B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} \\
&\quad + B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}C_{t_0}C_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}, \\
&= I_{n\times n} + B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}\left(-2\left(A_{t_0}A_{t_0}^T\right) + B_{t_0}B_{t_0}^T + C_{t_0}C_{t_0}^T\right)\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}, \\
&= I_{n\times n} - B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}, \\
&\quad (\text{note that } \left(\mathrm{A}_{t_0}\mathrm{A}_{t_0}^{\mathrm{T}}\right) = \mathrm{B}_{t_0}\mathrm{B}_{t_0}^{\mathrm{T}} + \mathrm{C}_{t_0}\mathrm{C}_{t_0}^{\mathrm{T}})
\end{aligned}$$

we have

$$\left(\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\right)\left(\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\right)^* \sim \begin{pmatrix} I_{n\times n} - B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0} & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore,

$$\begin{aligned}
1 - \|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\|^2 &= 1 - \|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\left(\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\right)^*\|, \\
&= 1 - \bar{\sigma}\left(I_{n\times n} - B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}\right), \\
&= \underline{\sigma}\left(B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}\right).
\end{aligned}$$

Since $\|\mathbf{P}_{\mathbf{X}_n}\mathbf{P}_{\mathbf{N}_{t_0}}\|^2 = 1$ when $\underline{\sigma}\left(B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}\right) = 0$, $\delta_0^T(u) = \infty$ if $B_{t_0}$ is not full rank. It is trivial to check that

$$\frac{\underline{\sigma}(B_{t_0}B_{t_0}^T)}{\bar{\sigma}\left(A_{t_0}A_{t_0}^T\right)} \leq \underline{\sigma}\left(B_{t_0}^T\left(A_{t_0}A_{t_0}^T\right)^{-1}B_{t_0}\right) \leq \frac{\underline{\sigma}(B_{t_0}B_{t_0}^T)}{\underline{\sigma}\left(A_{t_0}A_{t_0}^T\right)}.$$

This completes the proof.

□

When $\mathbf{X}_n = \mathbf{L}_n = sp\{1, z, \ldots, z^{n-1}\}$ and $\{f_i\} = \{z^i\}$, the operator $\mathbf{P}_{[t_0,t_0+T)}\Phi_u$ has the following matrix representation,

$$M_{t_0} = \begin{pmatrix} u(t_0) & u(t_0-1) & \cdots \\ u(t_0+1) & u(t_0) & \cdots \\ \vdots & \vdots & \vdots \\ u(t_0+T-1) & u(t_0+T-2) & \cdots \end{pmatrix}. \tag{4.21}$$

We denote the first $n$ columns of $M_{u,t_0}$ by $U_{t_0}(n)$, i.e.,

$$U_{t_0}(n) = \begin{pmatrix} u(t_0) & u(t_0-1) & \cdots & u(t_0-n+1) \\ u(t_0+1) & u(t_0) & \cdots & u(t_0-n+2) \\ \vdots & \vdots & \vdots & \vdots \\ u(t_0+T-1) & u(t_0+T-2) & \cdots & u(t_0+T-n) \end{pmatrix} \tag{4.22}$$

Set $\Psi_{u,t_0} = M_{u,t_0}M_{u,t_0}^T$ and $\Phi_{u,t_0}(n) = U_{t_0}(n)^T U_{t_0}(n)$. Then we have the following corollary to Proposition 4.3

**Corollary 4.1** *Under the conditions of Proposition 4.3, if $\mathbf{X}_n = \mathbf{L}_n$, then*

$$c\left(\frac{\underline{\sigma}(\Psi_{u,t_0})}{\underline{\sigma}(\Phi_{u,t_0}(n))}\right)^{1/2} \leq \delta_0^T(u) \leq c\left(\frac{\bar{\sigma}(\Psi_{u,t_0})}{\underline{\sigma}(\Phi_{u,t_0}(n))}\right)^{1/2}, \tag{4.23}$$

When an input is not in $\ell^2$, the null space $\mathbf{N}_{t_0}$ may not be closed in $\ell^2$. Hence, the characterization in terms of projection operators may not hold. However, a similar relation between the inherent error and the ratio of the largest and the smallest singular values of $\Phi_u$ still holds. For this, we introduce the notion of

## 4.3 Spectrum Flatness and Mixing Rate

**Definition 4.2** *Let $f(n)$, $g(n) \in \ell^\infty(-\infty,\infty)$. The T-crosscorrelation of $f$ and $g$ is a function $\phi_{fg}(\tau,t) : [1-T, T-1] \times \mathbb{Z} \longrightarrow \mathbb{R}$ defined by*

$$\phi_{fg}(\tau,t) := \sum_{i=0}^{T-1} f(t+i)g(t+\tau+i). \tag{4.24}$$

*The T-autocorrelation of $f$ is defined as $\phi_{ff}(\tau,t)$.*

For a given input we form the T-autocorrelation matrix

$$\Phi_{u,t_0}(n) = U_{t_0}(n)^T U_{t_0}(n), \tag{4.25}$$

$$= \begin{pmatrix} \phi_{uu}(0,t_0) & \phi_{uu}(-1,t_0) & \cdots & \phi_{uu}(-n+1,t_0) \\ \phi_{uu}(-1,t_0) & \phi_{uu}(0,t_0-1) & \cdots & \phi_{uu}(-n+2,t_0-1) \\ \vdots & \vdots & \cdots & \vdots \\ \phi_{uu}(-n+1,t_0) & \cdots & \cdots & \phi_{uu}(0,t_0-n+1) \end{pmatrix}. \tag{4.26}$$

**Definition 4.3** *If $u \in \ell^\infty(-\infty,\infty)$ and $\inf_{t_0} \underline{\sigma}(\Phi_{u,t_0}(n)) \neq 0$, we define the n-th degree spectrum flatness of $u$ by*

$$\nu_u(n,T) = \left(\frac{\sup_{t_0} \bar{\sigma}(\Phi_{u,t_0}(n))}{\inf_{t_0} \underline{\sigma}(\Phi_{u,t_0}(n))}\right)^{1/2}, \tag{4.27}$$

*otherwise, $\nu_u(n,T) = \infty$.*

We call $\nu_u(n,T)$ spectrum flatness for two reasons. a) For periodic signals $\nu_u(T,T)$ equals the ratio of the largest and smallest spectral values of the signal. b) For "stationary" signals, $\nu_u(n,T)$ also equals the ratio of the largest and smallest spectral values when $T$ and $n$ tends to infinity. These are shown in Appendix B. The short-time spectrum and local correlation are well understood subjects in signal processing and spectrum estimation.

For stochastic processes, spectrum flatness reflects their "regularity", which is related to the concept "mixing" [5, 9]. Here, we are going to introducing a quantity which measures the regularity of a deterministic signal and call it the *mixing rate*.

**Definition 4.4** *Let $f \in \ell^\infty(-\infty,\infty)$. If*

$$\inf_{t\in\mathbf{Z}}\left(\phi_{ff}(0,t) - \max_{0\le\tau\le T-1}\sum_{i=\tau-T+1,i\neq 0}^{\tau}|\phi_{ff}(i,t)|\right) > 0, \tag{4.28}$$

*then we say $f$ is T-mixing. The quantity*

$$\gamma_f(n,T) := \inf_{t\in\mathbf{Z}}\left(\phi_{ff}(0,t) - \max_{0\le\tau\le n-1}\sum_{i=\tau-n+1,i\neq 0}^{\tau}|\phi_{ff}(i,t)|\right)^{\frac{1}{2}}, \tag{4.29}$$

*where $0 < n \le T$, is called the n-th degree T-mixing rate.*

**Example 4.1** A sequence of impulses with period T is T-mixing. Let $u(t) = \sum_{k=-\infty}^{\infty}\delta(t-kT)$. Obviously $\phi_{uu}(\tau) = \delta(\tau)/T$. It follows that $\gamma_u(n,T) = 1/T > 0$ for all $n \le T$.

**Example 4.2** A pseudo-random binary sequence $u(t)$ with period T is T-mixing. It has been shown that $\phi_{uu}(0) = 1$ and $\phi_{uu}(\tau) = 1/T$ if $\tau \neq 0$. It follows that $\gamma_u(n,T) = (T+1-n)/T$. This implies that the PRBS has a larger mixing rate than the sequence of impulses if $n < T$.

The mixing rate is related to the spectrum flatness as follows.

**Proposition 4.4** *If $u$ is a T-mixing function, then*

$$\nu_u^2(n,T) \le \frac{2\sup_t \phi_{uu}(0,t)}{\gamma_u^2(n,T)} - 1. \tag{4.30}$$

**Proof** Consider the matrix $\Phi_{u,t}(n)$. By Gershgorin's theorem, we have

$$\inf_t \underline{\sigma}(\Phi_{u,t}(n)) \ge \gamma_u^2(n,T),$$

and

$$\begin{aligned}\sup_t \bar{\sigma}(\Phi_{u,t}(n)) &\le \sup_t(\phi_{uu}(0,t) + \max_{0\le\tau\le n-1}\sum_{i=\tau-n+1,i\neq 0}^{\tau}|\phi_{uu}(i,t)|), \\ &= 2\sup_t \phi_{uu}(0,t) - \gamma_u^2(n,T).\end{aligned}$$

The relation in (4.30) follows.

□

## 4.4 Spectrum Flatness and the Inherent Error

Input design in terms of the usual spectral properties has been studied in the past e.g., [31, 10]. The input is usually designed to achieve certain asymptotic identification criteria. Here, the objective of input design is to achieve fast identification and the emphasis is on the effects of unmodeled dynamics.

In this section, we show that spectrum flatness can be used as a criterion for input design. Spectrum flatness is an ensemble property of the inputs. There is a set of inputs all having the same spectrum flatness. To characterize the identifying capability of the inputs by spectrum flatness, we consider the worst-case of inputs in that set.

### 4.4.1 The $\ell^2$ Case

**Theorem 4.2** *Let* $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\ell^2}$. *If* $\mathbf{S}_n = \mathbf{L}_n + \mathbf{B}(\epsilon)$, *where* $\mathbf{B}(\epsilon)$ *is the closed* $\epsilon$*-ball in* $\ell^1$, *then*

$$\sup_{\nu_u(n,T)=\nu} \delta_0^T(u) = \sup_{\nu_u(n,T)=\nu} \bar{\delta}_0^T(u) = \epsilon\sqrt{1+\nu^2}. \tag{4.31}$$

**Remark** This theorem indicates that if we can modify the input so that its spectrum flatness is close to 1, then the inherent error is approximately $\sqrt{2}$ times of the representation error. On the other hand, if the spectrum flatness of the input is large, we may not be able to identify the system well. In the worst-case, the inherent error is $\sqrt{1+\nu^2}$ times of the representation error. Since $\|\cdot\|_{\ell^1} \geq \|\cdot\|_{\mathbf{H}^\infty} \geq \|\cdot\|_{\ell^2}$, this theorem also gives a lower bound on the inherent error in the $\mathbf{H}^\infty$ and $\ell^1$ norms.

**Proof of Theorem 4.2** Fix $t_0$. Let $k \in \mathbf{N}_{t_0}$. Then

$$U_{t_0}(n)(\mathbf{P}_{[0,n-1]}(k)) + \sum_{i=1}^{\infty} U_{t_0-in}(n)(\mathbf{P}_{[in,(i+1)n]}(k)) = 0. \tag{4.32}$$

This implies that

$$\begin{aligned}
\|U_{t_0}(n)(\mathbf{P}_{[0,n-1]}(k))\|_2 &\le \sum_{i=1}^{\infty}\|U_{t_0-in}(n)(\mathbf{P}_{[in,(i+1)n]}(k))\|_2,\\
&\le \sup_t\sqrt{\bar\sigma(\Phi_{u,t}(n))}\sum_{i=1}^{\infty}\|\mathbf{P}_{[in,(i+1)n]}(k)\|_2,\\
&\le \sup_t\sqrt{\bar\sigma(\Phi_{u,t}(n))}\sum_{i=1}^{\infty}\|\mathbf{P}_{[in,(i+1)n]}(k)\|_1,\\
&= \sup_t\sqrt{\bar\sigma(\Phi_{u,t}(n))}\|\mathbf{P}_{[n,\infty)}(k)\|_1.
\end{aligned}$$

Since $k\in\mathbf{S}_a$ implies that $\|\mathbf{P}_{[n,\infty)}(k)\|_1\le\epsilon$, we have

$$\|\mathbf{P}_{[0,n-1]}(k)\|_2\le\epsilon\left(\frac{\sup_t\bar\sigma(\Phi_{u,t}(n))}{\underline{\sigma}(\Phi_{u,t_0}(n))}\right)^{1/2}\le\epsilon\nu_u(n,T). \tag{4.33}$$

It follows that $\|k\|_2=(\|\mathbf{P}_{[0,n-1]}(k)\|_2^2+\|\mathbf{P}_{[n,\infty)}(k)\|_2^2)^{1/2}\le\epsilon\sqrt{1+\nu_u(n,T)^2}$. Noticing that the above formula is independent of $t_0$, we get the upper bound in (4.31).

For the lower bound, consider the input

$$u(t)=\begin{cases}1 & t=t_0+iT,\ i\in\mathbb{Z}/\{0\},\\ 1/\nu & t=t_0,\\ 0 & otherwise.\end{cases} \tag{4.34}$$

It is easy to check that $\nu_u(n,T)=\nu$. Put

$$k(\tau)=\begin{cases}\nu\epsilon & \tau=0,\\ -\epsilon & \tau=T,\\ 0 & otherwise.\end{cases} \tag{4.35}$$

The fact that $k\in\mathbf{N}_{t_0}\bigcap\mathbf{S}_a$ implies that

$$\delta_0^T(u)\ge\|k\|_2=\epsilon\sqrt{1+a^2}. \tag{4.36}$$

This completes the proof.

□

### 4.4.2 The $\mathbf{H}^\infty$ and $\ell^1$ cases

Now we study the inherent error for some more realistic a priori uncertainty sets in the $\mathbf{H}^\infty$ and $\ell^1$ norms. For this we will need the following two lemmas relating the $\mathbf{H}^\infty$ and $\ell^1$ norms to the $\ell^2$ norm. They indicate that if the frequency response of a system is smooth or its impulse response decays exponentially, then its $\mathbf{H}^\infty$ or $\ell^1$ norms are not too much larger than its $\ell^2$ norm.

**Lemma 4.1** *If $K'(z) \in \mathbf{H}^\infty(D)$, then $K(z) \in \mathbf{H}^p(D)$ $\forall p \in [1,\infty]$ and*

$$\|K(z)\|_\infty \leq \left(\frac{\|K'\|_\infty}{2(p+1)}\right)^{\frac{1}{p+1}} \|K\|_p^{\frac{p}{p+1}} + \left(\frac{1}{2\pi}\right)^{\frac{1}{p}} \|K\|_p. \tag{4.37}$$

**Proof** By Hardy's inequality, $K'(z) \in \mathbf{H}^\infty(D)$ implies that $K(z) = \sum_{n=0}^\infty a_n z^n$ and $\sum_{n=0}^\infty |a_n| < \infty$. It follows that $K(z) \in \mathbf{H}^p(D)$ $\forall p \in [1,\infty]$ and the nontangential limit $K(e^{i\theta})$ of $K(z)$ exists for all $\theta \in [-\pi,\pi]$ and $K(e^{i\theta}) = \lim_{r\to 1} K(re^{i\theta})$.

It has been shown in the proof of Proposition 3.3 part (iii) that $K(e^{i\theta})$ is Lipschitz continuous with Lipschitz constant $\|K'\|_\infty$. Therefore, $K(e^{i\theta})$ achieves its supremum and infimum. Without loss of generality, we assume $|K(e^{i0})| = |K(1)| = \|K\|_\infty$ and $|K(e^{i\theta_1})| = \inf_\theta |K(e^{i\theta})|$, where $|\theta_1| \leq \pi$. Put $\Delta\theta = (|K(1)| - |K(e^{i\theta_1})|)/\|K'\|_\infty$. Since $|K(e^{i0}) - K(e^{i\theta_1})| \leq \|K'\|_\infty |\theta_1|$, by Lipschitz continuity, $\Delta\theta \leq |\theta_1| \leq \pi$. Also, Lipschitz continuity of $K(e^{i\theta})$ implies that $\forall\theta \in [-\pi,\pi]$, $|K(e^{i\theta})| \geq \|K\|_\infty - \|K'\|_\infty|\theta|$. Therefore,

$$\begin{aligned}
\int_{-\Delta\theta}^{\Delta\theta} |K(e^{i\theta})|^p d\theta &\geq \int_{-\Delta\theta}^{\Delta\theta} (\|K\|_\infty - \|K'\|_\infty|\theta|)^p d\theta, && (4.38)\\
&= 2\int_0^{\Delta\theta} (\|K\|_\infty - \|K'\|_\infty\theta)^p d\theta, && (4.39)\\
&= \frac{2}{(p+1)\|K'\|_\infty}\left(\|K\|_\infty^{p+1} - (\|K\|_\infty - \|K'\|_\infty\Delta\theta)^{p+1}\right), && (4.40)\\
&= \frac{2}{(p+1)\|K'\|_\infty}\left(\|K\|_\infty^{p+1} - |K(e^{i\theta_1})|^{p+1}\right). && (4.41)
\end{aligned}$$

This implies that

$$\|K\|_p^p \geq \frac{2}{(p+1)\|K'\|_\infty}\left(\|K\|_\infty^{p+1} - (1/2\pi)^{\frac{p+1}{p}}\|K\|_p^{p+1}\right). \tag{4.42}$$

It follows that

$$\|K\|_\infty^{p+1} \leq (p+1)\|K'\|_\infty/2\|K\|_p^p + (1/2\pi)^{\frac{p+1}{p}}\|K\|_p^{p+1}. \tag{4.43}$$

Noticing that $(a+b)^{1/p} \leq a^{1/p} + b^{1/p}$, $\forall a > 0,\ b > 0,\ p > 1$, we complete the proof.

□

**Lemma 4.2** *Let* $p \geq 1,\ 0 < r < 1$. *If* $k \in \ell^p[0,\infty)$ *and* $|k(\tau)| \leq Cr^\tau\ \forall \tau \in [0,\infty)$, *then*

$$\|k\|_1 \leq \left(\frac{C}{e(1-r)\ln r^{-1}}\right)^{1-1/p}\|k\|_p^{1/p} + 2\|k\|_p. \tag{4.44}$$

**Proof** Since $\|k\|_p \leq (\sum_{\tau=0}^{\infty}(Cr^\tau)^p)^{1/p} < C/(1-r)$, there exists an integer $N$ such that

$$Cr^N/(1-r) \leq \|k\|_p \leq Cr^{(N-1)}/(1-r). \tag{4.45}$$

Hence,

$$\begin{aligned}
\|k\|_1 &= \sum_{\tau=0}^{N-1}|k(\tau)| + \sum_{\tau=N}^{\infty}|k(\tau)|,\\
&\leq (N)^{1-1/p}\left(\sum_{\tau=0}^{N-1}|k(\tau)|^p\right)^{1/p} + Cr^N/(1-r),\\
&\leq N^{1-1/p}\|k\|_p + \|k\|_p, \quad \text{(by (4.45),)}\\
&\leq (N-1)^{1-1/p}\|k\|_p + 2\|k\|_p,\\
&= ((N-1)\|k\|_p)^{1-1/p}\|k\|_p^{1/p} + 2\|k\|_p,\\
&\leq \left((N-1)r^{N-1}\frac{C}{1-r}\right)^{1-1/p}\|k\|_p^{1/p} + 2\|k\|_p, \quad \text{(by (4.45),)}\\
&\leq \left(\frac{C}{e(1-r)\ln r^{-1}}\right)^{1-1/p}\|k\|_p^{1/p} + 2\|k\|_p, \quad \left((N-1)r^{N-1} \leq (e\ln r^{-1})^{-1}\right).
\end{aligned}$$

□

**Proposition 4.5** *Let* $\mathbf{S}_a = \mathbf{S}_{a,1}$ *as in part (i) of Example 2.1 with* $C = 1$, *and* $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\ell^1}$. *Then*

$$\min\left\{(1+\nu)d^T, \left(\frac{1-r^T}{1-r}\right)\right\} \le \sup_{\nu_u(n,T)=\nu} \delta_0^T(u) \le \sup_{\nu_u(n,T)=\nu} \delta_0^T(u) \quad (4.46)$$

$$\le \rho(1+\nu^2)^{1/4} d_n^{1/2} + 2(1+\nu^2)^{1/2} d_n, \quad (4.47)$$

*where* $\rho = (\frac{1}{e(1-r)\ln r^{-1}})^{1/2}$, $d^T = d^T(\mathbf{S}_{a,1}, \ell^1) = r^T/(1-r)$, *and* $d_n = d_n(\mathbf{S}_{a,1}, \ell^1) = r^n/(1-r)$.

**Proof** Fix $t_0$. It has been shown in Corollary 3.7 that $\mathbf{S}_{a,1} \subset \mathbf{L}_n + \mathbf{B}(d_n)$. By Theorem 4.2, $\forall k \in \mathbf{S}_{a,1} \bigcap \mathbf{N}_{t_0}$,

$$\|k\|_{\ell^2} \le d_n\sqrt{1+\nu^2}. \quad (4.48)$$

Since $k \in \mathbf{S}_{a,1}$, by Lemma 4.2,

$$\|k\|_{\ell^1} \le \rho d_n^{1/2}(1+\nu^2)^{1/4} + 2d_n(1+\nu^2)^{1/2}. \quad (4.49)$$

Since the above inequality holds for all $k \in \mathbf{S}_{a,1} \bigcap \mathbf{N}_{t_0}$ and $t_0$, we get the upper bound (4.47).

To show the lower bound, we consider the same input as in (4.34) whose spectrum flatness is $\nu$. For such an input, $k \in \mathbf{N}_{t_0}$ if and only if

$$k(t) = -\nu \sum_{\tau=1}^{\infty} k(t+\tau T) \quad \forall t \in [0, T-1].$$

We will construct a system which is in $\mathbf{S}_{a,1} \bigcap \mathbf{N}_{t_0}$ in each of the following two cases. (i) If $1 - \nu\left(\frac{r^T}{1-r^T}\right) \ge 0$, then put

$$k(\tau) = \begin{cases} -\nu r^\tau \left(\frac{r^T}{1-r^T}\right) & \tau \le T-1, \\ r^\tau & \tau > T-1. \end{cases}$$

In this case we have

$$\begin{aligned} \|k\|_1 &= \nu\left(\frac{r^T}{1-r^T}\right)\left(\frac{1-r^T}{1-r}\right) + \frac{r^T}{1-r}, \\ &= d^T(1+\nu) \end{aligned}$$

(ii) If $1 - \nu\left(\frac{r^T}{1-r^T}\right) < 0$, put

$$k(\tau) = \begin{cases} r^\tau & \tau \leq T-1, \\ -\frac{1-r^T}{\nu r^T} r^\tau & \tau > T-1. \end{cases}$$

In this case we have $\|k\|_1 \geq \left(\frac{1-r^T}{1-r}\right)$. It is easy to check that in both cases $k \in \mathbf{S}_{a,1} \bigcap Null\left(\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u\right)$. Combining (i) and (ii), we get the lower bound (4.46).

□

**Proposition 4.6** *Let* $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{H}^\infty}$ *and*

$$\mathbf{S}_a := \left\{k \in \ell^1[0,\infty) : \quad \|\mathbf{P}_{[\tau,\infty)}(k)\|_1 \leq 1/\tau \ \forall \tau > 0, \ \textit{and} \ \|K'\|_\infty \leq 1\right\}, \tag{4.50}$$

*where $K'$ is the derivative of the z-transform of $k$. Then*

$$\frac{\nu+1}{\nu+T} \leq \sup_{\nu_u(n,T)=\nu} \delta_0^T(u) \leq \sup_{\nu_u(n,T)=\nu} \bar{\delta}_0^T(u) \leq \left(\frac{1+\nu^2}{6n^2}\right)^{\frac{1}{3}} + \left(\frac{1+\nu^2}{2\pi n^2}\right)^{\frac{1}{2}}. \tag{4.51}$$

**Proof** Fix $t_0$. It has been shown in Corollary 3.9 that $\mathbf{S}_a \subset \mathbf{L}_n + \mathbf{B}(1/n)$. By Theorem 4.2, $\forall k \in \mathbf{S}_a \bigcap \mathbf{N}_{t_0}$,

$$\|k\|_{\ell^2} \leq 1/n\sqrt{1+\nu^2}. \tag{4.52}$$

Since $k \in \mathbf{S}_a$, by Lemma 4.1,

$$\|k\|_{\mathbf{H}^\infty} \leq \left(\frac{1+\nu^2}{6n^2}\right)^{\frac{1}{3}} + \left(\frac{1+\nu^2}{2\pi n^2}\right)^{\frac{1}{2}}. \tag{4.53}$$

Since the above inequality holds for all $k \in \mathbf{S}_a \bigcap \mathbf{N}_{t_0}$ and $t_0$, we get the upper bound in (4.51).

To show the lower bound, we consider the same input as in (4.34) whose spectrum flatness is $\nu$. Put

$$k(\tau) = \begin{cases} \frac{-\nu}{\nu+T} & \tau = 0, \\ \frac{1}{\nu+T} & \tau = T, \\ 0 & otherwise. \end{cases}$$

It is easy to check that $k \in \mathbf{S}_u \bigcap \mathbf{N}_{t_0}$ and $\|k\|_{\mathbf{H}^\infty} = \frac{\nu+1}{\nu+T}$. This gives the lower bound in (4.51).

□

In Proposition 4.5 and Proposition 4.6, the inherent errors have upper and lower bounds increasing with the spectrum flatness of the input.

### 4.4.3 The Case of Finite Power Disturbances

In this section, we consider the inherent error when the disturbance has finite $\ell^2$ norm on an interval of fixed length. For this, we assume that

$$\mathbf{V} = \{v \in \ell^\infty(-\infty,\infty): \quad \|\mathbf{P}_{t_0,t_0+T}(v)\|_2 \leq \eta\}. \tag{4.54}$$

The next proposition shows the relation between the spectrum flatness of the input and $\delta^T(u)$ in the case when $\mathbf{L} = \ell^2$ and $\mathbf{S}_u = \mathbf{X}_n + \mathbf{B}(\epsilon)$, where $\mathbf{B}(\epsilon)$ is the closed $\epsilon$-ball in $\ell^1$. The other cases can be delt similarly.

**Proposition 4.7** *Under the conditions of Theorem 4.2, given the ensemble of inputs* $\mathbf{U}(\nu,\gamma) = \{u \in \ell^\infty(-\infty,\infty): \ \nu_u(n,T) = \nu, \gamma_u(n,T) = \gamma\}$, *the worst-case inherent error achievable with* $\mathbf{U}(\nu,\gamma)$ *is*

$$\sup_{u\in\mathbf{U}(\nu,\gamma)} \bar{\delta}^T(u) = \left((\eta/\gamma + \nu\epsilon)^2 + \epsilon^2\right)^{1/2}. \tag{4.55}$$

**Proof** Fix $t_0$. By definition of $\delta^T(u)$,

$$\delta^T(u) = \|\mathbf{S}_u \bigcap \mathbf{S}_{e,t_0}\|_2,$$

where

$$\mathbf{S}_{e,t_0} := \{k \in \mathbf{L}: \ \|\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u(k)\|_2 \leq \eta\}.$$

For $k \in \mathbf{S}_u \bigcap \mathbf{S}_{\epsilon,t_0}$, by a similar argument to that in the proof of Theorem 4.2, it can be shown that

$$U_{t_0}(n)(\mathbf{P}_{[0,n-1]}(k)) + M_{u,t_0-n}(\mathbf{P}_{[n,\infty)}(k)) = \mathbf{P}_{[t_0,t_0+T)}(v) \tag{4.56}$$

for some $v \in \mathbf{V}$. It follows that

$$\|U_{t_0}(n)(\mathbf{P}_{[0,n-1]}(k))\|_2 \le \sup_t \sqrt{\bar{\sigma}(\Phi_{u,t}(n))}\|\mathbf{P}_{[n,\infty)}(k)\|_2 + \eta. \tag{4.57}$$

Hence,

$$\|\mathbf{P}_{[0,n-1]}(k)\|_2 \le \epsilon\nu_u(n,T) + \eta/\sqrt{\underline{\sigma}(\Phi_{u,t}(n))} \le \epsilon\nu_u(n,T) + \eta/\gamma, \tag{4.58}$$

as $\sqrt{\underline{\sigma}(\Phi_{u,t}(n))} \ge \gamma_u(n,T)$. This yields the upper bound.

For the lower bound, consider the input

$$u(t) = \begin{cases} \nu\gamma & t = iT,\ i \in \mathbb{Z}/\{0\}, \\ \gamma & t = 0, \\ 0 & otherwise. \end{cases} \tag{4.59}$$

It is easy to check that $\nu_u(n,T) = \nu$ and $\gamma_u(n,T) = \gamma$. Put

$$k(\tau) = \begin{cases} \eta/\gamma + \nu\epsilon & \tau = 0, \\ -\epsilon & \tau = T, \\ 0 & otherwise. \end{cases} \tag{4.60}$$

The fact that $k \in \mathbf{S}_u \bigcap \mathbf{S}_{\epsilon,T-1}$ implies that

$$\delta^T(u) \ge \|k\|_2 = \left((\eta/\gamma + \nu\epsilon)^2 + \epsilon^2\right)^{1/2}. \tag{4.61}$$

This completes the proof.

□

**Remark** The uncertainty caused by the additive noise can be reduced by increasing the magnitude of the input, while the one caused by multiplicative uncertainty can not.

It should be noted that the noise energy in general increases with the length of the observation interval $T$. However, if the noise can be averaged out, a long observation interval can be decomposed into several shorter intervals. Then worst-case identification can be performed on the basis of ensemble averaged observations. Although it is not realistic to assume noise can be vanquished by averaging on a short interval, additive noise does have this property on a long interval in practical problems [26].

# Chapter 5

# Several Identification Algorithms

In this Chapter, we propose several algorithms for estimation of the modeled dynamics. These estimates will be used as nominal approximations to the true system. The estimation problem is first formulated as a convex optimization problem. Then two robust algorithms based on the analytic center are given. Another result involves an algorithm for impulse response estimation, for which an $\ell^1$ error bound is given.

## 5.1 From Worst-Case Identification to Convex Optimization

As in the previous chapter, we first assume the a priori uncertainty set has the form of a finite dimensional subspace plus an $\epsilon$-ball in $\mathbf{L}$, i.e., $\mathbf{S}_a = \mathbf{X}_n + \mathbf{B}(\epsilon)$, with $\mathbf{X}_n$ and $\mathbf{B}(\epsilon)$ representing the modeled and ummodeled dynamics respectively. Then, the true system $k_{true} \in \mathbf{S}_a$ has a decomposition $k = k_1 + k_2$, where $k_1 \in \mathbf{X}_n$ and $k_2 \in \mathbf{B}(\epsilon)$. The objective of identification is to estimate the modeled dynamics $k_1$ from the output observations on a finite time interval, $[t_0, t_0 + T)$. Given an input $u$, on the basis of these observations, the location of $k_1$ is narrowed down from $\mathbf{X}_n$ to a smaller set

$$\mathbf{S}_1(y) := \{k_1 \in \mathbf{X}_n : \mathbf{P}_{[t_0,t_0+T)}(\Phi_u(k_1) - y) = \mathbf{P}_{[t_0,t_0+T)}(v + \Phi_u(k_2))$$
$$\textit{for some } v \in \mathbf{V} \textit{ and } k_2 \in \mathbf{B}(\epsilon)\} . \tag{5.1}$$

The worst-case error in an estimate $k_{est}$ is

$$\mathcal{J}(k_{est}) := \sup_{k \in \mathbf{S}_1(y)} \|k_{est} - k\|_{\mathbf{L}}. \tag{5.2}$$

$\mathcal{J}(k_{est})$ gives a criterion for nominal system selection. If the estimate is to minimize the worst-case identification error, the identification problem is equivalent to the optimization problem

$$\min_{k_{est} \in \mathbf{X}_n} \mathcal{J}(k_{est}). \tag{5.3}$$

The optimal estimate with respect to $\mathcal{J}$ is called the *Chebechev Center*[1] of $\mathbf{S}_1(y)$. For a fixed $k$, $\|k_{est} - k\|_{\mathbf{L}}$ is a convex function in $k_{est}$. As it is the supremum of a class of convex functions, $\mathcal{J}$ is itself a convex function in $k_{est}$. Therefore, the optimization problem (5.3) is a problem of minimizing a convex function on a linear subspace.

Instead of the Chebechev center of $\mathbf{S}_1(y)$, one could choose other kind of centers as nominals, e.g., the analytic center (which will be defined later), which may have some advantages over the Chebechev center. Moreover, it is trivial to show that if the nominal is chosen in $\mathbf{S}_1(y)$, then the worst-case estimation error $\mathcal{J}(k_{est})$ is within a factor of two of the optimal one.

**Proposition 5.1** *If $k_{est} \in \mathbf{S}_1(y)$, then*

$$\mathcal{J}(k_{est}) \leq 2 \min_{k_{est} \in \mathbf{X}_n} \mathcal{J}(k_{est}). \tag{5.4}$$

Although, problem (5.3) is a convex optimization problem which is easy to solve in theory, the computation might be quite complicated because the construction of

[1]The Chebechev center of a convex set in a normed space is the center of the smallest ball covering the set.

the convex function $\mathcal{J}$ can be computationally involved. However, in some special cases which are of interest in control, the computation can be done easily.

## 5.2 Two Algorithms Based on Convex Optimization

First, we examine two cases where the a posteriori uncertainty set $\mathbf{S}_1(y)$ is a polytope.

**Example 5.1** Let $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\ell^1}$, $\mathbf{X}_n = \mathbf{L}_n = sp\{1, z, \ldots, z^{n-1}\}$, and $\mathbf{B}(\epsilon)$ be the $\epsilon$-ball in $\ell^1$. The noise $v$ has $\ell^\infty$ bound $\eta$, i.e., $\mathbf{V} = \{v \in \ell^\infty : \|v\|_\infty \leq \eta\}$. The input is given on the interval $[t_0 - n + 1, t_0 + T)$, and the output is observed on the interval $[t_0, t_0 + T)$. The input past is unknown except being bounded by $C_u := \|u\|_\infty$. In this case, it can be shown that the a posteriori uncertainty set

$$\mathbf{S}_1(y) = \left\{ k \in \mathbf{L}_n : \left| \sum_{\tau=0}^{n-1} k(\tau)u(t-\tau) - y(t) \right| \leq \eta + \epsilon C_u \ \forall t \in [t_0, t_0 + T) \right\}. \tag{5.5}$$

To show $\mathbf{S}_1(y)$ is contained in the set on the right side of (5.5), we notice that, for $k_1 \in \mathbf{S}_1(y)$, $\exists v \in \mathbf{V}$ and $k_2 \in \mathbf{B}(\epsilon)$ such that

$$\sum_{\tau=0}^{n-1} k(\tau)u(t-\tau) - y(t) = v(t) + \sum_{\tau=0}^{\infty} k_2(\tau)u(t-\tau) \quad \forall t \in [t_0, t_0 + T). \tag{5.6}$$

This implies that

$$\left| \sum_{\tau=0}^{n-1} k(\tau)u(t-\tau) - y(t) \right| \leq \eta + \epsilon C_u \quad \forall t \in [t_0, t_0 + T). \tag{5.7}$$

On the other hand, if a system $k \in \mathbf{L}_n$ satisfies (5.7), then

$$\sum_{\tau=0}^{n-1} k(\tau)u(t-\tau) - y(t) = v(t) + y(t) \quad \forall t \in [t_0, t_0 + T). \tag{5.8}$$

for some $v(t) \in \mathbf{V}$ and $y_2(t)$ such that $\|y_2\|_\infty \leq \epsilon C_u$. Put

$$k_2(\tau) = \begin{cases} \epsilon & \tau = T + n - 1, \\ 0 & otherwise, \end{cases} \tag{5.9}$$

and the input past

$$u(t) = \begin{cases} y(t+T+n-1)/\epsilon & t \in [t_0 - T - n + 1, t_0 - n], \\ 0 & t < t_0 - T - n + 1. \end{cases} \tag{5.10}$$

Then $y_2(t) = \sum_{\tau=0}^{\infty} k_2(\tau) u(t-\tau)$ $\forall t \in [t_0, t_0 + T)$, i.e., $y_2$ is the output of a system in $\mathbf{B}(\epsilon)$ driven by an input whose $\ell^\infty$ norm is less than $C_u$. Therefore, $k$ is in $\mathbf{S}_1(y)$. This completes the proof for (5.5).

Put

$$\vec{k} = [k(0), \ldots, k(n-1)]^T \tag{5.11}$$

$$\vec{y} = [y(t_0), \ldots, y(t_0 + T - 1)]^T \tag{5.12}$$

$$\vec{v} = [v(t_0), \ldots, v(t_0 + T - 1)]^T \tag{5.13}$$

$$\vec{y}_2 = [y_2(t_0), \ldots, y_2(t_0 + T - 1)]^T \tag{5.14}$$

Let $U_{t_0}(n)$ be defined as in (4.22). Then, $\mathbf{S}_1(y)$ can be written as

$$\begin{aligned} \mathbf{S}_1(y) &= \Big\{ \vec{k} \in \mathbb{R}^n : \ U_{t_0}(n)\vec{k} = \vec{y} + \vec{v} + \vec{y}_2 \\ &\qquad \vec{v},\ \vec{y}_2 \in \mathbb{R}^T,\ \|\vec{v}\|_\infty \le \eta,\ \|\vec{y}_2\|_\infty \le \epsilon C_u \Big\}. \\ &= \Big\{ \vec{k} \in \mathbb{R}^n : \ U_{t_0}(n)\vec{k} - \vec{y} \le [\eta + \epsilon C_u, \ldots, \eta + \epsilon C_u]^T \ \ and \\ &\qquad -U_{t_0}(n)\vec{k} + \vec{y} \le [\eta + \epsilon C_u, \ldots, \eta + \epsilon C_u]^T \Big\}, \end{aligned}$$

which is a polytop.

**Example 5.2** Let $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\ell^1}$, $\mathbf{X}_n = \mathbf{L}_n = sp\{1, z, \ldots, z^{n-1}\}$, and $\mathbf{V} = \{v \in \ell^\infty : \|v\|_\infty \le \eta\}$. We assume that the entire input is given. The output observations are made on the interval $[t_0, t_0 + T)$. The unmodeled dynamics are assumed to be in the $\epsilon$-ball of the set of causal BIBO stable systems which includes the time varying systems,(see Chapter 8 for details on time-varying systems,) which is still denoted by $\mathbf{B}(\epsilon)$, i.e.,

$$\mathbf{B}(\epsilon) := \Big\{ k(\cdot,\cdot) : \mathbb{Z} \times \mathbb{Z}_+ \to \mathbb{R} : \ \sup_t \|k(t,\cdot)\|_1 \le \epsilon \Big\}. \tag{5.15}$$

We are going to show that, in this case, $\mathbf{S}_1(y)$ is also a polytope, i.e.,

$$\mathbf{S}_1(y) = \left\{ k_1 \in \mathbf{L}_n : \left| \sum_{\tau=0}^{n-1} k_1(\tau)u(t-\tau) - y(t) \right| \le \eta + \epsilon \|\mathbf{P}_{(-\infty,t]}(u)\|_\infty \ \forall t \in [t_0, t_0+T) \right\}. \tag{5.16}$$

By definition of $\mathbf{S}_1(y)$ in (5.1), for $k_1 \in \mathbf{S}_1(y)$, there exist a system $k_2 \in \mathbf{B}(\epsilon)$ and a disturbance $v \in \mathbf{V}$ such that

$$\sum_{\tau=0}^{n-1} k_1(\tau)u(t-\tau) - y(t) = v(t) + \sum_{\tau=0}^{\infty} k_2(t,\tau)u(t-\tau) \ \forall t \in [t_0, t_0+T), \tag{5.17}$$

which implies that

$$\left| \sum_{\tau=0}^{n-1} k_1(\tau)u(t-\tau) - y(t) \right| \le \eta + \epsilon \|\mathbf{P}_{(-\infty,t]}(u)\|_\infty \ \forall t \in [t_0, t_0+T). \tag{5.18}$$

Conversely, if a system in $\mathbf{L}_n$ satisfies (5.18), then there exist $v \in \mathbf{V}$ and $y_2$, $\|\mathbf{P}_{(-\infty,t]}(y_2)\|_\infty \le \epsilon \|\mathbf{P}_{(-\infty,t]}(u)\|_\infty$, such that

$$\sum_{\tau=0}^{n-1} k_1(\tau)u(t-\tau) - y(t) = v(t) + y_2(t). \tag{5.19}$$

To show $k_1 \in \mathbf{S}_1(y)$, it is enough to show that there is a system $k_2 \in \mathbf{B}(\epsilon)$ such that $y_2(t) = \mathbf{K}_2(u)(t)$ on $[t_0, t_0+T)$. Next, we will construct a time-varying system $k \in \mathbf{B}(\epsilon)$ satisfying this condition. Assume that the supremum of the input pasts $\mathbf{P}_{(-\infty,t]}(u)$ is achieved at time $m_t$, i.e., $|u(m_t)| = \|\mathbf{P}_{(-\infty,t]}(u)\|_\infty$. Put

$$k_2(t,\tau) = \begin{cases} y_2(t)/u(m_t) & \tau = t - m_t, \\ 0 & otherwise. \end{cases} \tag{5.20}$$

$k \in \mathbf{B}(\epsilon)$ as $\|k\| = \sup_t |y_2(t)/u(m_t)| \le \epsilon$. Also, $\sum_{\tau=0}^{\infty} k_2(t,\tau)u(t-\tau) = y_2(t)$. This proves our claim.

In general, $\mathbf{S}_1(y)$ may not be a polytope, but often it is possible to find a polytope covering $\mathbf{S}_1(y)$ tightly. In these cases, an algorithm based on polytope can be used to construct a slightly inaccurate estimate of the nominal and related bounds.

### 5.2.1 An Algorithm Based on the Analytic Center

In the following, we will define the analytic center of a polytope and derive an algorithm which chooses the analytic center as the estimate. The analytic center is always in the polytope. Therefore, as shown in Proposition 5.1, the worst-case error for this estimate is within a factor of two of the optimal. Moreover, the analytic center is robust with respect to the inaccuracies in the a priori information.

Given a polytope in $\mathbb{R}^n$ defined by a set of linear inequalities,

$$\mathbf{S}_1(y) = \left\{\vec{k} \in \mathbb{R}^n : \quad \vec{a}_i^T \vec{k} \le b_i, \quad i = 1, \dots, 2T\right\}, \tag{5.21}$$

its barrier function is defined as

$$\psi(\vec{k}) := \begin{cases} -\log\left(\prod_{i=1}^{2T}(b_i - \vec{a}_i^T \vec{k})\right) & \vec{k} \in \mathbf{S}_1(y), \\ \infty & \vec{k} \notin \mathbf{S}_1(y). \end{cases} \tag{5.22}$$

It can be shown that $\psi$ is analytic and strictly convex on $\mathbf{S}_1(y)$. The analytic center of $\mathbf{S}_1(y)$ is defined as the unique minimizer of $\psi$, which is denoted by $\vec{k}_*$, i.e.,

$$\psi(\vec{k}_*) = \min_{\vec{k} \in \mathbf{S}_1(y)} \psi(\vec{k}). \tag{5.23}$$

Equivalently, $\vec{k}_*$ is the maximizer of $\prod_{i=1}^{2T}(b_i - \vec{a}_i^T \vec{k})$, i.e.,

$$\prod_{i=1}^{2T}(b_i - \vec{a}_i^T \vec{k}_*) = \max_{\vec{k} \in \mathbf{S}_1(y)} \prod_{i=1}^{2T}(b_i - \vec{a}_i^T \vec{k}). \tag{5.24}$$

There are several efficient algorithms for finding the analytic center of a convex set. Details of the general definition of the analytic center and algorithms to find it can be found in [4] and the references therein.

The analytic center $\mathbf{S}_1(y)$ is the point in the polytope furthest away from the boundary, as it maximizes the product of the distances to the constraint planes $\vec{a}_i^T \vec{k} = b_i$. This implies that the nominal given by the analytic center is the system in the a posteriori uncertainty set which is the least sensitive to inaccuracies in the a priori assumptions. Since the $b_i$'s in $\mathbf{S}_1(y)$ are given by the measurements and the a priori information on the noise and unmodeled dynamics, any inaccuracy in the a priori information will cause the boundary planes to move. Some points which are in the polytope will be eliminated when the boundary moves. The analytic center is the point in the polytope least likely to be eliminated.

Once the analytic center is chosen to be the nominal, the worst-case estimation error is given by

$$\sup_{k \in \mathbf{S}_1(y)} \|k - k_{est}\|_{\mathbf{L}}. \tag{5.25}$$

(5.25) is a problem of maximizing a convex function on a polytope. It is well known that the maximum is achieved at the vertices of the polytope. Assume that $k_1, \ldots, k_m$ are the vertices of $\mathbf{S}_1(y)$. The worst-case estimation error can be computed via the following formula:

$$\max_{1 \leq i \leq m} \|k_i - k_{est}\|_{\mathbf{L}}. \tag{5.26}$$

Since $\mathbf{S}_1(y)$ is a polytope in $\mathbb{R}^n$, the number of its vertices will not exceed $(2T)^n$. In fact, in most practical cases, the number of vertices is much smaller than this upper bound. Statistical results on the number of vertices can be found in [30]. The upper bound will have a polynomial order increase with the number of measurements. The polynomial property is an advantage of choosing a nominal in a predetermined finite parameter model set. Otherwise, as shown in [6], the computation complexity for the worst-case estimation error increases combinatorically with the number of measurements. Standard algorithms for finding the vertices of a polytope can be found in a survey paper by Matheiss and Rubin [30].

In fact, once the vertices of $\mathbf{S}_1(y)$ are found, the problem (5.3) of minimizing the worst-case uncertainty is readily solved. In this case,

$$\mathcal{J}(k_{est}) = \max_{1 \leq i \leq m} \|k_i - k_{est}\|_{\mathbf{L}}. \tag{5.27}$$

Since $\mathcal{J}(k_{est})$ is the maximum of $m$ convex functions in $k_{est}$, the problem (5.3) can be solved easily by any one of the convex optimization algorithms, e.g., in the book by Boyd and Barratt.

In summary, we propose the following identification algorithm:

(i) Form the polytope $\mathbf{S}_1(y)$;

(ii) Find the analytic center $k_{est}$ of $\mathbf{S}_1(y)$;

(iii) Find the vertices $\{k_i\}_1^m$ of $\mathbf{S}_1(y)$;

(iv) Compute the worst-case estimation error, $\max_{1\leq i\leq m}\|k_i - k_{est}\|_{\mathbf{L}}$.

### 5.2.2 A Minimum Description-Length Algorithm

Instead of the analytic center, we can choose a system in $\mathbf{S}_1(y)$ which needs the fewest parameters to describe, i.e., the lowest order FIR system in $\mathbf{S}_1(y)$.

Let $\mathbf{L}_i$ be the $i$th order FIR model set, i.e., $\mathbf{L}_i := sp\{1,\ldots,z^i\}$. To find the lowest order FIR system which is consistent with the a posteriori information, we check whether the set $\mathbf{S}_1(y)\bigcap\mathbf{L}_i$ is empty, starting from $i = 0$. If $\mathbf{S}_1(y)\bigcap\mathbf{L}_i$ is empty, we increase the order of the model set by 1 and check whether $\mathbf{S}_1(y)\bigcap\mathbf{L}_{i+1}$ is empty. This procedure ends within $n$ steps. There are several algorithms available for checking whether a convex set is empty. One of them is the ellipsoid algorithm (see [3] for details). Once a non-empty intersection $\mathbf{S}_1(y)\bigcap\mathbf{L}_i$ is found, its analytic center can be chosen as a nominal. This nominal is one of the lowest order FIR systems in $\mathbf{S}_1(y)$. Since each step is a convex optimization problem, to find the estimate we only need to solve at most $n$ convex optimization problems.

In summary, we give the following algorithm:

(i) Form the polytope $\mathbf{S}_1(y)$ and set $i = 0$;

(ii) Form the set $\mathbf{S}_1(y)\bigcap \mathbf{L}_i$ by adding on the constraints $k(i) = \ldots = k(n-1) = 0$;

(iii) Check whether $\mathbf{S}_1(y)\bigcap \mathbf{L}_i$ is empty. If it is, set $i = i+1$ and go to step (ii); otherwise, continue;

(iv) Find the analytic center $k_{est}$ of $\mathbf{S}_1(y)\bigcap \mathbf{L}_i$;

(v) Find the vertices $\{k_i\}_1^m$ of $\mathbf{S}_1(y)$;

(vi) Compute the worst-case estimation error $\max_{1\leq i\leq m} \|k_i - k_{est}\|_{\mathbf{L}}$.

## 5.3 An Algorithm for an Input with Flat Spectrum

As shown in Chapter 3, the model set $\mathbf{L}_n = sp\{1, \ldots, z^{n-1}\}$ is the optimal $n$-parameter affine model set for a class of a priori data sets. This suggests that for systems in these data sets, estimation of the first $n$ coefficients of their impulse responses is the most efficient way of identifying and representing the systems. A well known method [26] for estimating the coefficients of an impulse response is to compute the inner products of the output and delayed inputs on the observation interval, i.e.,

$$\hat{k}(i) = \frac{1}{\phi_{uu}(0, t_0)} \sum_{t=t_0}^{t_0+T-1} y(t)u(t-i). \qquad (5.28)$$

When the input is a white noise and the disturbance is uncorrelated with the input, $\hat{k}(i)$ converges to the $k(i)$ as $T \to \infty$. In fact, it can be shown that even on a finite interval, these estimates will be quite accurate when the input is highly mixing. In the next proposition, we give a hard bound on the worst-case estimation error in the $\ell^1$ norm for a special a priori data set which has been considered in Example 2.1. Similar results can be obtained for the other cases.

**Proposition 5.2** *Let*

$$\mathbf{S}_a = \left\{k \in \ell^1 : \ |k(\tau)| \le r^\tau \ \forall \tau \in \mathbb{Z}_+\right\}, \tag{5.29}$$

*and*

$$\mathbf{V} = \left\{v \in \ell^\infty : \ \|\mathbf{P}_{[t_0,t_0+T)}(v)\|_2 \le \eta\right\}. \tag{5.30}$$

*If $u$ is $T$-mixing, then the worst-case estimation error*

$$\mathcal{J}(k_{est}) := \sup_{k \in \mathbf{S}_1(y)} \|k - k_{est}\|_1 \tag{5.31}$$

*of*

$$k_{est} := \sum_{i=0}^{n-1} \hat{k}(i) z^i \tag{5.32}$$

*has upper bound*

$$\mathcal{J}(k_{est}) \le \left(1 - \frac{\gamma_u(n,T)}{\phi_{uu}(0,t_0)}\right)\frac{\sqrt{n}}{1-r} + \left(\frac{n\nu_u(n,T)}{\phi_{uu}(0,t_0)}\right)^{1/2}\eta + \nu_u(n,T)\left(\frac{r^n\sqrt{n}}{1-r}\right). \tag{5.33}$$

**Proof** By definition,

$$\mathbf{S}_1(y) = \Big\{k_1 \in \mathbf{P}_{[0,n)}(\mathbf{S}_a) : \quad \mathbf{P}_{[t_0,t_0+T)}(\Phi_u(k_1) - y) = \mathbf{P}_{[t_0,t_0+T)}(-\Phi_u(k_2) + v)$$
$$\textit{for some } k_2 \in \mathbf{P}_{[n,\infty)}(\mathbf{S}_a) \textit{ and } v \in \mathbf{V}\Big\}. \tag{5.34}$$

Fix $k_1 \in \mathbf{S}_1(y)$, there exist $k_2 \in \mathbf{P}_{[n,\infty)}(\mathbf{S}_a)$ and $v \in \mathbf{V}$, such that

$$\sum_{\tau=0}^{n-1} k_1(\tau)u(t-\tau) - y(t) = v(t) - \sum_{\tau=n}^{\infty} k_2(\tau)u(t-\tau) \ \forall t \in [t_0, t_0+T). \tag{5.35}$$

Set $y_2(t) := -\sum_{\tau=n}^{\infty} k_2(\tau)u(t-\tau)$ and $\vec{k}_1, \vec{y}, \vec{v}, \vec{y}_2$ as in (5.11)-(5.14). Then (5.35) is equivalent to

$$U_{t_0}(n)\vec{k}_1 - \vec{y} = \vec{v} + \vec{y}_2. \tag{5.36}$$

Multiply $U_{t_0}(n)^T$ on both sides of (5.36), we get

$$\Phi_{u,t_0}(n)\vec{k}_1 - U_{t_0}(n)^T\vec{y} = U_{t_0}(n)^T\vec{v} + U_{t_0}(n)^T\vec{y}_2. \tag{5.37}$$

It is trivial to check that

$$\vec{k}_{est} := [\hat{k}(0), \ldots, \hat{k}(n-1)]^T = U_{t_0}(n)^T \vec{y} / \phi_{uu}(0, t_0). \tag{5.38}$$

Therefore,

$$\begin{aligned} \|k_1 - k_{est}\|_1 &= \left\| (I - \Phi_{u,t_0}(n)/\phi_{uu}(0, t_0))\, \hat{k}_1 + U_{t_0}(n)^T (\vec{v} + \vec{y}_2)/\phi_{uu}(0, t_0) \right\|_1 \quad (5.39) \\ &\leq \left\| (I - \Phi_{u,t_0}(n)/\phi_{uu}(0, t_0))\, \hat{k}_1 \right\|_1 + \left\| U_{t_0}(n)^T \vec{v}/\phi_{uu}(0, t_0) \right\|_1 \\ &\quad + \left\| U_{t_0}(n)^T \vec{y}_2/\phi_{uu}(0, t_0) \right\|_1 . \quad (5.40) \end{aligned}$$

Noticing that $\underline{\sigma}(\Phi_{u,t_0}(n)) > \gamma_u(n, T)$ and $\bar{\sigma}(\Phi_{u,t_0}(n))/\phi_{uu}(0, t_0) \leq \nu_u(n, T)$, we get the desired upper bound from (5.40).

□

It will be shown in Chapter 8 that this algorithm is easy to implement for time-varying system identification.

# Chapter 6

# Fast Identification of Continuous-Time Systems

In this chapter, we first formulate the fast identification problem for continuous-time systems in a way similar to that in Chapter 2. Then we introduce two measures of identification speed similar to the time n-widths. Their properties and relation to Gel'fand n-width are studied.

The fast identification problem in the continuous-time case is different from its discrete-time counterpart. On a continuous-time interval, it is possible to collect an unlimited amount of sampled data provided the sampling can be made arbitrarily fast. It remains unclear, however, whether arbitrarily accurate identification can be achieved on the basis of this large amount of data. In Section 6.3, it is shown, that in the noise free case, one can identify a stable continuous-time LTI system exactly on an arbitrarily short time interval, provided the entire segment of the output on the interval is available and the input is chosen properly. Here, the only a priori information is that the system is BIBO stable. No structural information or quantitative information about the system is required. A logarithmic integral condition on the inputs involved is obtained via quasianalyticity theory.

In such a case, however, accurate identification becomes impossible when the measurements are even slightly corrupted by noise. Similarly, the inherent error can be large if only samples of the output on a interval are available. An example is given where the inherent error is the same as the a priori uncertainty no matter how fast the sampling. Nevertheless, it is shown in Section 6.4 that, if the system is known to be in a compact set, (in either the $\mathbf{H}^\infty$ norm or $\mathbf{L}^1$ norm,) then the inherent error can be made arbitrarily small provided there are enough sampling points in an interval in which the noise-to-signal ratio is small enough.

With the above results in mind, one can ask: is identification speed still restricted? Is the metric complexity still a factor limiting identification speed? The answer to these questions is affirmative. It is shown in Section 6.5 that for a fixed noise level, even if the sampling rate is infinitely high, there is an irreducible uncertainty whenever the a priori uncertainty set contains a smooth subset (e.g. a set of low-pass functions,) of positive Gel'fand n-width. The higher the metric complexity, the slower the identification. Finally, the irreducible uncertainty is obtained for a set of approximately band-limited and time-limited systems in an example.

This chapter is based on [25].

## 6.1 Formulation

We will consider continuous-time LTI systems with integrable impulse responses. We denote the space of such systems, equipped with either the $\mathbf{L}^1$ norm of the impulse responses or the $\mathbf{H}^\infty$ norm of the transfer functions, by $\mathbf{L}$. The admissible input signals are those in the unit ball of $\mathbf{L}^\infty(-\infty,\infty)$, denoted by $\mathbf{B_U}$. The system with impulse response $k$ acts on an input $u$ in the following convolution form:

$$y(t) = \int_0^\infty k(\tau)u(t-\tau)d\tau + v(t). \tag{6.1}$$

where $v(t)$ is the measurement noise. Sometimes we denote this relation in a more compact way, as $y(t) = (k * u)(t) + v(t)$.

We are given the *a priori information* that a system $k$ lies in a subset $\mathbf{S}_a$ of $\mathbf{L}$ and the measurement noise has bounded $\mathbf{L}^\infty$ norm, i.e., $\|v\|_\infty \leq \epsilon$. It is assumed that $\mathbf{S}_a$ is convex and symmetric, i.e., $k \in \mathbf{S}_a \Rightarrow -k \in \mathbf{S}_a$.

Given an input $u$ in $\mathbf{B_U}$, the objective of identification is to estimate the system from the noise corrupted observations of the output $y$ at sample points $t_i = t_0 + i/s$, $i = 0, \ldots, \zeta - 1$, in the time interval $[t_0, t_0 + T]$, where $s$ is the sampling frequency and $\zeta$ is the number of sampling points in the interval, $\zeta = Int(sT)$, the smallest integer strictly greater than $sT$. On the basis of these observations, $y(t_i) = (k_{true} * u)(t) + v(t_i)$, $i = 0, \ldots, \zeta - 1$, the location of the true kernel, $k_{true}$, is narrowed down from the a priori data set $\mathbf{S}_a$ to a smaller a posteriori set, $\mathbf{S}(y)$,

$$\mathbf{S}(y) := \{k \in \mathbf{S}_a : \ |(k_{true} * u - k * u)(t_i) + v(t_i)| \leq \epsilon \ \ for \ i = 0, \ldots, \zeta - 1\}. \quad (6.2)$$

As in the discrete-time case, we are going to represent the a posteriori information by a nominal and a ball in the normed space $\mathbf{L}$, and the nominal will be chosen from a finite parameter model set.

**Definition 6.1** *A subset $\mathbf{X}_n$ of $\mathbf{L}$ is called a n-parameter model set if it is in the range of a mapping from $\mathbb{R}^n$ to the set of real functions on $\mathbb{R}$.*

If the estimated kernel, $k_{est}$ is optimally chosen from a $n$-parameter model set $\mathbf{X}_n$ for $\mathbf{S}(y)$, then the worst-case identification error

$$e^T(u, \mathbf{X}_n, s, \epsilon) := \sup_{k_{true} \in \mathbf{S}_a} \ \sup_{\|v\|_\infty \leq \epsilon} \ \inf_{k_{est} \in \mathbf{X}_n} \ \sup_{k \in \mathbf{S}(y)} \|k - k_{est}\|_{\mathbf{L}} \quad (6.3)$$

is a function of the input $u$ and model set $\mathbf{X}_n$ when the sampling rate and noise level are fixed. If, instead of $\zeta$ samples on the interval $[t_0, t_0 + T]$, the entire segment of the output is available, then the worst-case uncertainty is denoted by $e^T(u, \mathbf{X}_n, \infty, \epsilon)$.

When the observation interval is not fixed, we consider the shift-invariant worst case identification error

$$\bar{e}^T(u, \mathbf{X}_n, s, \epsilon) := \sup_{t_0 \in \mathbf{Z}} e^T(u, \mathbf{X}_n, s, \epsilon), \tag{6.4}$$

As in the discrete-time case, the worst-case identification error can be split into two terms, the inherent error

$$\delta^T(u, s, \epsilon) := \sup \left\{ \|k\|_{\mathbf{L}} : \ k \in \mathbf{S}_a, \ |(k * u)(t_i)| \le \epsilon \ \ for\ i = 0, \ldots, \zeta - 1 \right\}, \tag{6.5}$$

and the representation error

$$dist(\mathbf{S}_a, \mathbf{X}_n) := \sup_{k \in \mathbf{S}_a} \ \inf_{g \in \mathbf{X}_n} \|k - g\|_{\mathbf{L}}. \tag{6.6}$$

Using an argument similar to that in Proposition 2.1, we can show that

$$\max\left\{\delta^T(u, s, \epsilon), dist(\mathbf{S}_a, \mathbf{X}_n)\right\} \le e^T(u, \mathbf{X}_n, s, \epsilon) \le 3 \max\left\{\delta^T(u, s, \epsilon), dist(\mathbf{S}_a, \mathbf{X}_n)\right\}, \tag{6.7}$$

and

$$\max\left\{\bar{\delta}^T(u, s, \epsilon), dist(\mathbf{S}_a, \mathbf{X}_n)\right\} \le \bar{e}^T(u, \mathbf{X}_n, s, \epsilon) \le 3 \max\left\{\bar{\delta}^T(u, s, \epsilon), dist(\mathbf{S}_a, \mathbf{X}_n)\right\}, \tag{6.8}$$

where

$$\bar{\delta}^T(u, s, \epsilon) := \sup_{t_0 \in \mathbb{R}} \delta^T(u, s, \epsilon). \tag{6.9}$$

In this chapter, we will mainly study the dependency of the inherent error on the input, noise, and length of the observation interval. The representation and estimation problems will be investigated in Chapter 7.

## 6.2 Two Measures of Identification Speed

In the following we introduce two notions, similar to the time n-widths, as measures of identification speed. The first one is the best achievable inherent error on a fixed observation interval,

$$\vartheta^T(s,\epsilon) := \inf_{u\in \mathbf{B}_{\mathbf{U}}} \delta^T(u,s,\epsilon). \tag{6.10}$$

The second is the best achievable shift-invariant inherent error,

$$\bar{\vartheta}^T(s,\epsilon) := \inf_{u\in \mathbf{B}_{\mathbf{U}}} \sup_{t_0\in \mathbb{R}} \delta^T(u,s,\epsilon) \tag{6.11}$$

Obviously, $\vartheta^T(s,\epsilon) \leq \bar{\vartheta}^T(s,\epsilon)$, and both functions are monotone decreasing in $s$ and increasing in $\epsilon$.

It should be noted that these two notions are slightly different from the time n-widths. In both $\vartheta$ and $\bar{\vartheta}$, the infimum is taken over all inputs in the unit ball of $\ell^\infty$ instead of the whole space. Although this is not important in the noise-free case, it does make a difference in the noisy case, especially when the complete segment of the output is available on an interval.

In the rest of the chapter, the following questions will be answered:

**Problem 1:** In the noise free case, what is the best achievable inherent error and the best achievable shift-invariant inherent error when the complete segment of the output is available on an interval of length $T$, i.e., what is $\vartheta^T(\infty,0)$ and $\bar{\vartheta}^T(\infty,0)$? What are the optimal inputs?

**Problem 2:** Is it possible to approach the best achievable inherent error and the best achievable shift-invariant inherent error in Problem 1 by increasing the sampling rate and signal-noise ratio, i.e., will $\vartheta^T(s,\epsilon)$ and $\bar{\vartheta}^T(s,\epsilon)$ tend to $\vartheta^T(\infty,0)$ and $\bar{\vartheta}^T(\infty,0)$ respectively as $s\to\infty$ and $\epsilon\to 0$?

**Problem 3:** Is the length of the observation interval $T$ a limiting factor on the best achievable worst-case uncertainty, i.e., will $\vartheta^T(s,\epsilon)$ and $\bar{\vartheta}^T(s,\epsilon)$ be large when $T$ is small?

## 6.3 Exact Identification in the Noise-Free Case

In this section, Problem 1 is solved by showing that if $\mathbf{S}_a$ is a bounded set in $\mathbf{L}^1$, then there exists a class of bounded inputs which make $\delta^T(u,\infty,0) = 0$ for all $T > 0$. This implies that $\tilde{\vartheta}^T(\infty,0) = \vartheta^T(\infty,0) = 0\ \forall T > 0$.

To show the existence of such inputs, we first derive a condition on input signals which enables us to identify a system exactly if the entire output on $(-\infty,\infty)$ is known exactly. Then we will derive another condition on the inputs, using quasianalyticity theory [21], which enables us to recover the entire output from its values on any interval of positive length. Finally, we construct a class of bounded test inputs based on these results which allow exact identification on any interval.

**Lemma 6.1** *Assume $u(t) \in \mathbf{L}^\infty(-\infty,\infty)$ and the autocorrelation function*

$$\phi_u(t) := \lim_{A\to\infty} \frac{1}{2A}\int_{-A}^{A} u(t+\tau)\overline{u}(\tau)d\tau \qquad (6.12)$$

*exists for almost all $t \in \mathbb{R}$, and that $\phi_u \in \mathbf{L}^1$. If the spectrum of $u$, i.e., the support of the Fourier transform of $\phi_u$, $\Phi_u(\omega)$, is of positive measure in $\mathbb{R}$, then $u$ defines an injective mapping $\mathcal{U}$ from $\mathbf{L}^1[0,\infty)$ to $\mathbf{L}^\infty(-\infty,\infty)$, $\mathcal{U} : k \to y$, by*

$$(\mathcal{U}(k))(t) := y(t) = \int_0^\infty k(\tau)u(t-\tau)d\tau. \qquad (6.13)$$

**Proof** To show $\mathcal{U}$ is injective, it is enough to show that if $\mathcal{U}(k) = 0$ for some $k \in \mathbf{L}$, then $k = 0$. Since $\mathcal{U}(k) = 0$ implies that

$$K(\omega)\Phi_u(\omega) = 0, \qquad (6.14)$$

where $K(\omega)$ is the Fourier transform of k, $K(\omega) = 0$ on the support of $\Phi_u(\omega)$. Since by the hypothesis the support of $\Phi_u(\omega)$ is of positive measure, $K(\omega) = 0$ on a set of positive measure. Since $K$ is the boundary function of an analytic function on the right half plane, $K(\omega)$ can not vanish on a set of positive measure unless $K = 0$. This implies the injectiveness of $\mathcal{U}$.

□

This lemma shows that when the spectrum of the input is of positive measure, one can identify any system in $\mathbf{L}^1$ accurately on $(-\infty, \infty)$.

Generally speaking, it is impossible to recover a bounded function from its values on an interval. However, recovery becomes possible if we know that the function is the restriction of an analytic function. In fact, it is enough for the function to be quasianalytic.

**Definition 6.2** *Given any interval $I \subset \mathbb{R}$ and a sequence of numbers $M_n > 0$, we say that a function $f$, infinitely differentiable on $I$, belongs to the class $C_I(\{M_n\})$ if there are two numbers $c$ and $\rho$, depending on $f$, such that*

$$|f^{(n)}(x)| \leq c\rho^n M_n \tag{6.15}$$

*for $x \in I$ and $n = 0, 1, 2, \ldots$*

**Definition 6.3** *A class $C_I(\{M_n\})$ is called quasianalytic if, given any $x_0 \in I$, the only function $f \in C_I(\{M_n\})$ such that $f^{(n)}(x_0) = 0$, $n = 0, 1, \ldots$, has $f(x) \equiv 0$, $x \in I$.*

A quasianalytic class is similar to an analytic one in the sense that (a) its members are infinitely differentiable, and (b) if a function in such a class vanishes on a subinterval of $I$, then the function vanishes on $I$.

### 6.3.1 A Logarithmic Integral Condition

The next lemma gives a necessary and sufficient condition for a class to be quasianalytic in terms of a logarithmic integral. The proof can be found in [21].

**Lemma 6.2** *(Carleman's Criterion) Given any interval $I$ of positive length, the class $C_I(\{M_n\})$ is quasianalytic iff*

$$\int_0^\infty \left( \frac{\log F(t)}{1-t^2} \right) dt = \infty, \tag{6.16}$$

*where*

$$F(t) = \sup_{n \geq 0} \frac{t^n}{M_n}. \tag{6.17}$$

Using the logarithmic intergal condition in (6.16), we can characterize a class of inputs which can identify any stable system exactly on an arbitrarily small interval.

**Theorem 6.1** *Let $u \in \mathbf{L}^\infty(-\infty, \infty)$ be infinitely differentiable and*

$$F_u(t) = \sup_{n \geq 0} \frac{t^n}{\|u^{(n)}\|_\infty}, \tag{6.18}$$

*where $u^{(n)}$ denotes the $n$-th order derivative of $u$. If the spectrum of $u$ has positive measure in the sense of Lemma 6.1, and*

$$\int_0^\infty \left( \frac{\log F_u(t)}{1-t^2} \right) dt = \infty, \tag{6.19}$$

*then for all $T > 0$ and $\mathbf{S}_a$ bounded in $\|\cdot\|_{\mathbf{L}^1}$,*

$$\bar{\delta}^T(u, \infty, 0) = 0. \tag{6.20}$$

**Remarks** (i) Since $\delta^T(u,\infty,0) \leq \bar{\delta}^T(u,\infty,0)$, (6.20) implies that $\delta^T(u,\infty,0) = 0$ for all $t_0$.

(ii) The boundedness condition on the a priori uncertainty set $\mathbf{S}_a$ only requires the existence of such a bound, not an explicit value of the bound. Therefore, knowledge that the system is BIBO stable suffices.

(iii) Theorem 6.1 implies that, under it hypotheses, the uncertainty about a system can be reduced from $\mathbf{S}_a$ to a singleton in an arbitrarily short time.

**proof** It will be shown that $\delta^T(u,\infty,0) = 0$ for all $t_0 \in \mathbb{R}$. For this, first we show that the set of outputs

$$\mathcal{U}(\mathbf{S}_a) := \left\{ y:\ y(t) = \int_0^\infty k(\tau)u(t-\tau)d\tau,\ k \in \mathbf{S}_a \right\}$$

is a subset of a quasianalytic class. Since $\mathbf{S}_a$ is bounded in $\mathbf{L}^1$ norm, there exists $M \in \mathbb{R}$ such that $\sup\{\|k\|_1 :\ k \in \mathbf{S}_a\} \le M$. Let

$$y(t) = \int_0^\infty k(\tau)u(t-\tau)d\tau.$$

Since (by Lebesgue's Dominated Convergence Theorem,)

$$y^{(n)}(t) = \int_0^\infty k(\tau)u^{(n)}(t-\tau)d\tau,$$

for $k \in \mathbf{S}_a$, we have

$$\|y^{(n)}\|_\infty \le \|k\|_1\|u^{(n)}\|_\infty \le M\|u^{(n)}\|_\infty. \tag{6.21}$$

This implies that the set of outputs $\mathcal{U}(\mathbf{S}_a)$ is a subset of $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$. By the assumption (6.19) and Lemma 6.2, $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$ is quasianalytic.

Next we show that if an output in $\mathcal{U}(\mathbf{S}_a)$ equals zero on an interval $[t_0, t_0+T]$, then it is zero on $(-\infty,\infty)$. To see this, let $y \in \mathcal{U}(\mathbf{S}_a)$ and $y(t) = 0\ t \in [t_0, t_0+T]$. Since $y$ is infinitely differentiable, $y(t) = 0$ on $[t_0, t_0+T]$ implies that $y^{(n)}(t_0+T/2) = 0,\ n = 0,1,2,\ldots$. Being a member of quasianalytic class, $y^{(n)}(t_0+T/2) = 0,\ n = 0,1,2,\ldots$ only when $y(t) \equiv 0$, by definition of quasianalytic class. Therefore,

$$\begin{aligned} \delta^T(u,\infty,0) &:= \{\|k\|_{\mathbf{L}} :\ k \in \mathbf{S}_a,\ y(t) = 0\ \ \forall t \in [t_0, t_0+T]\}, && (6.22) \\ &= \{\|k\|_{\mathbf{L}} :\ k \in \mathbf{S}_a,\ (\mathcal{U}(k))(t) \equiv 0\ \forall t \in (-\infty,\infty)\}. && (6.23) \end{aligned}$$

Since the spectrum of $u$ has positive measure, Lemma 6.1 implies that

$$\{k :\ k \in \mathbf{S}_a,\ (\mathcal{U}(k))(t) \equiv 0\ \forall t \in (-\infty,\infty)\ \} = \{0\}. \tag{6.24}$$

Therefore, $\delta^T(u,\infty,0) = 0$.

□

### 6.3.2 A Class of Quasianalytic Inputs

It will be shown in this section that inputs satisfying the conditions specified in Theorem 6.1 can be generated by smoothing rich signals with a Gaussian filter. For this we need the following lemma.

**Lemma 6.3** *Let $g(t) = e^{-t^2/2}$. Then $\forall v \in \mathbf{L}^\infty$, the convolution $u := g * v$ is infinitely differentiable, $u^{(n)} = g^{(n)} * v$, and $\|u^{(n)}\|_\infty \le c2^n n!$ for some constant $c \in \mathbb{R}$.*

**Proof** First we prove that $u$ is infinitely differentiable and $u^{(n)} = g^{(n)} * v$ by induction. This equation is obviously true for $n = 0$. Assume this is true for $n-1$. Then the $n$-th order derivative

$$\begin{aligned} u^{(n)}(t) &:= \lim_{\Delta t \to 0} \frac{u^{(n-1)}(t+\Delta t) - u^{(n-1)}(t)}{\Delta t}, \\ &= \lim_{\Delta t \to 0} \int_{-\infty}^{\infty} \left( \frac{g^{(n-1)}(t-\tau+\Delta t) - g^{(n-1)}(t-\tau)}{\Delta t} \right) v(\tau) d\tau. \\ &\quad \text{(by hypothesis)} \end{aligned}$$

Since $g^{(n)}(x) = \exp(-x^2/2)P(x)$, where $P(x)$ is a $n$th order polynomial, there exists $X > 0$, such that $|g^{(n)}|$ is monotone decreasing with the rate $|\exp(-x^2/2)x^n|$ on $[X, \infty)$ and monotone increasing on $(-\infty, -X]$ with the same rate. Noticing that the following relations hold, (i) for all $t$ and $\tau$,

$$\begin{aligned} \left| \frac{g^{(n-1)}(t-\tau+\Delta t) - g^{(n-1)}(t-\tau)}{\Delta t} \right| &= \left| \int_{t-\tau}^{t-\tau+\Delta t} g^{(n)}(x) dx \right|, \\ &= \left| g^{(n)}(\xi) \right|, \quad (\ \xi \in [t-\tau, t-\tau+\Delta t]\ ) \\ &\le \|g^{(n)}\|_\infty; \end{aligned}$$

(ii) for $t - \tau \ge X$ and $|\Delta t| \le 1$,

$$\begin{aligned} \left| \frac{g^{(n-1)}(t-\tau+\Delta t) - g^{(n-1)}(t-\tau)}{\Delta t} \right| &\le \left| g^{(n)}(\xi) \right|, \quad (\ \xi \in [t-\tau, t-\tau+\Delta t]\ ) \\ &\le \left| g^{(n)}(t-\tau-1) \right| \quad (|g^{(n)}(x)| \text{ decreasing}); \end{aligned}$$

(iii) for $t - \tau \le -X$ and $|\Delta t| \le 1$,

$$\begin{aligned}\left|\frac{g^{(n-1)}(t-\tau+\Delta t) - g^{(n-1)}(t-\tau)}{\Delta t}\right| &\le \left|g^{(n)}(\xi)\right|, \quad (\ \xi \in [t-\tau, t-\tau+\Delta t]\ ) \\ &\le \left|g^{(n)}(t-\tau+1)\right| \quad (|g^{(n)}(x)| \text{ increasing});\end{aligned}$$

we get, for $|\Delta t| \le 1$,

$$\left|\frac{g^{(n-1)}(t-\tau+\Delta t) - g^{(n-1)}(t-\tau)}{\Delta t}\right| \le \begin{cases} \left|g^{(n)}(t-\tau+1)\right| & -\infty < t-\tau \le -X, \\ \|g^{(n)}\|_\infty & -X < t-\tau \le X, \\ \left|g^{(n)}(t-\tau+1)\right| & X < t-\tau \le \infty. \end{cases} \tag{6.25}$$

Knowing that the function on the right hand side of the above inequality is in $\mathbf{L}^1$, by Lebesgue's Dominated Convergence Theorem, we have

$$\lim_{\Delta t \to 0} \int_{-\infty}^{\infty} \left(\frac{g^{(n-1)}(t-\tau+\Delta t) - g^{(n-1)}(t-\tau)}{\Delta t}\right) v(\tau)d\tau = \int_{-\infty}^{\infty} g^{(n)}(t-\tau)v(\tau)d\tau. \tag{6.26}$$

Since this holds for all $t \in \mathbb{R}$, we have our claim.

Next, we show that $\|u^{(n)}\|_\infty \le c2^n n!$ for some constant $c$.

Since $u^{(n)} = g^{(n)} * v$, $\|u^{(n)}\|_\infty \le \|g^{(n)}\|_1 \|v\|_\infty$. Noticing that

$$\begin{aligned} g^{(n)} &= \left(e^{-x^2/4} \times e^{-x^2/4}\right)^{(n)} \\ &= \sum_{m=0}^{n} C_n^m \left(e^{-x^2/4}\right)^{(m)} \left(e^{-x^2/4}\right)^{(n-m)}, \end{aligned}$$

we have

$$\begin{aligned} \|g^{(n)}\|_1 &\le \sum_{m=0}^{n} C_n^m \| \left(e^{-x^2/4}\right)^{(m)} \left(e^{-x^2/4}\right)^{(n-m)} \|_1, \\ &\le \sum_{m=0}^{n} C_n^m \| \left(e^{-x^2/4}\right)^{(m)} \|_2 \| \left(e^{-x^2/4}\right)^{(n-m)} \|_2 \quad \text{(by Schwarz inequality)}, \\ &= c_1 \sum_{m=0}^{n} C_n^m \|\omega^m e^{-\omega^2}\|_2 \|\omega^{n-m} e^{-\omega^2}\|_2 \quad \text{(by Parseval's identity)} \end{aligned}$$

where $c_1$ is a constant. Since

$$\begin{aligned}\|\omega^m e^{-\omega^2}\|_2^2 &= \int_{-\infty}^{\infty} \omega^{2m} e^{-2\omega^2} d\omega \\ &= 2\sqrt{\frac{\pi}{2}}\frac{(2m-1)!}{2^{3m+1}m!},\end{aligned}$$

we have

$$\|\omega^m e^{-\omega^2}\|_2 \|\omega^{n-m} e^{-\omega^2}\|_2 \le \sqrt{\frac{\pi}{2}}n!. \tag{6.27}$$

It follows that

$$\begin{aligned}\|g^{(n)}\|_1 &\le c_1 \sum_{m=0}^{n} C_n^m \sqrt{\frac{\pi}{2}}n! & (6.28)\\ &= c_2 2^n n!, & (6.29)\end{aligned}$$

where $c_2$ is a constant. Therefore, for all $v \in \mathbf{L}^\infty$, there exists a constant $c = c_2\|v\|_\infty$, such that $\|u^{(n)}\|_\infty \le c2^n n!$.

□

**Proposition 6.1** *If the spectrum of $v \in \mathbf{L}^\infty$ has positive measure, and $u = g * v$, where $g(t) = e^{-t^2/2}$, then $u$ satisfies the condition of Theorem 6.1.*

**Proof** Since $\Phi_u(\omega) = G(\omega)\Phi_v(\omega)$ and $G(\omega)$ does not vanish anywhere, the spectrum of $u$ is the same as the spectrum of $v$, which is of positive measure. By Lemma 6.3, $u$ is infinitely differentiable and $\|u^{(n)}\|_\infty \le c2^n n!$ for some constant $c$.

Now instead of going through the tedious procedure of checking that $F_u$ satisfies the logarithmic integral condition (6.19), we will show that functions in $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$ are in fact analytic in a region $R$ containing the real line, i.e., $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right) \subset \mathbf{H}(R)$, using a device from [21]. This will imply that $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$ is quasianalytic. For an analytic function $f \in \mathbf{H}(R)$, $t \in \mathbb{R}$, $f^{(n)}(t) = 0,\ n = 0, 1, 2, \ldots$ implies that $f \equiv 0$. Therefore, by definition, $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right) \subset \mathbf{H}(R)$ is quasianalytic. Since the logarithmic integral condition is also a necessary condition for a set to be quasianalytic, quasianalyticity of $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$ will imply the satisfaction of the condition.

Let $f \in \mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$. Then, by definition of $\mathcal{C}_{\mathbb{R}}\left(\left\{\|u^{(n)}\|_\infty\right\}\right)$, there are two positive numbers $\alpha$ and $\rho$, such that

$$\begin{aligned}\|f^{(n)}\|_\infty &\leq \alpha\rho^n\|u^{(n)}\|_\infty \\ &\leq c\alpha(2\rho)^n n!.\end{aligned}$$

Therefore the Taylor expansion

$$f(t) = f(t_0) + \sum_{n=1}^{\infty} \frac{f^{(n)}(t_0)}{n!}(t-t_0)^n \tag{6.30}$$

converges for $|t - t_0| < 1/2\rho$ and $t_0 \in \mathbb{R}$. This implies that $f(t)$ is in fact the restriction to $\mathbb{R}$ of a function analytic in $R := \{z : |Im(z)| < 1/2\rho\}$.

□

**Example 6.1** Let $\lambda \in (0,1)$ be a real number with the binary expansion $0.a_1a_2\ldots$. Set

$$v(t) = \begin{cases} 2a_{2n+1} - 1 & n < t \leq n+1, \\ 2a_{2n} - 1 & -n < t \leq 1-n. \end{cases} \tag{6.31}$$

Choose $\lambda$ such that the autocorrelation function of $v$,

$$\phi_v(t) = \begin{cases} 1 - |t| & |t| \leq 1, \\ 0 & |t| \geq 1. \end{cases} \tag{6.32}$$

It is shown in [50, p.151] that almost all $\lambda \in (0,1)$ satisfies this condition. Therefore,

$$\Phi_v(\omega) = \frac{\sin^2(\omega/2)}{(\omega/2)^2}, \tag{6.33}$$

whose support has positive measure. By the above proposition and Theorem 6.1, if $u = g * v$, then $\bar{\delta}^T(u, \infty, 0) = 0$.

From this example, we get

**Theorem 6.2** *If $\mathbf{S}_a$ is a bounded set in $\mathbf{L}^1$, then for any $T > 0$,*

$$\vartheta^T(\infty, 0) = 0. \tag{6.34}$$

## 6.4 Arbitrarily Accurate Identification on a Finite Interval

In the proof of Theorem 6.1, we exploited the fact that in the noise free case, if the input is infinitely differentiable then the measured output is also infinitely differentiable; and if the output equals zero on an interval then the n-th derivative of the output is zero at the center of the interval, for all n. This will not be true when the output is even slightly corrupted by noise. Therefore the questions raised in Problem 2 arise: is Problem 1 well posed, i.e., given any interval, is it possible to identify a system arbitrarily accurately if the sufficiently high sampling rate and signal-noise ratio are sufficiently high?

Generally speaking, $\lim_{s\to\infty}\lim_{\epsilon\to 0}\bar{\vartheta}^T(s,\epsilon)$ may not be zero even when $\vartheta^T(\infty, 0) = 0$. The next example, derived from Proposition 6.3 in the next section, shows this clearly.

**Example 6.2** Let $\mathbf{S}_a$ be the unit ball of $\mathbf{L}^1$. Then

$$\bar{\vartheta}^T(s,\epsilon) \geq d^\zeta(\mathbf{S}_a, \mathbf{L}^1) = 1 \quad \textit{for all } s,\ \epsilon \in \mathbb{R}. \tag{6.35}$$

where $d^\zeta(\mathbf{S}_a, \mathbf{L}^1)$ is the Gel'fand n-width of $\mathbf{S}_a$ in $\mathbf{L}^1$, and $\zeta = Int(sT)$.

However, as shown in the next section, the limit converges to zero if $\mathbf{S}_a$ is compact in $\mathbf{L}^1$. The above example shows that the compactness condition is not dispensable.

**Proposition 6.2** *Let* $\mathbf{S}_a \subset \mathbf{L}$ *be compact in* $\|\cdot\|_{\mathbf{L}^1}$. *If* $u$ *satisfies the conditions in Theorem 6.1, then for all* $t_0 \in \mathbb{R}$, $T > 0$,

$$\lim_{s\to\infty}\lim_{\epsilon\to 0}\delta^T(u,s,\epsilon) = 0. \tag{6.36}$$

**Proof** We will show that for all $u$ satisfying the conditions in Theorem 6.1,

$$\lim_{s\to\infty}\lim_{\epsilon\to 0}\delta^T(u,s,\epsilon) = \delta^T(u,\infty,0). \tag{6.37}$$

Therefore, by Theorem 6.1, we have (6.36).

First we show that

$$\lim_{\epsilon\to 0}\delta^T(u,\infty,\epsilon) = \delta^T(u,\infty,0). \tag{6.38}$$

Since $\delta^T(u,\infty,\epsilon)$ is a monotone increasing function in $\epsilon$, it is enough to show that there exists a sequence $\{m_i\} \subset \mathbb{Z}_+$ such that $\delta^T(u,\infty,1/m_i) \to \delta^T(u,\infty,0)$ as $i \to \infty$. By definition of

$$\delta^T(u,\infty,\epsilon) := \sup\{\|k\|_{\mathbf{L}} : \ k \in \mathbf{S}_a, \ |k * u(t)| \le \epsilon \ \forall t \in [t_0, t_0 + T]\}, \tag{6.39}$$

$\forall m \in \mathbb{Z}_+$, $\exists k_m \in \mathbf{S}_a$ such that $|k * u(t)| \le 1/m \ \forall t \in [t_0, t_0+T]$ and $\delta^T(u,\infty,1/m) \le \|k_m\|_{\mathbf{L}} + 1/m$. Since $\{k_m\} \subset \mathbf{S}_a$ and $\mathbf{S}_a$ is compact in $\mathbf{L}^1$, there exists a subsequence $\{k_{m_i}\} \subset \{k_m\}$ such that $k_{m_i} \to k \in \mathbf{S}_a$ in $\|\cdot\|_{\mathbf{L}^1}$ as $i \to \infty$. It follows that

$$\lim_{i\to\infty}\delta^T(u,\infty,1/m_i) \le \lim_{i\to\infty}(\|k_{m_i}\|_{\mathbf{L}} + 1/m_i) = \|k\|_{\mathbf{L}}.$$

It will be shown that $|k * u(t)| = 0 \ \forall t \in [t_0, t_0 + T]$, which implies that $\|k\|_{\mathbf{L}} \le \delta^T(u,\infty,0)$. This implies (6.38), noticing that $\delta^T(u,\infty,\epsilon)$ is monotone increasing in $\epsilon$.

To show $|k * u(t)| = 0 \ \forall t \in [t_0, t_0 + T]$, we notice that for all $t$, $u \in \mathbf{L}^\infty$ defines a bounded linear functional on $\mathbf{L}^1$ via the convolution $(k * u)(t)$. Therefore, $k_{m_i} \to k \in \mathbf{S}_a$ in $\|\cdot\|_{\mathbf{L}^1}$ implies that for $t \in [t_0, t_0 + T]$,

$$(k * u)(t) = \lim_{i\to\infty}(k_{m_i} * u)(t) = 0.$$

Now we show that (6.37) holds. Set $\zeta = Int(sT)$. By definition,

$$\delta^T(u,s,\epsilon) := \sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}_a, |k * u(t_i)| \leq \epsilon\ t_i = t_0 + i/s,\ i = 0,\ldots,\zeta - 1\}. \quad (6.40)$$

Since $\mathbf{S}_a$ is compact in $\mathbf{L}^1$ and the derivative of $u$ is bounded, for $k \in \mathbf{S}_a$, $k * u$ is differentiable and there exists a constant $M$ such that $\|(k * u)'\|_\infty \leq M < \infty$. Therefore, for $k \in \mathbf{S}_a$, $|k * u(t_i)| \leq \epsilon\ t_i = t_0 + i/s,\ i = 0,\ldots,\zeta - 1$ implies that $|(k * u)(t)| \leq \epsilon + M/s$ on $[t_0, t_0 + T]$. Hence,

$$\begin{aligned} \delta^T(u,s,\epsilon) &\leq \sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}_a,\ |(k * u)(t)| \leq \epsilon + M/s\ \ \forall t \in [t_0, t_0 + T]\}, \\ &= \delta^T(u, \infty, \epsilon + M/s). \end{aligned}$$

Taking the limits on both sides of the inequalities and applying (6.38), we get

$$\lim_{s\to\infty}\lim_{\epsilon\to 0} \delta^T(u,s,\epsilon) \leq \lim_{\epsilon + M/s \to 0} \delta^T(u,\infty,\epsilon + M/s) = \delta^T(u,\infty,0). \quad (6.41)$$

Since $\delta^T(u,s,\epsilon) \geq \delta^T(u,\infty,0)$ for all $s$ and $\epsilon$, (6.41) implies (6.37).

□

**Corollary 6.1** *If $\mathbf{S}_a$ is compact in $\mathbf{L}^1$, then for $T > 0$,*

$$\lim_{s\to\infty}\lim_{\epsilon\to 0} \vartheta^T(s,\epsilon) = 0. \quad (6.42)$$

In fact a stronger claim can be made.

**Theorem 6.3** *If $\mathbf{S}_a$ is compact in $\mathbf{L}^1$, then for $T > 0$,*

$$\lim_{s\to\infty}\lim_{\epsilon\to 0} \bar{\vartheta}^T(s,\epsilon) = 0. \quad (6.43)$$

**Proof** We will show that there is a sequence of input functions which will identify the system arbitrarily accurately as $\epsilon$ tends to 0 and $s$ tends to infinity.

Let

$$v_N(t) = \begin{cases} 1 & 2^N n \le t \le 2^N(n+1/2), \ n = 0, \pm 1, \ldots \\ 0 & otherwise. \end{cases} \tag{6.44}$$

Then $v_N$ has discrete spectrum at $\omega_m = 2m\pi/2^N$, $m = 0, \pm 1, \ldots$ and its Fourier coefficients

$$|V_N(\omega_m)| = 2\left|\frac{\sin(m\pi/2)}{\omega_m}\right|. \tag{6.45}$$

Let $g(t) = \exp(-t^2/2)$ and $u_N = g * v_N$. Since $G(\omega) \neq 0$, $\forall \omega \in \mathbb{R}$, $U_N(\omega_m) = G(\omega_m)V_N(\omega_m) \neq 0$ *for* $m = 0, \pm 1, \pm 3, \ldots$

First, we show that

$$\lim_{N\to\infty} \sup_{t_0} \delta^T(u_N, \infty, 0) = 0. \tag{6.46}$$

By Lemma 6.3, there exists a constant $c$ such that $\|u_N^{(n)}\|_\infty \le c2^n n!$ for all $N$. Using an argument similar to the proof of Proposition 6.1, we can show that the set $\mathcal{U}_N\,(\mathbf{S}_a)$ is quasianalytic and for all $t_0$,

$$\begin{aligned} \delta^T(u_N, \infty, 0) &:= \sup\{\|k\|_{\mathbf{L}} : \ k \in \mathbf{S}_a, \ (k * u_N)(t) = 0 \ \forall t \in [t_0, t_0 + T]\} \\ &= \sup\{\|k\|_{\mathbf{L}} : \ k \in \mathbf{S}_a, \ (k * u_N)(t) = 0 \ \forall t \in \mathbb{R}\}. \end{aligned}$$

$(k * u_N)(t) = 0 \ \forall t \in \mathbb{R}$ implies $K(\omega_m)U_N(\omega_m) = 0$ *for* $m = 0, \pm 1, \pm 3, \ldots$, where $K$ is the Fourier transform of $k$. Since $U_N(\omega_m) \neq 0$ *for* $m = 0, \pm 1, \pm 3, \ldots$, we have

$$\delta^T(u_N, \infty, 0) = \sup\{\|k\|_{\mathbf{L}} : \ k \in \mathbf{S}_a, \ K(\omega_m) = 0, \ m = 0, \pm 1, \pm 3, \ldots\}. \tag{6.47}$$

As $\delta^T(u_N, \infty, 0)$ is a monotone decreasing in $N$, it is enough to show that there is a sequence $\{N_i\}$ such that $\delta^T(u_{N_i}, \infty, 0)$ converges to 0 as $i \to \infty$.

Now, for each $N$, there is $k_N \in \mathbf{S}_a$, $K_N(\omega_m) = 0$, $\omega_m = 2m\pi/2^N \ m = 0, \pm 1, \pm 3, \ldots$, and $\delta^T(u_N, \infty, 0) \le \|k_N\|_{\mathbf{L}} + 1/N$. Being a sequence in the compact set $\mathbf{S}_a$, $\{k_N\}$ contains a subsequence $\{k_{N_i}\}$ which converges to a function $k \in \mathbf{S}_a$ in $\|\cdot\|_{\mathbf{L}^1}$, i.e., $\lim_{i\to\infty} \|k_{N_i} - k\|_{\mathbf{L}^1} = 0$. Therefore,

$$\lim_{i\to\infty} \delta^T(u_{N_i}, \infty, 0) \le \lim_{i\to\infty} \|k_{N_i}\|_{\mathbf{L}^1} = \|k\|_{\mathbf{L}^1}. \tag{6.48}$$

It remains to show that $k = 0$.

As $\|\cdot\|_{\mathbf{H}^\infty} \le \|\cdot\|_{\mathbf{L}^1}$, $\lim_{i\to\infty} k_{N_i} = k$ in $\mathbf{H}^\infty$. $k \in \mathbf{S}_a \subset \mathbf{L}^1$ implies that the Fourier transform $K$ of $k$ is continuous. If $k \ne 0$, there exists an interval $I$ and a constant $a$ such that $|K(\omega)| > a > 0$ $\forall \omega \in I$. It follows that $|K_{N_i}(\omega)| > a/2 > 0$ $\forall \omega \in I$ for $N_i$ large enough, which is a contradiction.

Now we show (6.43). To express the dependency of $\delta^T$ on $t_0$ explicitely, we denote $\delta^T$ by $\delta^T(u, s, \epsilon, t_0)$. The limit in (6.43) exists because $\vartheta$ is monotone decreasing in $s$ and increasing in $\epsilon$. Therefore, it is enough to show that for integers $m$ and $l$,

$$\lim_{m\to\infty} \lim_{l\to\infty} \bar{\vartheta}^T(m, 1/l) = 0 \tag{6.49}$$

To show (6.49) by contradiction, we assume the limit in (6.49) is greater than some constant $2\sigma > 0$. By (6.46), $\exists N > 0$ such that

$$\sup_{t_0\in\mathbb{R}} \delta^T(u_N, \infty, 0, t_0) \le \sigma/2 \tag{6.50}$$

Since $u$ is a periodic function with period $2^N$, we have

$$\sup_{t_0\in\mathbb{R}} \delta^T(u_N, m, 1/l, t_0) \le \sup_{t_0\in[-T,2^N+T]} \delta^T(u_N, m, 1/l, t_0) \tag{6.51}$$

Therefore, by definition of $\bar{\vartheta}^T$,

$$\lim_{m\to\infty} \lim_{l\to\infty} \bar{\vartheta}^T(m, 1/l) \le \lim_{m\to\infty} \lim_{l\to\infty} \sup_{t_0\in[-T,2^N+T]} \delta^T(u_N, m, 1/l, t_0) \tag{6.52}$$

By the hypothesis, for each $m$ and $l$, there exists a $t_0^{(m,l)} \in [-T, 2^N + T]$ such that

$$\delta^T(u_N, m, 1/l, t_0^{(m,l)}) > \sigma. \tag{6.53}$$

Being a sequence in a closed interval, $\{t_0^{(m,l)}\}$ contains a subsequence $\{t_0^{(m_i,l_j)}\}$ which converges to $t_0^*$ and $|t_0^{(m_i,l_j)} - t_0^*| \le T/4$ for all $i$ and $j$. By definition of $\delta^T$,

$$\delta^T(u_N, m_i, 1/l_j, t_0^{(m_i,l_j)}) \le \delta^{T/2}(u_N, m_i, 1/l_j, t_0^* + T/4)\ \forall i,\ j \in \mathbb{Z}_+.$$

Therefore, by (6.53),

$$\lim_{i\to\infty}\lim_{j\to\infty}\delta^{T/2}(u_N, m_i, 1/l_j, t_0^* + T/4) > \sigma. \tag{6.54}$$

However, from (6.37) in the proof of Proposition 6.2 and (6.50), we get

$$\lim_{i\to\infty}\lim_{j\to\infty}\delta^{T/2}(u_N, m_i, 1/l_j, t_0^* + T/4) = \delta^{T/2}(u_N, \infty, 0, t_0^* + T/4) \le \sigma/2,$$

which contradicts to (6.54).

□

The condition in the theorem can be relaxed. If $\mathbf{S}_a$ is merely compact in the $\mathbf{H}^\infty$ norm but still bounded in the $\mathbf{L}^1$ norm, then the convergence still holds in $\mathbf{H}^\infty$.

The order of the limits in all of the above results is not important. Using similar proofs, we can show that if the order of the limits is reversed the conclusions still hold.

## 6.5 Identification Speed and Gel'fand n-Width

In the noise-free case, the uncertainty of a system known in a compact set can be reduced to an arbitrarily small neighborhood around the true system almost immediately when the sampling rate is high enough. However, it will be shown in this section that in the case where the measurement is corrupted by additive noise, there is an irreducible uncertainty even when the sampling rate is infinitely high. Moreover, the irreducible uncertainty is large when the observation interval is short. This uncertainty is given in terms of the Gel'fand n-width. It gives a lower bound on the best achievable inherent error, which will be shown to be asymptotically accurate.

When the sampling rate is finite, the optimal inherent error is bounded below by the Gel'fand n-width in a way similar to that in the discrete-time case.

**Proposition 6.3** *If there exists $\lambda > 0$ such that $\|k\|_{\mathbf{L}} \geq \lambda \|k\|_{\mathbf{L}_1}$ for all $k \in \mathbf{S}_a$, then*

$$\vartheta^T(s, 0) \geq d^\zeta(\mathbf{S}_a, \mathbf{L}), \tag{6.55}$$

*where $\zeta = Int(sT)$.*

**Remark** The hypothesis of the proposition is automatically satisfied when $\|\cdot\|_{\mathbf{L}} = \|\cdot\|_{\mathbf{L}_1}$. The proof which is similar to the one for Proposition 3.2 is omitted.

This proposition shows that for a fixed sampling rate the optimal inherent error is bounded below by a quantity related to the metric complexity of $\mathbf{S}_a$. When $\mathbf{S}_a$ has positive Gel'fand n-width, there is an irreducible uncertainty even when the noise is zero.

To get a lower bound for $\vartheta^T(\infty, \epsilon)$, we need to study certain subsets of $\mathbf{S}_a$.

### 6.5.1 A Notion of Smooth Subsets

**Definition 6.4** *Let $\mathbf{S}_M$ be a subset of the $M$-ball of $\mathbf{L}^1[0, \infty)$, i.e., $\sup\{\|k\|_{\mathbf{L}^1} : k \in \mathbf{S}_M\} \leq M$, and $W$ be a convolution operator with kernel $w \in \mathbf{L}^1[0, \infty)$ such that $\|w(t) - w(t + \Delta t)\|_1 \leq \kappa |\Delta t|$. If $W\mathbf{S}_M := \{w * k : k \in \mathbf{S}_M\} \subset \mathbf{S}_a$, we call $W\mathbf{S}_M$ a smooth subset of $\mathbf{S}_a$.*

**Example 6.3** Let $\mathbf{S}_a$ be the set considered in [52], i.e., for $C_1 > 0$, $, C_2 > 0$, $a > 0$,

$$\mathbf{S}_a := \left\{ k \in \mathbf{L} : \ |k(t)| \leq C_1 e^{-at} \ and \ |\omega K(j\omega)| \leq C_2 \right\} \tag{6.56}$$

Let

$$w(t) = \begin{cases} 1/T_1 & 0 \leq t \leq T_1 \\ 0 & t > T_1 \end{cases} \tag{6.57}$$

and $\mathbf{S}_M = \{k \in \mathbf{L} : \ \|k(t) e^{\alpha t}\|_1 < M\}$. It can be shown that $W\mathbf{S}_M$ is a smooth subset of $\mathbf{S}_a$ with $\kappa = 2/T_1$ for all $\alpha \geq a$ and $M \leq \min\{C_1 T_1 e^{-\alpha T_1}, C_2 T_1 / 2\}$.

For $k_1 \in W\mathbf{S}_M$, there exists a function $k \in \mathbf{S}_M$ such that $k_1 = w * k$. We notice that

$$\begin{aligned}
|(w * k)(t)| &\leq \frac{1}{T_1}\int_0^{T_1} |k(t-\tau)|d\tau, \\
&= \frac{e^{-\alpha\tau}}{T_1}\int_{t-T_1}^{t} |k(\tau)|e^{\alpha\tau}d\tau, \\
&\leq e^{-\alpha t}\frac{e^{\alpha\tau}}{T}M, \quad (\text{because } k \in \mathbf{S}_M) \\
&\leq C_1 e^{-\alpha t},
\end{aligned}$$

and

$$\begin{aligned}
|\omega W(j\omega)K(j\omega)| &\leq |\omega W(j\omega)|\|K\|_\infty, \\
&\leq M\left|\omega\frac{\sin(\omega T_1/2)}{\omega T_1/2}\right|, \\
&\leq 2M/T_1, \\
&\leq C_2.
\end{aligned}$$

Therefore, $k_1$ is in $\mathbf{S}_a$.

### 6.5.2 The Optimal Inherent Error with Finite Signal to Noise Ratio

With the notion of smooth subsets, we can get a lower bound on the optimal inherent error in the noisy case, where the sampling rate can be infinite.

**Proposition 6.4** *If conditions in Proposition 6.3 are satisfied, then*

$$\vartheta^T(\infty,\epsilon) \geq \sup_{W\mathbf{S}_M \subset \mathbf{S}_a} d^N(W\mathbf{S}_M, \mathbf{L}), \tag{6.58}$$

*where* $W\mathbf{S}_M$ *is a smooth subset of* $\mathbf{S}_a$ *and* $N = Int(\frac{\kappa MT}{\epsilon}) + 1$.

**Proof** First we prove that for all $WS_M \subset S_n$,

$$\begin{aligned}
\mathbf{S}_p &:= \{k \in W\mathbf{S}_M : |(k*u)(t)| \le \epsilon \ \forall t \in [t_0, t_0+T]\} \\
&\supset \{k \in W\mathbf{S}_M : (k*u)(t_i) = 0 \ for\ i = 0,\ldots,N-1\} \\
&=: \tilde{\mathbf{S}}_p
\end{aligned}$$

If $k \in \tilde{\mathbf{S}}_p$, $\exists k_1 \in \mathbf{S}_M$ such that $k = w * k_1$. This implies that

$$\begin{aligned}
|y(t) - y(t+\Delta t)| &:= |(k*u)(t) - (k*u)(t+\Delta t)|, \\
&= |(k_1 * (w*u))(t) - (k_1*(w*u))(t+\Delta t)|, \\
&\le \|k_1\|_1 \|(w*u)(t) - (w*u)(t+\Delta t)\|_\infty, \\
&\le M\|(w*u)(t) - (w*u)(t+\Delta t)\|_\infty
\end{aligned}$$

Since for $u \in \mathbf{B_U}$,

$$\begin{aligned}
& |(w*u)(t) - (w*u)(t+\Delta t)| \\
\le\ & \left|\int_{-\infty}^{t} (w(t-\tau) - w(t+\Delta t-\tau))u(\tau)d\tau\right| + \left|\int_{t}^{t+\Delta t} w(t+\Delta t-\tau)u(\tau)d\tau\right|, \\
\le\ & 2\|w(t) - w(t+\Delta t)\|_1 \|u\|_\infty, \\
\le\ & 2\kappa|\Delta t|, \qquad \text{(by the assumptions on } w)
\end{aligned}$$

we have $|y(t) - y(t+\Delta t)| \le 2\kappa M|\Delta t|$. $k \in \tilde{\mathbf{S}}_p$ implies that $y(t_i) = 0$ and hence $|y(t_i+\Delta t) \le 2\kappa M|\Delta t|$. Since $\forall t \in [t_0, t_0+T]$, $\exists t_i$ such that $|t_i - t| \le T/2(N-1)$, it follows that $|y(t)| \le \kappa MT/(N-1) \le \epsilon$. This implies that $k \in \mathbf{S}_p$.

Therefore,

$$\sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}_p\} \ge \sup\left\{\|k\|_{\mathbf{L}} : k \in \tilde{\mathbf{S}}_p\right\} \tag{6.59}$$

Since this holds for all $u \in \mathbf{B_U}$, we have

$$\begin{aligned}
\vartheta^T(\infty,\epsilon) &= \inf_{u \in \mathbf{U}} \delta^T(u,\infty,\epsilon), \\
&\ge \inf_{u \in \mathbf{B_U}} \sup\{\|k\|_{\mathbf{L}} : k \in \mathbf{S}_p\}, \quad (\text{because } \mathbf{WS_M} \subset \mathbf{S}_n,)
\end{aligned}$$

$$\geq \inf_{u \in \mathbf{B_U}} \sup \left\{ \|k\|_{\mathbf{L}} : \; k \in \tilde{\mathbf{S}}_p \right\},$$

$$\geq d^N(W\mathbf{S}_M, \mathbf{L}), \qquad \text{(by a similar argument as in Proposition 6.3.)}$$

The proof is completed by noticing that the above inequalities hold for all smooth subsets of $\mathbf{S}_a$.

□

This proposition shows that for a fixed noise level $\epsilon$, the optimal inherent error is bounded below by the Gel'fand n-widths of the smooth subsets of $\mathbf{S}_a$. Since the Gel'fand n-width $d^N$ is large when $N$ is small and $N$ is proportional to the length of the observation interval, the optimal inherent error can be quite large if the observation interval is short, no matter how fast the sampling. This coincides with the well known fact that it is impossible to estimate the low frequency components of a signal on a short time interval.

After combining Proposition 6.3 and Proposition 6.4, we get

### 6.5.3 An Asymptotically Accurate Lower Bound on the Optimal Inherent Error

**Theorem 6.4** *Under the conditions of Proposition 1,*

$$\vartheta^T(s,\epsilon) \geq \max\left( d^N(\mathbf{S}_a, \mathbf{L}), \sup_{W\mathbf{S}_M \subset \mathbf{S}_a} d^\zeta(W\mathbf{S}_M, \mathbf{L}) \right) \tag{6.60}$$

*where $n$, $W\mathbf{S}_M$ and $N$ are as in Proposition 6.4 and 6.3.*

Obviously, $\bar{\vartheta}^T(s,\epsilon)$ is also bounded below by the max in (6.60). This lower bound is asymptotically accurate in the following sense. For all $T > 0$,

$$\lim_{\substack{s\to\infty \\ \epsilon\to 0}} \bar{\vartheta}^T(s,\epsilon) = \lim_{\substack{s\to\infty \\ \epsilon\to 0}} \max\left( d^\zeta(\mathbf{S}_a, \mathbf{L}), \sup_{W\mathbf{S}_M \subset \mathbf{S}_a} d^N(W\mathbf{S}_M, \mathbf{L}) \right) = 0. \tag{6.61}$$

For all $0 < s < \infty$ and $\epsilon > 0$,

$$\begin{aligned}\lim_{T\to 0} \vartheta^T(s,\epsilon) &= \lim_{T\to 0} \max\left(d^\zeta(\mathbf{S}_a,\mathbf{L}),\ \sup_{W\mathbf{S}_M\subset\mathbf{S}_a} d^N(W\mathbf{S}_M,\mathbf{L})\right) \\ &= \|\mathbf{S}_a\|_{\mathbf{L}},\end{aligned} \tag{6.62}$$

where $\|\mathbf{S}_a\|_{\mathbf{L}} := \{\|k\|_{\mathbf{L}} :\ k\in\mathbf{S}_a\}$ is the a priori inherent error.

### 6.5.4 An Example

**Example 6.4** Let $\mathbf{S}_a$ be as in Example 6.3, with $C_1/C_2 = \epsilon/2$. We will show that

$$\vartheta^T(s,\epsilon) \geq \max\left\{\frac{e^{-C_2T/\epsilon}}{3 + C_2T/\epsilon}\frac{C_2}{2e^3a^2},\quad \frac{C_1e^{-sT}}{(ae)^2(sT+1)}\right\}, \tag{6.63}$$

where $M = \min\{C_1/ae, C_2/2a\}$.

First, we find a lower bound for the Gel'fand n-width $d^n(W\mathbf{S}_M,\mathbf{L}_1)$, where $w$ and $\mathbf{S}_M$ are as in Example 6.3.

By Proposition 3.5 in [38],

$$d^n(W\mathbf{S}_M,\mathbf{L}_1) \geq \sup_{X_{n+1}} \inf_{k\in\partial(W\mathbf{S}_M\bigcap X_{n+1})} \|k\|_1, \tag{6.64}$$

where $X_{n+1}$ is a $n+1$ dimensional subspace of $\mathbf{L}^1$. For $i = 1,\ldots,n+1$ and $T_2 > 0$, set

$$k_i(t) = \begin{cases} 1 & (i-1)(T_1+T_2) \leq t \leq (i-1)(T_1+T_2)+T_2, \\ 0 & otherwise, \end{cases} \tag{6.65}$$

$Y_{n+1} = sp\{k_1,\ldots,k_{n+1}\}$ and $X_{n+1} = Wsp\{k_1,\ldots,k_{n+1}\}$. It follows that

$$d^n(W\mathbf{S}_M,\mathbf{L}_1) \geq \sup_{T_2>0} \inf_{k\in\partial(W\mathbf{S}_M\bigcap X_{n+1})} \|k\|_1. \tag{6.66}$$

Since $k\in\partial(W\mathbf{S}_M\bigcap X_{n+1})$ if and only if there exists $h\in\partial(\mathbf{S}_M\bigcap Y_{n+1})$ such that $k = Wh$, we have

$$d^n(W\mathbf{S}_M,\mathbf{L}_1) \geq \sup_{T_2>0} \inf_{k\in\partial(\mathbf{S}_M\bigcap Y_{n+1})} \|Wk\|_1. \tag{6.67}$$

It can be shown that for $k \in Y_{n+1}$ $\|Wk\|_1 = \|k\|_1$. To see this, let $k = \sum_{i=1}^{n+1} \alpha_i k_i$. It follows that $\|Wk\|_1 = \|\sum_{i=1}^{n+1} \alpha_i W k_i\|_1$. From (6.57) and (6.65), we get (i) for $T_2 < T_1$,

$$(Wk_i)(t) = \begin{cases} 0 & 0 \le t \le (i-1)(T_1+T_2) \\ \frac{t-(i-1)(T_1+T_2)}{T_1} & (i-1)(T_1+T_2) < t \le (i-1)(T_1+T_2)+T_2 \\ T_2/T_1 & (i-1)(T_1+T_2)+T_2 < t \le (i-1)(T_1+T_2)+T_1 \\ \frac{i(T_1+T_2)-t}{T_1} & (i-1)(T_1+T_2)+T_1 < t \le i(T_1+T_2) \\ 0 & i(T_1+T_2) < t \end{cases} \tag{6.68}$$

and (ii) for $T_2 \ge T_1$,

$$(Wk_i)(t) = \begin{cases} 0 & 0 \le t \le (i-1)(T_1+T_2) \\ \frac{t-(i-1)(T_1+T_2)}{T_1} & (i-1)(T_1+T_2) < t \le (i-1)(T_1+T_2)+T_1 \\ 1 & (i-1)(T_1+T_2)+T_1 < t \le (i-1)(T_1+T_2)+T_2 \\ \frac{i(T_1+T_2)-t}{T_1} & (i-1)(T_1+T_2)+T_2 < t \le i(T_1+T_2) \\ 0 & i(T_1+T_2) < t \end{cases} \tag{6.69}$$

In both cases, the support of $Wk_i$ is disjoint with the support of $Wk_j$ if $i \neq j$. Therefore, $\|Wk\|_1 = \sum_{i=1}^{n+1} |\alpha_i| \|Wk_i\|_1$ Also in both cases, $\|Wk_i\|_1 = T_2 = \|k_i\|_1$. Hence, $\|Wk\|_1 = \sum_{i=1}^{n+1} |\alpha_i| \|k_i\|_1 = \|k\|_1$. It follows that

$$d^n(W\mathbf{S}_M, \mathbf{L}_1) \ge \sup_{T_2>0} \inf_{k \in \partial(\mathbf{S}_M \bigcap Y_{n+1})} \|k\|_1. \tag{6.70}$$

It can be shown that $\inf_{k \in \partial(\mathbf{S}_M \bigcap Y_{n+1})} \|k\|_1 \ge T_2 e^{-\alpha(n+1)T_2} e^{-\alpha n T_1} M$. To show this, it is enough to show that for $k = \sum_{i=1}^{n+1} \alpha_i k_i \in Y_{n+1}$, if $\|k\|_1 \le T_2 e^{-\alpha(n+1)T_2} e^{-\alpha n T_1} M$, then $k$ is in $\mathbf{S}_M$. Since for $k \in Y_{n+1}$

$$\begin{aligned} \int_0^\infty |k(\tau) e^{\alpha\tau}| d\tau &= \sum_{i=1}^{n+1} |\alpha_i| \int_{(i-1)(T_1+T_2)}^{(i-1)(T_1+T_2)+T_2} e^{\alpha\tau} d\tau, \\ &\le \sum_{i=1}^{n+1} |\alpha_i| e^{(i-1)(T_1+T_2)+T_2}, \end{aligned}$$

$$\leq \frac{1}{T_2} e^{\alpha(n+1)T_2} e^{\alpha n T_1} \|k\|_1.$$

$\|k\|_1 \leq T_2 e^{-\alpha(n+1)T_2} e^{-\alpha n T_1} M$ implies $k \in \mathbf{S}_M$.

By (6.70), we have

$$d^n(W\mathbf{S}_M, \mathbf{L}_1) \geq \sup_{T_2>0} T_2 e^{-\alpha(n+1)T_2} e^{-\alpha n T_1} M = \frac{e^{-\alpha n T_1}}{\alpha(n+1)} \frac{M}{e} \tag{6.71}$$

By Proposition 6.4,

$$\begin{aligned}
\vartheta^T(\infty, \epsilon) &\geq \sup_{T_1>0, \alpha\geq a, M\leq \min\{C_1 T_1 e^{-\alpha T_1}, C_2 T_1/2\}} \frac{e^{-\alpha N T_1}}{\alpha(N+1)} \frac{M}{e} \\
&\geq \sup_{T_1>0, M\leq \min\{C_1 T_1 e^{-a T_1}, C_2 T_1/2\}} \frac{e^{-a N T_1}}{a(N+1)} \frac{M}{e} \quad \text{(by letting } \alpha = a\text{)} \\
&\geq \sup_{M\leq \min\{C_1/ea, C_2/2a\}} \frac{e^{-N}}{a(N+1)} \frac{M}{e} \quad \text{(by letting } T_1 = 1/a\text{)} \\
&\geq \sup_{M\leq \min\{C_1/ea, C_2/2a\}} \frac{e^{-2aMT/\epsilon}}{a(3+2aMT/\epsilon)} \frac{M}{e^3} \quad \text{(because } N \leq 2 + 2aMT/\epsilon\text{)}.
\end{aligned}$$

If $C_1/ea = C_2/2a$, then by letting $M = C_2/2a$ we get a lower bound on $\vartheta^T(\infty, \epsilon)$, i.e.,

$$\vartheta^T(\infty, \epsilon) \geq \frac{e^{-C_2 T/\epsilon}}{3 + C_2 T/\epsilon} \frac{C_2}{2e^3 a^2}. \tag{6.72}$$

On the other hand, by, Proposition 6.3,

$$\vartheta^T(s, 0) \geq d^\zeta(\mathbf{S}_a, \mathbf{L}) \geq d^\zeta(W\mathbf{S}_M, \mathbf{L}_1), \tag{6.73}$$

where $\zeta = Int(sT)$. Applying (6.71) with $\alpha = a$, $T_1 = 1/a$, and $M = C_1/ae$, we get

$$\vartheta^T(s, 0) \geq \frac{C_1 e^{-sT}}{(ae)^2(sT+1)}. \tag{6.74}$$

Combining (6.72) and (6.74), we get the desired lower bound for $\vartheta^T(s, \epsilon)$.

# Chapter 7

# Representation and Estimation of Continuous-Time Systems

In this chapter, we study the problem of representation and estimation of continuous-time systems by sampling[1]. One of the key questions is: once a model set and estimate are obtained for the sampled data system, how should their continuous-time counterparts be constructed? Here, we give a causal procedure for the construction of a continuous-time estimate from a discrete one. Representation and estimation errors are given in the $\mathbf{L}_1$ norm which is an upper bound on the $\mathbf{H}^\infty$ norm. (Parts of this chapter are in [46].)

Continuous and discrete-time signals and systems will appear in pairs. To distinguish them, we use the carat symbol $\hat{}$ to denote the discrete-time quantities, e.g., $\hat{y}$ denotes the sampled version of the continuous-time output $y$. To guarantee that a sampled version is well defined, the continuous-time signals are constrained to be piecewise continuous. $\mathbf{L}^\infty$ denotes the set of piecewise continuous bounded real function on $\mathbb{R}$.

[1]Sampled data systems have been extensively studied in the past, e.g., [1]. Recently, they have been re-investigated in the robust control context [7].

## 7.1 Sampled Data Systems

A mapping $\mathbf{M} : \mathbf{L}^\infty(-\infty,\infty) \to \ell^\infty(-\infty,\infty)$ is *causal* if

$$\hat{\mathbf{P}}_{(-\infty,t]}\mathbf{M}\left(\mathbf{I} - \mathbf{P}_{(-\infty,t]}\right) = 0 \quad \forall t \in \mathbb{Z}, \tag{7.1}$$

where $\mathbf{I}$ is the identity operator on $\mathbf{L}^\infty$; and $\widehat{\mathbf{M}} : \ell^\infty(-\infty,\infty) \to \mathbf{L}^\infty(-\infty,\infty)$ is *causal* if

$$\mathbf{P}_{(-\infty,t]}\widehat{\mathbf{M}}\left(\hat{\mathbf{I}} - \hat{\mathbf{P}}_{(-\infty,m]}\right) = 0 \quad \forall t \in \mathbb{R}, \tag{7.2}$$

where $\hat{\mathbf{I}}$ is the identity operator on $\ell^\infty$, and $m = \sup\{l \in \mathbb{Z} : l \leq t\}$.

A *sampler* is any causal bounded map from $\mathbf{L}^\infty$ to $\ell^\infty$; and a *hold* is any causal bounded map from $\ell^\infty$ to $\mathbf{L}^\infty$. In particular, for a fixed *sampling period* $T_p > 0$, we define the (usual) *synchronized sampler* $\mathbf{S}_{T_p}$ by

$$(\mathbf{S}_{T_p}u)(m) = u(mT_p), \quad m \in \mathbb{Z},\ u \in \mathbf{L}^\infty; \tag{7.3}$$

and the (usual) *zero-order hold* $\mathbf{H}_{T_p}$

$$(\mathbf{H}_{T_p}\hat{u})(t) = \hat{u}(m), \quad t \in (mT_p, (m+1)T_p],\ \hat{u} \in \ell^\infty. \tag{7.4}$$

The subscript $T_p$ in $\mathbf{H}_{T_p}$ and $\mathbf{S}_{T_p}$ is assumed fixed and usually suppressed from notation.

$\mathbf{H}$ admits a convolution sum representation. Let $h_0 \in \mathbf{L}^1$ be the pulse function

$$h_0(t) = \begin{cases} 1, & 0 < t \leq T_p; \\ 0, & \textit{otherwise}. \end{cases} \tag{7.5}$$

Embed $\mathbb{Z}$ into $\mathbb{R}$ as usual. Define $h : \mathbb{R} \times \mathbb{Z} \to \mathbb{R}$ by

$$h(t,\tau) = h_0(t - \tau T_p), \quad t \in \mathbb{R},\ \tau \in \mathbb{Z}.$$

Then, $\mathbf{H}$ can be expressed as

$$(\mathbf{H}\hat{u})(t) = \sum_{\tau=-\infty}^{\infty} h(t,\tau)\hat{u}(\tau), \quad t \in \mathbb{R} \tag{7.6}$$

for $\hat{u} \in \ell^\infty$, where the series converges absolutely.

Let $\hat{\mathbf{T}}$ be the right shift operator in $\ell^\infty$,

$$(\hat{\mathbf{T}}\hat{x})(t) = \hat{x}(t-1), \quad t \in \mathbb{Z}, \ \hat{x} \in \ell^\infty;$$

and $\mathbf{T}$ the $T_p$-shift operator in $\mathbf{L}^\infty$,

$$(\mathbf{T}x)(t) = x(t - T_p), \quad t \in \mathbb{R}, \ x \in \mathbf{L}^\infty.$$

Then $\mathbf{S}$ and $\mathbf{H}$ satisfies

$$\mathbf{H}\hat{\mathbf{T}} = \mathbf{T}\mathbf{H}, \quad \mathbf{S}\mathbf{T} = \hat{\mathbf{T}}\mathbf{S}. \tag{7.7}$$

Moreover,

$$\mathbf{S}\mathbf{H} = \hat{\mathbf{T}}. \tag{7.8}$$

A sampler and hold pair define a mapping $\mathbf{D}$ from the set of continuous-time systems to the set of discrete-time systems by

$$\mathbf{D}(k) := \mathbf{S}\mathbf{K}\mathbf{H}. \tag{7.9}$$

We call this map *discretization map* and the discrete-time system $\hat{k} = \mathbf{D}(k)$ the *sampled system* of $k$.

**Proposition 7.1** $\mathbf{D}$ *is a contractive linear operator from* $\mathbf{L}^1$ *to* $\ell^1$, *i.e.,* $\|\mathbf{D}\| \leq 1$, *and for all* $k \in \mathbf{L}^1$,

$$\hat{k}(t) = \mathbf{D}(k)(t) = \int_{(t-1)T_p}^{tT_p} k(\tau)d\tau \quad t \in \mathbb{Z}. \tag{7.10}$$

**Proof** First, we show that if $k$ is shift invariant, then $\mathbf{D}(k)$ is also shift invariant.

$$\begin{aligned}
\widehat{\mathbf{K}}\hat{\mathbf{T}} &= \mathbf{S}\mathbf{K}\mathbf{H}\hat{\mathbf{T}} && \text{by (7.9)} \\
&= \mathbf{S}\mathbf{K}\mathbf{T}\mathbf{H} && \text{by (7.7)} \\
&= \mathbf{S}\mathbf{T}\mathbf{K}\mathbf{H} && \text{since } \mathbf{K} \text{ is shift invariant} \\
&= \hat{\mathbf{T}}\mathbf{S}\mathbf{K}\mathbf{H} && \text{by (7.7)} \\
&= \hat{\mathbf{T}}\widehat{\mathbf{K}} && \text{by (7.9),}
\end{aligned}$$

which proves that $\widehat{\mathbf{K}}$ is shift invariant.

Moreover, for $\hat{u} \in \ell^\infty$ and $t \in \mathbb{Z}$,

$$\begin{aligned}
\hat{y}(t) &= (\widehat{\mathbf{K}}\hat{u})(t) \\
&= (\mathbf{SKH}\hat{u})(t) \\
&= (\mathbf{KH}\hat{u})(tT_p) \quad \text{by the definition of } \mathbf{S} \\
&= [\mathbf{K}(\sum_{\tau=-\infty}^{\infty} h(\cdot,\tau)\hat{u}(\tau))](tT_p) \quad \text{by (7.6)} \\
&= \sum_{\tau=-\infty}^{\infty} [\mathbf{K}h(\cdot,\tau)\hat{u}(\tau)](tT_p) \quad \text{by linearity.}
\end{aligned}$$

However,

$$\begin{aligned}
[\mathbf{K}h(\cdot,\tau)\hat{u}(\tau)](tT_p) &= [\mathbf{K}h(\cdot,\tau)](tT_p)\hat{u}(\tau) \\
&= [\mathbf{K}h_0(\cdot - \tau T)](tT_p)\hat{u}(\tau) \quad \text{by the definition of } h \\
&= [\mathbf{K}h_0]((t-\tau)T)\hat{u}(\tau) \quad \text{since } \mathbf{K} \text{ is shift invariant} \\
&= \hat{k}(t-\tau)\hat{u}(\tau),
\end{aligned}$$

where

$$\begin{aligned}
\hat{k}(t) &= (\mathbf{K}h_0)(tT_p) \\
&= \int_{-\infty}^{\infty} k(tT_p - \tau)h_0(\tau)d\tau \\
&= \int_0^T k(tT_p - \tau)d\tau \\
&= \int_{(t-1)T_p}^{tT_p} k(\tau)d\tau.
\end{aligned}$$

To show $\|\mathbf{D}\| \le 1$, it is enough to show that for all $k \in \mathbf{L}^1$, $\|\hat{k}\|_{\ell^1} \le \|k\|_{\mathbf{L}^1}$. This follows directly from (7.10), as

$$\begin{aligned}
\|\hat{k}\|_{\ell^1} &= \sum_{t=0}^{\infty} |\hat{k}(t)| \\
&= \sum_{t=1}^{\infty} \left| \int_{(t-1)T_p}^{tT_p} k(\tau)d\tau \right|
\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{i=1}^{\infty} \int_{(i-1)T_p}^{iT_p} |k(\tau)|d\tau \\
&= \int_0^{\infty} |k(\tau)|d\tau \\
&= \|k\|_{\mathbf{L}^1}.
\end{aligned}
$$

□

The following assumptions are made about the identification problem:

(i) The a priori information about the true system is given in continuous-time, i.e., $k_{true} \in \mathbf{S}_a \subset \mathbf{L}^1$;

(ii) The identification experiment is performed on the sampled system $\hat{k}_{true}$, i.e., the input is generated by a computer as a sequence of real numbers and then applied to the true system through a zero-order hold; the output observations are collected through a sampler.

The objective of the identification is to construct an estimate of $k_{true}$ in a selected model set in $\mathbf{L}^1$ and give error bounds in the $\mathbf{L}^1$ norm.

One way to achieve this objective is as the follows:

(i) Transform the a priori uncertainty set $\mathbf{S}_a$ in continuous-time to an a priori uncertainty set $\hat{\mathbf{S}}_a$ in discrete-time by using the map $\mathbf{D}$, i.e., $\hat{\mathbf{S}}_a := \mathbf{D}(\mathbf{S}_a)$;

(ii) Select an optimal (or suboptimal) model set $\widehat{\mathbf{X}}_n$ for $\hat{\mathbf{S}}_a$ in $\ell^1$;

(iii) Choose an estimate $\hat{k}_{est}$ of the sampled system $\hat{k}_{true}$ in $\widehat{\mathbf{X}}_n$ and compute the error bounds in the $\ell^1$ norm by using algorithms for discrete-time system identification;

(iv) Construct an estimate and compute the error bounds in $\mathbf{L}^1$.

The last step of the above procedure is related to the following topic.

## 7.2 Inversion of the Discretization Map

It is not difficult to verify that $\mathbf{D}$ has a non-trivial null space in $\mathbf{L}^1$. Therefore, it is not invertible on $\mathbf{L}^1$. Nevertheless, if a subspace $\mathbf{X} \subset \mathbf{L}$ is in the complement of $Null(\mathbf{D})$, i.e., $\mathbf{X} \bigcap Null(\mathbf{D}) = 0$, then the restrication of $\mathbf{D}$ on $\mathbf{X}$ is invertible. This implies that if both $\hat{\mathbf{S}}_a$ and $\widehat{\mathbf{X}}_n$ are in $\mathbf{D}(\mathbf{X})$, then an estimate for $k_{true}$ can be taken as $k_{est} = (\mathbf{D}|_{\mathbf{X}})^{-1}(\hat{k}_{est})$, and the estimation error has bounds

$$\|\hat{k}_{est} - \hat{k}_{true}\|_{\ell^1} \leq \|k_{est} - k_{true}\|_{\mathbf{L}^1} \leq \|(\mathbf{D}|_{\mathbf{X}})^{-1}\| \|\hat{k}_{est} - \hat{k}_{true}\|_{\ell^1}. \tag{7.11}$$

One of such subspaces $\mathbf{X}$ is the set of band-limited systems. Shannon's sampling theorem states that if the frequency response of $k$ is band-limited and the sampling frequency is larger than the band-width of $k$, then $k$ can be exactly recovered from $\hat{k}$. Although in practical control problems, systems are never band-limited, a system can be approximately recovered if it is approximately band-limited. More generally, we have

**Proposition 7.2** *Let* $\mathbf{X} \subset \mathbf{L}^1$, $\mathbf{D}(\mathbf{X}) = \ell^1$, *and* $\mathbf{X} \bigcap Null(\mathbf{D}) = 0$. *If* $\mathbf{S}_a \subset \mathbf{X} + \mathbf{B}(\epsilon)$, *then for* $k \in \mathbf{S}_a$ *and* $k_{est} := (\mathbf{D}|_{\mathbf{X}})^{-1}(\hat{k}_{est})$, *we have*

$$\|\hat{k}_{est} - \hat{k}\|_{\ell^1} \leq \|k_{est} - k\|_{\mathbf{L}^1} \leq \|(\mathbf{D}|_{\mathbf{X}})^{-1}\| \left(\|\hat{k}_{est} - \hat{k}\|_{\ell^1} + \epsilon\right) + \epsilon. \tag{7.12}$$

**Proof** The first inequality follows from Proposition 7.1. To show the second one, we notice that

$$\|k_{est} - k\|_{\mathbf{L}^1} \leq \|k_{est} - (\mathbf{D}|_{\mathbf{X}})^{-1}\mathbf{D}(k)\|_{\mathbf{L}^1} + \|(\mathbf{D}|_{\mathbf{X}})^{-1}\mathbf{D}(k) - k\|_{\mathbf{L}^1}. \tag{7.13}$$

Since $k \in \mathbf{S}_a$, $\exists k_1 \in \mathbf{X}$, $k_2 \in \mathbf{B}(\epsilon)$ such that $k = k_1 + k_2$. It follows that

$$(\mathbf{D}|_{\mathbf{X}})^{-1}\mathbf{D}(k) = k_1 + (\mathbf{D}|_{\mathbf{X}})^{-1}\mathbf{D}(k_2). \tag{7.14}$$

Therefore,

$$\begin{aligned}
\|k_{est} - k\|_{\mathbf{L}^1} &\leq \|(\mathbf{D}|_{\mathbf{X}})^{-1}(\hat{k}_{est} - \hat{k})\|_{\mathbf{L}^1} + \|k_1 - k + (\mathbf{D}|_{\mathbf{X}})^{-1}\mathbf{D}(k_2)\|_{\mathbf{L}^1} && (7.15)\\
&\leq \|(\mathbf{D}|_{\mathbf{X}})^{-1}\|\|\hat{k}_{est} - \hat{k}\|_{\ell^1} + \|k_2\|_{\mathbf{L}^1} + \|(\mathbf{D}|_{\mathbf{X}})^{-1}\|\|\mathbf{D}(k_2)\|_{\ell^1} && (7.16)\\
&\leq \|(\mathbf{D}|_{\mathbf{X}})^{-1}\|\left(\|\hat{k}_{est} - \hat{k}\|_{\ell^1} + \epsilon\right) + \epsilon. && (7.17)
\end{aligned}$$

This completes the proof.

□

If $\mathbf{D}$ is seen as a mapping from $\mathbf{L}^\infty$ on the $j\omega$ axis to $\mathbf{L}^\infty$ on the unit circle, then Shannon's reconstruction procedure has unit norm on the space of band-limited systems, i.e., $\|(\mathbf{D}|_{\mathbf{X}})^{-1}\| = 1$.

A severe shortcoming of Shannon's reconstruction procedure is that it is not causal. In other words, it can not be performed until the complete set of sampled data becomes available.[2] A Nehari approximation problem can be solved to obtain a causal estimate.

Instead of a subspace of band-limited systems, we will take the inverse of $\mathbf{D}$ on a subspace of step functions. It can be shown that, if the system's impulse response is smooth (similar to approximately band-limited), then it can be approximated by a step function. Moreover, the inversion procedure on such a subspace is causal and the norm of the inverse $\|(\mathbf{D}|_{\mathbf{X}})^{-1}\|$ equals one.

**Proposition 7.3** *Let*

$$f_i(t) := \begin{cases} 1/T_p & iT_p < t \leq (i+1)T_p, \\ 0 & \textit{otherwise}, \end{cases} \tag{7.18}$$

*and* $\mathbf{X}_s := \overline{sp}\{f_0, f_1, \ldots\}$. *Then* $\mathbf{D}|_{\mathbf{X}_s} : \mathbf{X}_s \to \ell^1$ *is one-one and onto, and*

$$(\mathbf{D}|_{\mathbf{X}_s})^{-1}(\hat{k}) = \sum_{i=0}^{\infty} \hat{k}(i) f_i. \tag{7.19}$$

*Moreover,* $\|(\mathbf{D}|_{\mathbf{X}_s})^{-1}\| = 1$.

[2] It is well known that the ideal rectangle filter used in the Shannon reconstruction is not causal.

**Proof** To show $\mathbf{D}|_{\mathbf{X}_s}$ is one to one, it is enough to show that $\mathbf{D}|_{\mathbf{X}_s}(k) = 0$ implies that $k = 0$. By Proposition 7.1, if $\mathbf{D}|_{\mathbf{X}_s}(k) = 0$, then

$$\int_{iT_p}^{(i+1)T_p} k(t)dt = 0 \quad \forall i \in \mathbf{Z}_+. \tag{7.20}$$

Since $k \in \mathbf{X}_s$, $k = \sum_{i=0}^{\infty} c_i f_i$. (7.20) implies that $c_i = 0$ for all $i$, i.e., $k = 0$.

$\mathbf{D}|_{\mathbf{X}_s}$ is onto as for any $\hat{k} \in \ell^1$, there exists $k = \sum_{i=0}^{\infty} \hat{k}(i) f_i \in \mathbf{X}_s$ such that $\mathbf{D}(k) = \hat{k}$.

$\|(\mathbf{D}|_{\mathbf{X}_s})^{-1}\| = 1$ because for any $k = \sum_{i=0}^{\infty} c_i f_i \in \mathbf{X}_s$,

$$\begin{aligned} \|\mathbf{D}|_{\mathbf{X}_s}(k)\|_{\ell^1} &= \sum_{i=0}^{\infty} \left| \int_{iT_p}^{(i+1)T_p} k(t)dt \right| && (7.21)\\ &= \sum_{i=0}^{\infty} \int_{iT_p}^{(i+1)T_p} |k(t)|dt \quad \text{as } k \text{ is constant on } (iT_p, (i+1)T_p]. && (7.22)\\ &= \|k\|_{\mathbf{L}^1}. && (7.23) \end{aligned}$$

□

The above inversion procedure will enable us to relate a representation problem in the continuous-time case to one in the discrete-time case.

**Proposition 7.4** *If* $\mathbf{S}_a \subset \mathbf{X}_s + \mathbf{B}(\epsilon)$, *then*

$$d_n\left(\hat{\mathbf{S}}_a, \ell^1\right) \le d_n\left(\mathbf{S}_a, \mathbf{L}^1\right) \le d_n\left(\hat{\mathbf{S}}_a, \ell^1\right) + 2\epsilon. \tag{7.24}$$

**Proof** By definition of $d_n(\mathbf{S}_a, \mathbf{L}^1)$, $\forall \eta > 0$, there exists a $n$-dimensional subspace $\mathbf{X}_n \subset \mathbf{L}^1$ such that

$$dist(\mathbf{S}_a, \mathbf{X}_n) \le d_n\left(\mathbf{S}_a, \mathbf{L}^1\right) + \eta. \tag{7.25}$$

Let $\hat{\mathbf{X}}_n := \mathbf{D}(\mathbf{X}_n)$. Then

$$\begin{aligned} d_n\left(\hat{\mathbf{S}}_a, \ell^1\right) &\le dist\left(\hat{\mathbf{S}}_a, \hat{\mathbf{X}}_n\right) && (7.26)\\ &= \sup_{k \in \mathbf{S}_a} \inf_{h \in \mathbf{X}_n} \|\mathbf{D}(k - h)\|_{\ell^1} && (7.27)\\ &\le dist(\mathbf{S}_a, \mathbf{X}_n) \quad \text{by Proposition 7.2,} && (7.28)\\ &\le d_n\left(\mathbf{S}_a, \mathbf{L}^1\right) + \eta. && (7.29) \end{aligned}$$

Since this is true for all $\eta > 0$, we get the first inequality in (7.24).

By definition of $d_n\left(\hat{\mathbf{S}}_a, \ell^1\right)$, $\forall \eta > 0$, there exists a $n$-dimensional subspace $\widehat{\mathbf{X}}_n \subset \ell^1$ such that

$$dist\left(\hat{\mathbf{S}}_a, \widehat{\mathbf{X}}_n\right) \leq d_n\left(\hat{\mathbf{S}}_a, \ell^1\right) + \eta. \tag{7.30}$$

Set $\mathbf{X}_n = (\mathbf{D}|_{\mathbf{X}_s})^{-1}(\widehat{\mathbf{X}}_n)$. Then

$$\begin{aligned} d_n\left(\mathbf{S}_a, \mathbf{L}^1\right) &\leq dist\left(\mathbf{S}_a, \mathbf{X}_n\right), & (7.31)\\ &= \sup_{k\in\mathbf{S}_a} \inf_{h\in\mathbf{X}_n} \left\|k - (\mathbf{D}|_{\mathbf{X}_s})^{-1}\mathbf{D}(h)\right\|_{\mathbf{L}^1}. & (7.32)\end{aligned}$$

By Proposition 7.2 and 7.3, for all $k \in \mathbf{S}_a$,

$$\left\|k - (\mathbf{D}|_{\mathbf{X}_s})^{-1}\mathbf{D}(h)\right\|_{\mathbf{L}^1} \leq \|\hat{k} - \hat{h}\|_{\ell^1} + 2\epsilon. \tag{7.33}$$

Therefore,

$$\begin{aligned} d_n\left(\mathbf{S}_a, \mathbf{L}^1\right) &\leq \sup_{k\in\mathbf{S}_a} \inf_{h\in\mathbf{X}_n} \|\hat{k} - \hat{h}\|_{\ell^1} + 2\epsilon. & (7.34)\\ &\leq d_n\left(\hat{\mathbf{S}}_a, \ell^1\right) + 2\epsilon + \eta. & (7.35)\end{aligned}$$

Since the above inequalities hold for all $\eta > 0$, we get the second inequality in (7.24).

□

## 7.3 An Example

Consider the identification of a continuous-time system in

$$\mathbf{S}_a := \left\{k \in \mathbf{L}^1 : \ |k(\tau)| \leq r^\tau, \ k \textit{ is abs. cont. and } \|\dot{k}\|_1 \leq c\right\} \tag{7.36}$$

To show that this problem can be reduced to the discrete-time case, it is enough to show that $\mathbf{S}_a$ is contained in $\mathbf{X}_s + \mathbf{B}(2T_pc)$, i.e., for all $k \in \mathbf{S}_a$, $\exists k_1 \in \mathbf{X}_s$ such that $\|k - k_1\|_{\mathbf{L}^1} \leq 2T_pc$.

For $k \in \mathbf{S}_a$, set $k_1 = (\mathbf{D}|_{\mathbf{X}_*})^{-1}\mathbf{D}(k)$. To prove that $\|k - k_1\|_{\mathbf{L}^1} \leq 2T_p c$, we notice that

$$k_1(t) = \frac{1}{T_p}\hat{k}(m) = \frac{1}{T_p}\int_{(m-1)T_p}^{mT_p} k(\tau)d\tau \qquad t \in (mT, (m+1)T]. \tag{7.37}$$

As a result,

$$\begin{aligned}
\|k - k_1\|_{\mathbf{L}^1} &= \int_0^\infty |k(t) - k_1(t)|dt \\
&= \sum_{m=0}^{\infty} \int_{mT_p}^{(m+1)T_p} |k(t) - k_1(t)|dt \\
&= \sum_{m=0}^{\infty} \int_{mT_p}^{(m+1)T_p} \left| k(t) - \frac{1}{T_p}\int_{(m-1)T_p}^{mT_p} k(\tau)d\tau \right| dt \\
&= \sum_{m=0}^{\infty} \int_{mT_p}^{(m+1)T_p} \frac{1}{T_p} \left| \int_{(m-1)T_p}^{mT_p} (k(t) - k(\tau))d\tau \right| dt. 
\end{aligned} \tag{7.38}$$

However,

$$\begin{aligned}
\int_{mT_p}^{(m+1)T_p} & \frac{1}{T_p} \left| \int_{(m-1)T_p}^{mT_p} (k(t) - k(\tau))d\tau \right| dt \\
&= \int_{mT_p}^{(m+1)T_p} \frac{1}{T_p} \left| \int_{(m-1)T_p}^{mT_p} \int_\tau^t \frac{dk(\tau')}{d\tau'} d\tau' d\tau \right| dt \\
&\leq \int_{mT_p}^{(m+1)T_p} \frac{1}{T_p} \int_{(m-1)T_p}^{mT_p} \int_{(m-1)T_p}^{(m+1)T_p} \left| \frac{dk(\tau')}{d\tau'} \right| d\tau' d\tau dt \\
&= T_p \int_{(m-1)T_p}^{(m+1)T_p} \left| \frac{dk(\tau')}{d\tau'} \right| d\tau'.
\end{aligned}$$

This implies that

$$\begin{aligned}
\|k - k_1\|_{\mathbf{L}^1} &\leq \sum_{m=0}^{\infty} T_p \int_{(m-1)T_p}^{(m+1)T_p} \left| \frac{dk(\tau')}{d\tau'} \right| d\tau' \\
&\leq 2T_p \|\dot{k}\|_{\mathbf{L}^1} \\
&\leq 2T_p c.
\end{aligned} \tag{7.39}$$

Therefore, by Proposition 7.2 and Proposition 7.4, the representation and estimation problem for $k \in \mathbf{S}_a$ can be done via the procedure proposed in Section 7.1

# Chapter 8

# The Intrinsic Uncertainty in Identification of Time-Varying Systems

In this chapter, we will apply the results for fast identification derived in the previous chapters to the identification of time-varying systems. The systems will be represented by their Volterra kernels. It will be shown that there is an irreducible identification error when a system's present behavior does not determine it future behavior completely. The faster the system's possible variation, the larger the error. Later, the identification of slowly time-varying systems will be discussed.

In this chapter, we will only deal with the discrete-time case. The continuous-time counterpart can be derived easily using the results in Chapter 6 and 7. This chapter is based on [24, 58].

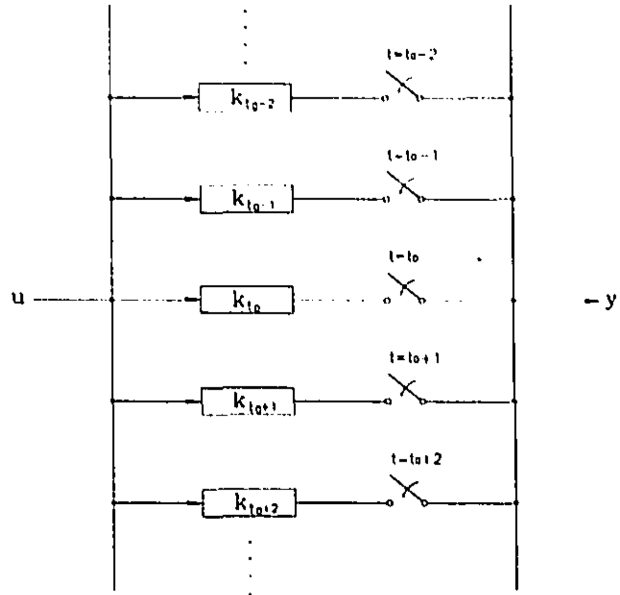


Figure 8.1: Linear time-varying systems represented by Volterra sums

## 8.1 Representation of Time-Varying Systems by Volterra Sums

Time-varying systems will be represented by Volterra sum operators, $\mathbf{K} : \mathbf{U} \to \mathbf{Y}$, where (as shown in Figure 8.1)

$$y(t) = \sum_{\tau=0}^{\infty} k(t,\tau)u(t-\tau), \quad t,\tau \in \mathbb{Z}. \tag{8.1}$$

Here, as in the time invariant case, $\mathbf{U}$ and $\mathbf{Y}$ are normed linear spaces contained in the set $\ell^\infty(-\infty,\infty)$. A distinction will be made between kernels $k(\cdot,\cdot) : \mathbb{Z} \times \mathbb{Z}_+ \to \mathbb{R}$ and the weighting functions that these kernel induce, denoted by $k_t(\cdot)$, $t \in \mathbb{Z}$, which satisfy $k_t(\tau) := k(t,\tau)$, $k_t(\cdot) : \mathbb{Z}_+ \to \mathbb{R}$. It will be assumed that kernels $k(\cdot,\cdot)$ belong to a normed algebra $\mathbf{B}$ satisfying the conditions,

$$\mathbf{B} := \left\{ k(\cdot,\cdot) : \; k_t(\cdot) \in \mathbf{L} \;\; \forall t \in \mathbb{Z} \text{ and } \sup_{t\in\mathbb{Z}} \|k_t(\cdot)\|_{\ell^1} < \infty \right\}, \tag{8.2}$$

where $\mathbf{L}$ is a convolution algebra defined as in the time invariant case. The norm on $\mathbf{B}$ is

$$\|k(\cdot,\cdot)\|_{\mathbf{B}} := \sup_{t\in\mathbf{Z}} \|k_t(\cdot)\|_{\mathbf{L}}. \tag{8.3}$$

The product in $\mathbf{B}$ of two kernels in $\mathbf{B}$ is the kernel of the product operator.

Identification will be considered in the $\mathbf{L}$-norm for weighting functions and/or the $\mathbf{B}$-norm for the kernels. The $\mathbf{B}$-norm is a natural choice where it coincides with the operator norm of the time-varying operator, as in the case where $\mathbf{L} = \ell^1$. More generally, the precise operator norm may be intractable, but the $\mathbf{B}$-norm is nevertheless suitable for the "frozen-time" analysis of systems, as in [58, 49]. There, the systems vary slowly with time or "approximately commute with the shift", and the $\mathbf{B}$-norm is an upper bound on the operator norms of the local[1] operators.

The rate of change $\rho$ of such a system is defined to be

$$\rho(k) = \sup_{t\in\mathbf{Z}} \|k_t(\cdot) - k_{t-1}(\cdot)\|_{\mathbf{L}} \tag{8.4}$$

For any[2] subset $\tilde{\mathbf{S}} \subset \mathbf{B}$, $\rho\left(\tilde{\mathbf{S}}\right) := \sup_{k\in\tilde{\mathbf{S}}} \rho(k)$.

Suppose that the a priori information concerning a system locates its weighting functions $k_t(\cdot)$ in a set $\mathbf{S}_a \subset \mathbf{L}$ and limits its rate of change, but does not otherwise constrain the manner in which it changes with time, i.e.,

$$\tilde{\mathbf{S}}_a = \{k \in \mathbf{B} :\ k_t \in \mathbf{S}_a \subset \mathbf{L}\ \forall t \in \mathbb{Z},\ and\ \rho(k) \le c < \infty\}, \tag{8.5}$$

which implies that $\rho\left(\tilde{\mathbf{S}}_a\right) = c < \infty$. Here $\mathbf{S}_a$ again satisfies Assumption 1 of Chapter 2 and is therefore a closed convex symmetric set.

[1] The local operator of $\mathbf{K}$ at time $t$ is the time-invariant operator $\mathbf{K}_t$ with the impulse response $k_t(\cdot)$.

[2] We will use capitals with tilde, such as $\tilde{\mathbf{S}}$, to denote set of kernels in $\mathbf{B}$, and capitals without tilde to denote the sets of corresponding weighting functions in $\mathbf{L}$.

## 8.2 Uncertainty Principles

First, we consider the identification of the weighting functions $k_t(\cdot)$ in the $\mathbf{L}$-norm. To get the most general lower bounds to uncertainty, assume that the entire histories of the input $u$ and output $y$ on $(-\infty, \infty)$ are known and there is no measurement noise. (Otherwise, a greater lower bound is possible.) Based on these observations of $u$ and $y$ the location of the true kernel $k_{true}$ is narrowed down, as in the time-invariant case, from $\tilde{\mathbf{S}}_a$ to a smaller set

$$\tilde{\mathbf{S}}\left(k_{true}\right) := \left\{k \in \tilde{\mathbf{S}}_a :\ \left(\mathbf{K}u - \mathbf{K}_{true}u\right)(t) = 0\ \forall t \in \mathbb{R}\right\}, \tag{8.6}$$

and the uncertainty in the corresponding weighting function at time $t_0$, is reduced from $\mathbf{S}_a$ to the subset

$$\mathbf{S}\left(k_{true}, t_0\right) := \left\{k_{t_0} \in \mathbf{S}_a :\ k_{t_0}(\cdot) = k(t_0, \cdot)\ \textit{for some}\ k(\cdot,\cdot) \in \tilde{\mathbf{S}}\left(k_{true}\right)\right\}, \tag{8.7}$$

Again, to get the most general lower bound, we assume that the nominal system for $k_{t_0}$ can be any system in $\mathbf{L}$. As in the time invariant case, the worst-case uncertainty in identifying the weighting function at $t_0$, for an optimally chosen estimate $(k_{est})_{t_0}$, is

$$e(u, t_0) := \sup_{(k_{true})_{t_0} \in \mathbf{S}_a} \ \inf_{(k_{est})_{t_0} \in \mathbf{L}} \ \sup_{k_{t_0} \in \mathbf{S}(k_{true}, t_0)} \left\|k_{t_0} - (k_{est})_{t_0}\right\|_{\mathbf{L}}, \tag{8.8}$$

and is a function of the input $u$. We would like to relate this uncertainty optimized over all inputs, i.e.,

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0\right) := \inf_{u \in \mathbf{U}} e(u, t_0), \tag{8.9}$$

to the n-width $\theta^n\left(\mathbf{S}_a, \mathbf{L}\right)$, and show that if the rate $\rho\left(\tilde{\mathbf{S}}_a\right)$ is greater than zero then there is an irreducible uncertainty in identification no matter what the input. For this we will need the following lower bound on $e(u, t_0)$ whose derivation is similar to (2.14) for the time-invariant case,

$$e(u, t_0) \geq \sup\left\{\|k_{t_0}\|_{\mathbf{L}} :\quad k_{t_0}(\cdot) = k(t_0, \cdot)\ \textit{for some}\ k \in \tilde{\mathbf{S}}_a\right.$$

$$\textit{for which } (\mathbf{K}u)(t) = 0 \ \forall t \in \mathbb{R}\} \tag{8.10}$$

Given $\theta^n(\mathbf{S}_a, \mathbf{L})$, introduce the function $\phi$ mapping the positive integers $\{1, 2, \ldots\}$ into $\mathbb{R}_+ \bigcup \{\infty\}$, $\phi(n) := \frac{1}{n}\theta^{2n-1}$. Then $\phi(n)$ is monotone decreasing in $n$. Let $\phi^{-1}$ be the inverse relation $\phi^{-1} : \mathbb{R}_+ \to \mathbb{Z}_+$,

$$\phi^{-1}(x) := \inf\{n : \ n > 0, \ \phi(n) \leq x\},$$

which is also monotone decreasing .

**Theorem 8.1** *If the a priori uncertainty set $\tilde{\mathbf{S}}_a$ has a rate of change $\rho\left(\tilde{\mathbf{S}}_a\right)$, then the optimal worst-case uncertainty in identification of the weighting function at time $t_0$, $\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0\right)$, has the lower bound*

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0\right) \geq \theta^{\left(2\phi^{-1}(\rho) \ - \ 1\right)} \qquad \forall t_0 \in \mathbb{Z} \tag{8.11}$$

*where $\theta := \theta(\mathbf{S}_a, \mathbf{L})$ and $\rho := \rho\left(\tilde{\mathbf{S}}_a\right)$.*

**Proof** By (8.10), it is enough to show that for all $u \in \mathbf{U}$ and $t_0 \in \mathbb{Z}$, there exists a null kernel, $k \in \tilde{\mathbf{S}}_a \bigcap Null(\Phi_u)$ whose frozen-time system $k_{t_0}$ is appropriately large. Choose $n = \phi^{-1}(\rho)$. By definition of $\theta^{2n-1}$, given $\epsilon > 0$, $u \in \mathbf{U}$, and $t_0 \in \mathbb{Z}$, there is an impulse response $k_{t_0} \in \mathbf{S}_a$ for which the (time-invariant) system operator $\mathbf{K}_{t_0}$ satisfies $(\mathbf{K}_{t_0}u)(t) = 0$ for $t_0 - n < t < t_0 + n$, and

$$\theta^{(2n-1)} - \epsilon \leq \|k_{t_0}\|_{\mathbf{L}} \leq \theta^{(2n-1)}. \tag{8.12}$$

Define $k \in \mathbf{B}$,

$$k(t, \cdot) := \begin{cases} k_{t_0}(\cdot)\left(1 - \frac{|t_0 - t|}{n}\right) & \textit{if } t_0 - n < t < t_0 + n, \\ 0 & \textit{elsewhere.} \end{cases} \tag{8.13}$$

The resulting (time-varying) operator $\mathbf{K}$ is null, i.e., $(\mathbf{K}u) = 0$, and $k(t, \cdot) \in \mathbf{S}_a$. Also, the choice $n = \phi^{-1}(\rho)$, together with (8.12) and (8.13) imply that $\mathbf{K}$ has an appropriate rate, i.e., $\rho(\mathbf{K}) \leq \frac{\theta^{2n-1}}{n} \leq \rho\left(\tilde{\mathbf{S}}_a\right)$. Hence $k \in \tilde{\mathbf{S}}_a \bigcap Null(\Phi_u)$. Finally,

$$\begin{aligned} e(u, t_0) &\geq \|k_{t_0}\|_{\mathbf{L}} \quad (\text{by } (8.10)), \\ &\geq \theta^{2n-1} - \epsilon, \quad (\text{by } (8.12)) \\ &= \theta^{2\phi^{-1}(\rho)-1} - \epsilon, \quad \left(\phi^{-1}(\rho) = n\right), \end{aligned}$$

and since this holds for all $\epsilon > 0$ and $u \in \mathbf{U}$, the theorem follows.

□

In fact, the uncertainty is greater than zero not only in the worst-case but for any system which is not too close to the boundary of $\tilde{\mathbf{S}}_a$. This is shown in the next corollary to Theorem 8.1. Denote the uncertainty for a weighting function $(k_{true})_{t_0}$ by

$$e(u, t_0, k_{true}) := \inf_{(k_{est})_{t_0} \in \mathbf{L}} \ \sup_{k_{t_0} \in \mathbf{S}(k_{true}, t_0)} \|k_{t_0} - (k_{est})_{t_0}\|_{\mathbf{L}}. \tag{8.14}$$

Also, denote the optimal uncertainty for $(k_{true})_{t_0}$ by,

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0, k_{true}\right) := \inf_{u \in \mathbf{U}} e(u, t_0, k_{true}). \tag{8.15}$$

**Corollary 8.1** *If* $(k_{true})_t \in \alpha \mathbf{S}_a \ \ \forall t \in \mathbb{Z}$, *and* $\rho(k_{true}) \leq \beta \rho\left(\tilde{\mathbf{S}}_a\right)$ *for some* $0 \leq \alpha < 1$, $0 \leq \beta < 1$, *then for all* $t_0 \in \mathbb{Z}$,

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0, k_{true}\right) \geq (1-\alpha)\theta^{\left\{2\phi^{-1}[(1-\beta)\rho] \ - \ 1\right\}}, \tag{8.16}$$

*where* $\theta := \theta(\mathbf{S}_a, \mathbf{L})$, $\rho := \rho\left(\tilde{\mathbf{S}}_a\right)$ *and* $\alpha\tilde{\mathbf{S}}_a := \left\{\alpha k : \ k \in \tilde{\mathbf{S}}_a\right\}$.

**Proof** It will be shown later that a closed convex symmetric set $\tilde{\mathbf{S}}_b \subset \mathbf{B}$ can be found such that $\left(k_{true} + \tilde{\mathbf{S}}_b\right) \subset \tilde{\mathbf{S}}_a$. It will follow that for all $u \in \mathbf{U}$

$$e(u, t_0, k_{true}) \geq \sup\left\{\|k_{t_0}\|_{\mathbf{L}} : \ k_{t_0}(\cdot) = k(t_0, \cdot) \ for \ some \ k \in \tilde{\mathbf{S}}_b \ for \ which \ \mathbf{K}u = 0\right\}. \tag{8.17}$$

To see that (8.17) will follow, note that $\left(k_{true}+\tilde{\mathbf{S}}_b\right)\subset\tilde{\mathbf{S}}_a$ implies that the set in (8.7) satisfies

$$\begin{aligned}\mathbf{S}\left(k_{true},t_0\right) \quad &\supset \quad \left\{k_{t_0}\in\mathbf{L}:\ k_{t_0}(\cdot)=k(t_0,\cdot)\ for\ some\ k\in\left(k_{true}+\tilde{\mathbf{S}}_b\right)\right. \\ &\qquad\qquad \left. for\ which\ \left(\mathbf{K}u-\mathbf{K}_{true}u\right)=0\right\} &(8.18)\\ &=:\quad \mathbf{S}_b\left(k_{true},t_0\right). &(8.19)\end{aligned}$$

Therefore, by (8.14),

$$e(u,t_0,k_{true})\geq\inf_{(k_{est})_{t_0}\in\mathbf{L}}\ \sup_{k_{t_0}\in\mathbf{S}_b(k_{true},t_0)}\left\|k_{t_0}-\left(k_{est}\right)_{t_0}\right\|_{\mathbf{L}}, \tag{8.20}$$

Since $\mathbf{S}_b\left(k_{true},t_0\right)$ is convex and symmetric around $\left(k_{true}\right)_{t_0}$, the optimal choise for $\left(k_{est}\right)_{t_0}$ for the right hand side of (8.20) is $\left(k_{true}\right)_{t_0}$, i.e.,

$$e(u,t_0,k_{true})\geq\sup_{k_{t_0}\in\mathbf{S}_b(k_{true},t_0)}\left\|k_{t_0}-\left(k_{true}\right)_{t_0}\right\|_{\mathbf{L}} \tag{8.21}$$

which implies (8.17) by definition of $\mathbf{S}_b\left(k_{true},t_0\right)$.

An appropriate set $\tilde{\mathbf{S}}_b$ is the subset of $\tilde{\mathbf{S}}_a$,

$$\tilde{\mathbf{S}}_b:=\left\{k:\ k_t\in(1-\alpha)\mathbf{S}_a\ \forall t\in\mathbb{Z},\ \rho(k)\leq(1-\beta)\rho\left(\tilde{\mathbf{S}}_a\right)\right\}$$

for then $k\in\tilde{\mathbf{S}}_b$ implies $\left(k+k_{true}\right)_t\in\left((1-\alpha)\mathbf{S}_a+\alpha\mathbf{S}_a\right)\subset\mathbf{S}_a\quad\forall t\in\mathbb{Z}$, and $\rho\left(k+k_{true}\right)\leq(1-\beta)\rho\left(\tilde{\mathbf{S}}_a\right)+\beta\rho\left(\tilde{\mathbf{S}}_a\right)=\rho\left(\tilde{\mathbf{S}}_a\right)$, which implies that $k+k_{true}\in\tilde{\mathbf{S}}_a$. Since (8.17) holds for all $u\in\mathbf{U}$, applying the method of proof of Theorem 8.1 to $\tilde{\mathbf{S}}_b$, we get the proposition.

□

When $\alpha=\beta=0$, the lower bound for $\Delta\left(\tilde{\mathbf{S}}_a,\mathbf{L},t_0,k_{true}\right)$ is the same as the one for the optimal worst-case uncertainty $\Delta\left(\tilde{\mathbf{S}}_a,\mathbf{L},t_0\right)$ in the theorem.

Now we consider the identification of a time-varying kernel in the normed space **B**. Based on the observations on the infinite interval $(-\infty,\infty)$, uncertainty as to the true kernel is reduced to $\tilde{\mathbf{S}}(k_{true})$ as in (8.6). The worst-case uncertainty for the optimally chosen estimate $k_{est}$ in the $\mathbf{B}$-norm is

$$e(u) := \sup_{k_{true} \in \tilde{\mathbf{S}}_a} \inf_{k_{est} \in \mathbf{B}} \sup_{k \in \tilde{\mathbf{S}}(k_{true})} \|k - k_{est}\|_{\mathbf{B}}. \tag{8.22}$$

Since the $\mathbf{B}$-norm of a time-varying kernel is the supremum of the $\mathbf{L}$-norms of its weighting functions, it is not difficult to show that $e(u) \geq e(u, t_0)$ for all $t_0$. Therefore, the lower bound in Theorem 8.1 is also a lower bound to the optimal worst-case uncertainty of the time-varying kernel $\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}\right) := \inf_{u \in \mathbf{U}} e(u)$; i.e.,

**Corollary 8.2** *Under the hypotheses of Theorem 8.1,*

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}\right) \geq \theta^{\left(2\phi^{-1}(\rho) - 1\right)}. \tag{8.23}$$

A result similar to Corollary 8.1 is also easy to obtain for time-varying kernels.

**Example 8.1** In the following we assume that the sets $\mathbf{S}_{a_i}$ $i = 1, 2, 3, 4$, are defined as in Example 2.1; $(k_{true})_t \in \alpha \mathbf{S}_a$ $\forall t \in \mathbb{Z}$; and $\rho(k_{true}) \leq \beta\rho\left(\tilde{\mathbf{S}}_a\right)$. For fixed $r > 0$, let $\psi$ be the function $\psi : [0, \infty) \to \mathbb{Z}_+$,

$$\psi(x) := \inf\left\{n : \frac{r^{2n-1}}{n} \leq x\right\}.$$

(i) If $\mathbf{S}_a = \mathbf{S}_{a_1}$, then it has been claimed that $\theta^{2n-1} = \frac{C}{1-r} r^{2n-1}$. By Corollary 8.1, for all $t_0 \in \mathbb{Z}$,

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0, k_{true}\right) \geq (1 - \alpha) \frac{C}{1 - r} r^{\left\{2\psi\left[(1-\beta)\frac{\rho}{C}(1-r)\right] - 1\right\}}.$$

(ii) If $\mathbf{S}_a = \mathbf{S}_{a_2}$ or $\mathbf{S}_a = \mathbf{S}_{a_3}$, then $\theta^{2n-1} = Cr^{2n-1}$. Hence, for all $t_0 \in \mathbb{Z}$,

$$\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0, k_{true}\right) \geq (1 - \alpha) C r^{\left\{2\psi\left[(1-\beta)\frac{\rho}{C}\right] - 1\right\}}.$$

(iii) If $\mathbf{S}_a = \mathbf{S}_{a_4}$, then $\theta^{2n-1} = \frac{C}{2n-1}$. By definition of $\phi$, $\phi(n) = \frac{C}{n(2n-1)}$. It follows that the inverse function

$$\begin{aligned}\phi^{-1}((1-\beta)\rho) &= \inf\left\{n \in \mathbb{Z}_+ : \frac{C}{n(2n-1)} \leq (1-\beta)\rho\right\} \\ &\leq \inf\left\{n \in \mathbb{Z}_+ : 2n-1 \geq \frac{C}{(1-\beta)\rho}\right\} \\ &\leq \frac{C}{2(1-\beta)\rho} + 3/2.\end{aligned}$$

As $\theta^n = \frac{C}{n}$, we have $\theta^{\{2\phi^{-1}[(1-\beta)\rho] - 1\}} = C\{2\phi^{-1}[(1-\beta)\rho] - 1\}^{-1}$. Corollary 8.1 now gives

$$\begin{aligned}\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0, k_{true}\right) &\geq \left(\frac{1-\alpha}{2}\right)\left(\frac{\phi^{-1}((1-\beta)\rho)}{C} - \frac{1}{2C}\right)^{-1} && \forall t_0 \in \mathbb{Z} \\ &\geq \left(\frac{1-\alpha}{2}\right)\left(\frac{1}{2(1-\beta)\rho} + \frac{1}{C}\right)^{-1} && \forall t_0 \in \mathbb{Z}.\end{aligned}$$

**Remarks on prediction uncertainty.** It might be expected that for quickly changing uncertain plants, observations from the remote past should contribute little to the identification of the present weighting function; i.e., the useful observation interval should get shorter as the time variation rate $\rho(\tilde{\mathbf{S}}_a)$ increases. This is borne out by Theorem 8.1 and the examples, which show that the optimal identification error is bounded below by a monotone increasing function of the time variation rate $\rho\left(\tilde{\mathbf{S}}_a\right)$. The error in predicting $(k_{true})_{t_0+1}$ from observations of $y$ on $(-\infty, t_0]$ is bounded below by $\Delta\left(\tilde{\mathbf{S}}_a, \mathbf{L}, t_0, k_{true}\right) + \rho$ at least[3]. If that bound exceeds $\|\mathbf{S}_a\|_{\mathbf{L}}$, then identification provides no information about future behavior, and it becomes impossible to construct a model with any predictive power. This happens whenever the rate $\rho$ satisfies $\rho \geq \rho_{max}$, where $\rho_{max}$ is the solution of

$$(1-\alpha)\,\theta^{\{2\phi^{-1}[(1-\beta)\rho_{max}] - 1\}} + \rho_{max} = \|\mathbf{S}_a\|_{\mathbf{L}},$$

where $k_{true}$ satisfies the hypotheses of the corollary.

[3]A greater lower bound can be obtained by exploiting the fact that observations are available only on $(-\infty, t_0)$.

## 8.3 Comments on Identification of Slowly Varying Systems

To estimate a time-varying system $k$ accurately in the $\mathbf{B}$ norm, it is necessary to estimate each frozen time system accurately in the $\mathbf{L}$ norm. For a fixed $t_0$, the frozen system $k_{t_0}$ is a time-invariant system. If we were able to carry out an input-output experiment on this frozen system, we could use any available time-invariant identification algorithm to construct a model for $k_{t_0}$. Unfortunately, this is not feasible. An experiment on the time-varying system can only give us the output of $k_{t_0}$ at time $t_0$, as shown in Figure 8.1. Nevertheless, if the system is known to be slowly time varying, we can estimate the output of $k_{t_0}$, $y_{t_0}(t)$, from the output $y(t)$ of $k$.

**Proposition 8.1** *Let* $k \in \mathbf{B}$ *and* $\|k(t,\cdot) - k(t+1,\cdot)\|_1 \leq \rho\ \forall t \in \mathbf{Z}$. *For* $u \in \ell^\infty(-\infty,\infty)$, *let*

$$y(t) = \sum_{\tau=0}^{\infty} k(t,\tau)u(t-\tau), \tag{8.24}$$

$$y_{t_0}(t) = \sum_{\tau=0}^{\infty} k(t_0,\tau)u(t-\tau). \tag{8.25}$$

*Then* $\forall t, t_0 \in \mathbf{Z}$

$$|y(t) - y_{t_0}(t)| < \rho|t-t_0|\|u\|_\infty. \tag{8.26}$$

**Proof** For simplicity, we prove the lemma for $t \leq t_0$ only.

$$\begin{aligned}
|y(t) - y_{t_0}(t)| &= |\sum_{\tau=0}^{\infty}(k(t,\tau) - k(t_0,\tau))u(t-\tau)|, && (8.27)\\
&\leq \sum_{\tau=0}^{\infty} |k(t,\tau) - k(t_0,\tau)|\|u\|_\infty, && (8.28)\\
&= \|k(t,\cdot) - k(t_0,\cdot)\|_1\|u\|_\infty, && (8.29)\\
&\leq \sum_{i=0}^{t_0-t+1} \|k(t+i,\cdot) - k(t+i+1,\cdot)\|_1\|u\|_\infty, && (8.30)\\
&\leq \sum_{i=0}^{t_0-t+1} \rho\|u\|_\infty, && (8.31)
\end{aligned}$$

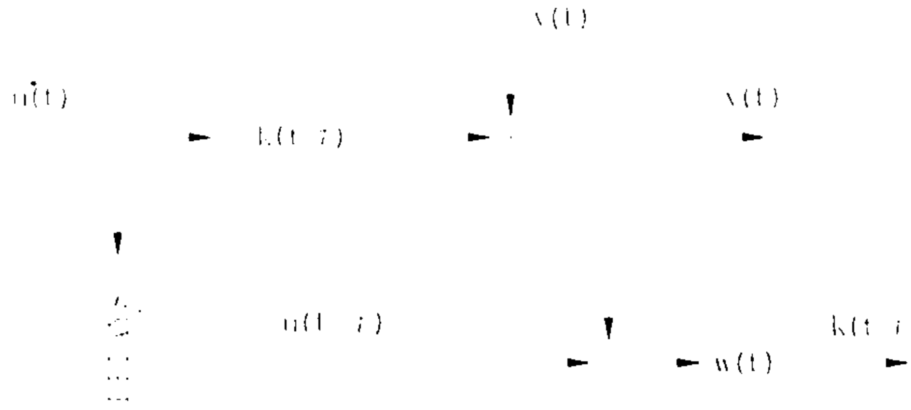

Figure 8.2: LTV system identification by smoothing

$$= \rho \|u\|_\infty (t_0 - t). \tag{8.32}$$

□

This Proposition indicates that if a system is slowly time varying, then the output of the system $y$ is close to the output $y_{t_0}$ of the frozen time system $k_{t_0}$ around $t_0$. Therefore, a window emphasising the output around $t_0$ should be applied in identifying the frozen time system $k_{t_0}$. When a rectangular window is used, the time-varying system identification problem is reduced to a sequence of time-invariant system identification problems on a sliding finite-time interval, and all the algorithms derived in Chapter 5 can be applied. In particular, the algorithm estimating the impulse response via inner products of the output and delayed inputs has the advantage of easy implementation. The impulse responses of the frozen time systems are obtained by filtering the products of the output and the delayed inputs with time-varying rectangular windows, as shown in Figure 8.2.

# Chapter 9

# Concluding Remarks

## 9.1 Summary of the Work

Motivated by the problems of robust adaptive control, we posed the problem of fast identification in this thesis. Time complexity and algorithmic complexity, two complexity issues related to fast identification, have been studied. The minimal time needed to identify a system to a specified accuracy in input-output behavior has been studied for both discrete and continuous-time cases. That time has been shown to depend on the metric complexity of the a priori data set, as measured by the Gel'fand n-width. The complexity of an identification algorithm was measured by the minimum number of parameters the algorithm needs to estimate in order to obtain a representation of the a posteriori information within a specified accuracy. This minimum number depends on the Kolmogorov n-width of the a priori data set when the model sets are restricted to be affine models. It is the relationship between identification speed and complexity that makes it fruitful to pose identification problems in the context of complexity theory. In that context, identification and feedback both serve the common purpose of reducing plant uncertainty and thereby reducing complexity.

In the discrete-time case, for a class of monotone decreasing data sets, it has been shown that the optimal input to achieve that minimal time is an impulse at the start of the observation interval, and the optimal affine model to achieve that simplest representation is the FIR model. In the continuous-time case, a class of optimal inputs has been characterized by a logarithmic integral condition using the theory of quasianalytic functions. A suboptimal affine representation for continuous-time data sets has been shown to be obtainable from the optimal representation of their sampled data systems. The optimality of such suboptimal representations is related to the smoothness of the impulse responses and the sampling period.

The input design problem has also been studied in a case where only certain ensemble properties of the input can be designed. Upper and lower bounds to the inherent error have been given in terms of the gap metric in general, and in terms of the spectrum flatness in several special cases. Two robust identification algorithms have also been proposed.

If a system changes while it is being identified, then there is an irreducible uncertainty as to its input-output behavior, which has been related to its rate of change. The irreducible identification errors derived here exist even if there is no additive noise. This intrinsic property of time-varying systems indicates that adaptive controllers must be designed for a set of systems, i.e., must be *robust*.

## 9.2 Directions for Future Research

Certain aspects of the research reported in this thesis are worth further investigation. Further research is contemplated on the following topics:

1. **The effects of stochastic additive noise on identification speed.** In Chapter 2 and Chapter 3, we concentrated on the effects of lack of data on the inherent error. The optimal inherent error was sought in the limiting case where the signal to noise ratio tends to infinity. When the signal to noise ratio

is limited, the results provided lower bounds on the best achievable inherent errors. It is unclear how the stochastic additive noise would affect the best achievable inherent error, and how one should exploit the stochastic property of the additive noise in input design. The results on ensemble input design should shed some light on these questions.

2. **Formulation of the fast identification problem in the stochastic setting.** In Chapter 2, we formulated the fast identification problem in the worst-case setting. A similar formulation can be achieved in the stochastic setting by embedding both system uncertainty and disturbance uncertainty into a probabilistic framework.

3. **Estimation of $\bar{\theta}^n$.** In Chapter 3, the upper and lower bounds on the n-width $\bar{\theta}^n$ have been given for several data sets. A challenging technical problem is to obtain the exact n-width $\bar{\theta}^n$.

4. **An information interpretation of the logarithmic integral condition.** In Chapter 6, a class of optimal inputs are characterized by a logarithmic integral condition. It is often speculated that logarithmic integrals are related to information theory. An interpretation of the results in Chapter 6 may reveal certain intrinsic relations between identification and information theory.

Besides the above direct extensions, the results in this thesis should be relevant to several long-term research topics, e.g., a unified theory of feedback and identification, an information-based adaptive control theory, optimal adaptive strategies based on optimal identification and optimal feedback, and iterative identification and control design.

# Appendix A

# Bounds on Inherent Error

Lemma 2.1 used in the proof of Proposition 2.1 can be shown using devices similar to those in [43].

**Proof of Lemma 2.1** If we can show that $\mathbf{S}(0)$ is symmetric around the origin, then for all $k_{est} \in \mathbf{L}$,

$$
\begin{aligned}
\sup_{k \in \mathbf{S}(0)} \|k - k_{est}\|_{\mathbf{L}} &= \frac{1}{2}\left(\sup_{k \in \mathbf{S}(0)} \|k - k_{est}\|_{\mathbf{L}} + \sup_{k \in \mathbf{S}(0)} \| - k - k_{est}\|_{\mathbf{L}}\right) \\
&\qquad \text{(as } \mathbf{S}(0) \text{ is symmetric),} \\
&\geq \frac{1}{2} \sup_{k \in \mathbf{S}(0)} \left(\|k - k_{est}\|_{\mathbf{L}} + \|k + k_{est}\|_{\mathbf{L}}\right), \\
&\geq \frac{1}{2} \sup_{k \in \mathbf{S}(0)} \|k - k_{est} + k + k_{est}\|_{\mathbf{L}}, \\
&= \sup_{k \in \mathbf{S}(0)} \|k\|_{\mathbf{L}}.
\end{aligned}
$$

Since the above inequality holds for all $k_{est} \in \mathbf{L}$, we get (2.15).

To show that $\mathbf{S}(0)$ is symmetric around the origin, we notice that

$$
\begin{aligned}
\mathbf{S}(0) &= \mathbf{S}_a \bigcap \left\{k \in \mathbf{L} : \ \mathbf{P}_{[t_0, t_0+T)} \mathbf{\Phi}_u k = \mathbf{P}_{[t_0, t_0+T)} v \ \textit{for some} \ v \in \mathbf{V}\right\}, \\
&=: \mathbf{S}_a \bigcap \mathbf{S}_c.
\end{aligned}
$$

Since $\mathbf{S}_a$ is symmetric by Assumption 1, it is enough to show that the set $\mathbf{S}_r$ symmetric. Let $k_1$ be an element in $\mathbf{S}_r$. Then there exist $v_1$ in $\mathbf{V}$ such that $(\mathbf{K}_1 u)(t) = v_1(t)\ \forall t \in [t_0, t_0 + T)$. This implies that $(-\mathbf{K}_1 u)(t) = -v_1(t)\ \forall t \in [t_0, t_0 + T)$. Since $\mathbf{V}$ is symmetric by Assumtion 1, $-v_1$ is in $\mathbf{V}$. This implies that $-k_1 \in \mathbf{S}_r$ and hence $\mathbf{S}_r$ is symmetric.

For (2.16), we show that for all $k_{true} \in \mathbf{S}_a$ and $v_{true} \in \mathbf{V}$,

$$\sup_{k_1, k_2 \in \mathbf{S}(y)} \|k_1 - k_2\|_{\mathbf{L}} \leq \sup_{h_1, h_2 \in \mathbf{S}(0)} \|h_1 - h_2\|_{\mathbf{L}}. \tag{A.1}$$

Let $k_1,\ k_2 \in \mathbf{S}(y)$. Then there exist $v_1,\ v_2 \in \mathbf{V}$ such that $\mathbf{P}_{[t_0,t_0+T)}(\mathbf{\Phi}_u(k_1) - y) = \mathbf{P}_{[t_0,t_0+T)}(v_1)$ and $\mathbf{P}_{[t_0,t_0+T)}(\mathbf{\Phi}_u(k_2) - y) = \mathbf{P}_{[t_0,t_0+T)}(v_2)$. It follows from the linearity of $\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u$ that $\mathbf{P}_{[t_0,t_0+T)}\mathbf{\Phi}_u\left(\frac{k_1-k_2}{2}\right) = \mathbf{P}_{[t_0,t_0+T)}\left(\frac{v_1-v_2}{2}\right)$. Since $\mathbf{V}$ is convex and symmetric, $\frac{v_1-v_2}{2} \in \mathbf{V}$, which implies that $\frac{k_1-k_2}{2} \in \mathbf{S}_r$. Since $k_1, k_2 \in \mathbf{S}_a$, and $\mathbf{S}_a$ is also convex and symmetic by Assumption 1, $\frac{k_1-k_2}{2}$ is also in $\mathbf{S}_a$. Hence, $\frac{k_1-k_2}{2} \in \mathbf{S}_a \bigcap \mathbf{S}_r = \mathbf{S}(0)$. By the symmetry of $\mathbf{S}(0)$, $\frac{k_2-k_1}{2} \in \mathbf{S}(0)$. (A.1) is proved by letting $h_1 = \frac{k_1-k_2}{2}$ and $h_2 = \frac{k_2-k_1}{2}$ on the right hand side.

Since $\mathbf{S}(0)$ is convex and symmetric, its diameter equals twice of its radius, i.e.,

$$\sup_{h_1, h_2 \in \mathbf{S}(0)} \|h_1 - h_2\|_{\mathbf{L}} = 2\delta(u).$$

This completes the proof.

$\square$

# Appendix B

# Comments on Spectrum Flatness

It can be shown that the spectrum flatness defined in the Chapter 4 is closely related to the ratio of the supremum and infimum of the spectrum of a signal in the standard sense.

**Proposition B.1** *If $u$ is a T-periodic signal, then*

$$\nu_u(T) = \frac{\min_m |\mathcal{U}(\omega_m)|}{\max_m |\mathcal{U}(\omega_m)|}, \tag{B.1}$$

*where* $\mathcal{U}(\omega_m) = \sum_{t=0}^{T-1} u(t)e^{-j\omega_m t}$, $m = 0, \cdots, T-1$, $\omega_m = 2\pi m/T$.

**Proof** Since $u$ is T-periodic, there exists a unitary matrix $V$ such that $U_{t_0}(n) = VU_{T-1}(T)$. It is trivial to check that for $t \in [0, T-1]$,

$$\sum_{\tau=0}^{T-1} u(T-1-t-\tau)e^{-j\omega_m(T-1-\tau)} = \mathcal{U}(\omega_m)e^{-j\omega_m t}. \tag{B.2}$$

Put

$$X = \begin{pmatrix} e^{-j\omega_0(T-1)} & e^{-j\omega_1(T-1)} & \ldots & e^{-j\omega_{T-1}(T-1)} \\ e^{-j\omega_0(T-2)} & e^{-j\omega_1(T-2)} & \ldots & e^{-j\omega_{T-1}(T-2)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \tag{B.3}$$

$$Y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{-j\omega_0} & e^{-j\omega_1} & \cdots & e^{-j\omega_{T-1}} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-j\omega_0(T-1)} & e^{-j\omega_1(T-1)} & \cdots & e^{-j\omega_{T-1}(T-1)} \end{pmatrix}. \tag{B.4}$$

Notice that both X and Y are unitary matrices. From (B.2) we have

$$U_{t_0}(n) = VY diag(\mathcal{U}(\omega_0), \cdots, \mathcal{U}(\omega_{T-1}))X^*. \tag{B.5}$$

Hence

$$\begin{aligned} \Phi_{u,t_0}(T) &= U_{t_0}(T)^T U_{t_0}(T), \\ &= X diag(|\mathcal{U}(\omega_0)|^2, \cdots, |\mathcal{U}(\omega_{T-1})|^2)X^*. \end{aligned}$$

Therefore

$$\inf\nolimits_{t_0} \underline{\sigma}(\Phi_{u,t_0}(T)) = \min\nolimits_m |\mathcal{U}(\omega_m)|^2,$$

$$\sup\nolimits_{t_0} \bar{\sigma}(\Phi_{u,t_0}(T)) = \max\nolimits_m |\mathcal{U}(\omega_m)|^2.$$

□

**Proposition B.2** *Assume $u \in \ell^\infty$ and its autocorrelation function*

$$\phi(t) = \lim_{T\to\infty} \frac{1}{2T} \sum_{\tau=-T}^{T} u(\tau)u(t+\tau) \tag{B.6}$$

*is in $\ell^1$. Let*

$$\Phi(e^{i\theta}) = \sum_{t=-\infty}^{\infty} e^{i\theta t}\phi(t), \tag{B.7}$$

$$m = \inf_{0\le\theta\le 2\pi} \Phi(e^{i\theta}), \tag{B.8}$$

$$M = \sup_{0\le\theta\le 2\pi} \Phi(e^{i\theta}). \tag{B.9}$$

*If there exists a function $f$ such that*

$$\left| \frac{1}{2T} \sum_{\tau=-T}^{T} u(\tau)u(t+\tau) - \frac{1}{2T} \sum_{\tau=t_0-T}^{t_0+T} u(\tau)u(t+\tau) \right| \le f(T) \quad \forall t_0, t \in \mathbb{Z}, \tag{B.10}$$

*and* $\lim_{T\to\infty} f(T) = 0$, *then*

$$\lim_{n\to\infty}\lim_{T\to\infty} \nu_u(n,T) = M/m. \tag{B.11}$$

To prove Proposition B.2, we need a standard result from Szego [13].

**Lemma B.1** *Let*

$$V_n = \begin{pmatrix} \phi(0) & \phi(-1) & \cdots & \phi(-n) \\ \phi(1) & \phi(0) & \cdots & \phi(1-n) \\ \vdots & \vdots & \cdots & \vdots \\ \phi(n) & \phi(n-1) & \cdots & \phi(0) \end{pmatrix} \tag{B.12}$$

*and* $\lambda_{n+1}^{(n)} \le \ldots \le \lambda_1^{(n)}$ *be eigenvalues of the Toeplitz form* $V_n$. *Under the hypotheses of the above proposition,*

$$\lim_{n\to\infty} \lambda_1^{(n)} = M, \tag{B.13}$$

*and*

$$\lim_{n\to\infty} \lambda_{n+1}^{(n)} = m. \tag{B.14}$$

**Proof of Proposition B.2** It will be shown that

$$\lim_{T\to\infty} \frac{1}{2T} \sup_{t_0} \bar{\sigma}(\Phi_{t_0}(n, 2T+1)) = \lambda_1^{(n)} \tag{B.15}$$

$$\lim_{T\to\infty} \frac{1}{2T} \inf_{t_0} \underline{\sigma}(\Phi_{t_0}(n, 2T+1)) = \lambda_{n+1}^{(n)}. \tag{B.16}$$

Then, by the above lemma, the proposition follows from the definition of $\nu_u(n,T)$.

Put $E_{t_0} = \frac{1}{2T}\Phi_{t_0}(n,2T+1) - V_n$. Since $\Phi_{t_0}(n,2T+1)$ and $V_n$ are both Hermitian matrices, so is $E_{t_0}$. Since $\Phi_{t_0}(n,2T+1)$ is a Hermitian matrix, its singular values are the same as its eigenvalues. If we see $\frac{1}{2T}\Phi_{t_0}(n,2T+1)$ as $V_n$ being perturbed by $E_{t_0}$, then by a standard result in matrix analysis, we have

$$\lambda_i(V_n) + \lambda_{min}(E_{t_0}) \le \frac{1}{2T}\lambda_i\left(\Phi_{t_0}(n,2T+1)\right) \le \lambda_i(V_n) + \lambda_{max}(E_{t_0}) \tag{B.17}$$

It follows that

$$\lambda_i(V_n) + \inf_{t_0} \lambda_{min}(E_{t_0}) \leq \tfrac{1}{2T}\sup_{t_0} \lambda_i(\Phi_{t_0}(n, 2T+1)) \leq \lambda_i(V_n) + \sup_{t_0} \lambda_{max}(E_{t_0}) \quad \text{(B.18)}$$

$$\lambda_i(V_n) + \inf_{t_0} \lambda_{min}(E_{t_0}) \leq \tfrac{1}{2T}\inf_{t_0} \lambda_i(\Phi_{t_0}(n, 2T+1)) \leq \lambda_i(V_n) + \sup_{t_0} \lambda_{max}(E_{t_0}) \quad \text{(B.19)}$$

If we can show that $\lim_{T\to\infty} \inf_{t_0} \lambda_{min}(E_{t_0}) = 0$ and $\lim_{T\to\infty} \sup_{t_0} \lambda_{max}(E_{t_0}) = 0$ then we can get (B.15) and (B.16) by taking the limit $T \to \infty$ in the above inequalities.

For each element of $E_{t_0}$, $e_{i,j}$, we have

$$\begin{aligned} e_{i,j} &= \frac{1}{2T}\left(\sum_{\tau=t_0+i-1-T}^{t_0+i-1+T} u(\tau)u(\tau+i-j)\right) - \frac{1}{2T}\left(\sum_{\tau=-T}^{T} u(\tau)u(\tau+i-j)\right) \\ &\quad + \frac{1}{2T}\left(\sum_{\tau=-T}^{T} u(\tau)u(\tau+i-j)\right) - \phi(i-j) \end{aligned}$$

Therefore, by (B.10), we have

$$|e_{i,j}| \leq f(T) + \left|\frac{1}{2T}\left(\sum_{\tau=-T}^{T} u(\tau)u(\tau+i-j)\right) - \phi(i-j)\right| \quad \text{(B.20)}$$

By the hypotheses, $\lim_{T\to\infty} f(T) = 0$ and

$$\lim_{T\to\infty}\left|\frac{1}{2T}\left(\sum_{\tau=-T}^{T} u(\tau)u(\tau+i-j)\right) - \phi(i-j)\right| = 0,$$

hence, $\lim_{T\to\infty} |e_{i,j}| = 0$. Since the eigenvalues of a matrix are continuous functions of its elements, we conclude that

$$\lim_{T\to\infty} \inf_{t_0} \lambda_{min}(E_{t_0}) = 0$$

and

$$\lim_{T\to\infty} \sup_{t_0} \lambda_{max}(E_{t_0}) = 0.$$

This completes the proof.

□

# Bibliography


[1] K. J. Astrom and B. Wittenmark. *Computer Controlled Systems: Theory and Design.* Englewood Cliffs, NJ, Prentice-Hall, 1984.

[2] K. J. Astrom and B. Wittenmark. *Adaptive Control.* Addison-Wesley, Reading, Massachusetts, 1989.

[3] S. Boyd and C. Barratt. *Linear Control System Design: Limits of Performance.* Prentice-Hall, New Jersey, 1991.

[4] S. Boyd and L. El Ghaoui. Method of Centers for Minimizing Generalized Eigenvalues. Preprint.

[5] P. E. Caines. *Linear Stochastic Systems.* Wiley, 1988.

[6] J. Chen, C. N. Nett, and Michael K. H. Fan. Optimal Non-Parametric System Identification From Arbitrary Corrupt Finit Time Series: A Control-Oriented Approach. In *Proceedings of ACC*, 1992.

[7] T. Chen and B. A. Francis. Input-Output Stability of Sampled-Data Systems. *IEEE*, AC-36:50–58, 1991.

[8] J. R. Deller. Set Membership Identification in Digital Signal Processing. *IEEE ASSP Magazine*, Oct. 1989.

[9] L. Gerencser. On a Class of Mixing Processes. *Stochastics*, 26:165 191, 1989.

[10] M. Gevers and L. Ljung. Optimal Experiment Desings with Respect to the Intended Model Application. *Automatica*, 22(5):543–554, 1986.

[11] G. Goodwin and R. L. Payne. *Dynamic System Identification*. Academic Press, New York, 1977.

[12] G. C. Goodwin, Michel Gevers, and B. Ninness. Quantifying the Error in Estimated Transfer Functions with Application to Model Order Selection. *IEEE*, AC-37(7):913–928, July 1992.

[13] U. Grenander and G. Szego. *Toeplitz Forms and their Applications*. Berkeley, University of California Press, 1958.

[14] G. Gu and P. P. Khargonekar. Linear and Non-Linear Algorithms for Identification in $\mathbf{H}^\infty$ With Error Bounds. *IEEE*, AC-37(7):953–963, July 1992.

[15] G. Gu, P. P. Khargonekar, and E. B. Lee. Approximation of Infinite-Dimensional Systems . *IEEE*, AC-34(6):610–618, June 1989.

[16] S. Gunnarsson and B. Wahlberg. Some Asymptotic Results in Recursive Identification Using Laguerre Models. In *Proc. of the 29th CDC*, pages 1068–1073, 1990.

[17] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. $\mathbf{H}^\infty$ Identification of Stable LSI Systems: A Scheme with Direct Application to Controller Design. In *Proceedings of ACC*, pages 1428–1434, 1989.

[18] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. Identification in $\mathbf{H}^\infty$: A Robustly Convergent Nonlinear Algorithm. In *Proceedings of ACC*, pages 386–391, 1990.

[19] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. Control Oriented System Identification: A Worst-Case/Deterministic Approach in $\mathbf{H}^\infty$. *IEEE*, AC-36:1163–1176, 1991.

[20] B. Kacewicz and M. Milanese. Optimal Finite Sample Experiment Design in Worst-Case $\ell^1$ System Identification. In *Proceedings of the 31st Conference on Decision and Control*, pages 56–61, 1992.

[21] P. Koosis. *The Logarithmic Integal I.* Cambridge University Press, 1988.

[22] R. L. Kosut, M. Lau, and S. Boyd. Set Membership Identification of Systems with Parametric and Nonparametric Uncertainty . *IEEE*, AC-37(7):929–941, July 1992.

[23] L. Lin and L. Y. Wang. Time Complexity and Model Complexity of Fast Identification. In *Proceedings of the 32nd CDC*, 1993.

[24] L. Lin, L. Y. Wang, and G. Zames. Uncertainty Principles and Identification n-Widths for LTI and Slowly Varying Systems. In *Proceedings of ACC*, pages 296–300, 1992.

[25] L. Lin, L. Y. Wang, and G. Zames. Fast Identification n-Widths and Quasianalytic Inputs for Continuous LTI Systems. In *Proceedings of ACC*, 1993.

[26] L. Ljung. *System Identification: Theory for the Users.* Reading, MA: Addison-Wesley, 1987.

[27] D. G. Luenberger. *Optimization by Vector Space Methods.* John Wiley, 1968.

[28] P. M. Mäkilä. On Identification of Stable Systems and Optimal Approximation. *Automatica*, 27:663–676, 1991.

[29] P. M. Makila and J. R. Partington. Robust Approximation and Identification in $\mathbf{H}^\infty$. In *Proc. of ACC*, 1991.

[30] T. H. Matheiss and D. S. Rubin. A Survey and Comparison of Methods for Finding All Vertices of Convex Polyhedral Sets. *Mathematics of Operations Research*, 5(2), May 1985.

[31] R. K. Mehra. Optimal Input Signals for Parameter Estimation in Dynamic Systems— Survey and New Results. *IEEE*, AC-19(6):753–768, 1974.

[32] M. Milanese, R. Tempo, and A Vicino. *Robustness in Identification and Control*. Plenum Press, 1989.

[33] A. S. Morse. Towards a Unified Theory of Parameter Adaptive Control: Tunability. *IEEE*, AC-35(9):1002–1012, Sept. 1990.

[34] A. S. Morse. Towards a Unified Theory of Parameter Adaptive Control—Part II: Certainty Equivalence and Implicit Tuning. *IEEE*, AC-37(1):15–29, Jan. 1992.

[35] N. K. Nikol'skii. *A Treatise on the Shift Operator*. Springer Verlag, 1986.

[36] J. P. Norton and S. M. Veres. Developments in Parameter Bounding. In *Empty*, 1991.

[37] P. Parker and R. Bitmead. Adaptive Frequency Response Estimation. In *Proceedings of the 26th Conference on Decision and Control*, pages 348–353, 1987.

[38] J. R. Partington and P. M. Makila. Robust Identification of Stabilizable Systems. Report of the University of Leeds, 1991.

[39] A. Pinkus. *n-Widths in Approximation Theory*. Springer-Verlag, 1985.

[40] K. Poolla, P. P. Khargonekar, A. Tikku, J. Krause, and K. Nagpa. A Time Domain Approach to Model Validation. In *Proc. of ACC*, pages 313–317, 1992.

[41] K. Poolla and A. Tikku. On the Time Complexity of Worst-Case System Identification. In *Proc. of 1993 ACC*, 1993.

[42] J. Rissanen. *Stochastic Complexity in Statistical Inquiry.* World Scientific, 1989.

[43] R. J. P. Shrama and P. M. J. Van den Hof. An Iterative Model Scheme for Identification and Cotrol Design Based on Coprime Factorization. In *Proceedings of ACC*, 1992.

[44] R. S. Smith and J. C. Doyle. Model Validation: A Connection Between Robust Control and Identification. *IEEE*, AC-37(7):942–952, July 1992.

[45] D. N. C. Tse, M. A. Dahleh, and J. N. Tsitsiklis. Optimal Asymptotic Identification Under Bounded Disturbances. In *Proceedings of the 30th Conference on Decision and Control*, pages 623–628, 1991.

[46] A. G. Vitushkin. *Theory of the Transmission and Processing of Information.* Pergamon, 1961.

[47] B. Wahlberg and L. Ljung. Hard Frequency-Domain Model Error Bounds from Least-Squares Like Identification Techniques. *IEEE*, AC-37(7):900–912, July 1992.

[48] L. Y. Wang and L. Lin. Worst Case Identification in Discrete and Continuous Time Systems. In *Proceedings of the 30th Conference on Decision and Control*, pages 38–43, 1991.

[49] L. Y. Wang and L. Lin. Complexity and Entropy of Identification and Adaptation. In *Proceedings of the 31st Conference on Decision and Control*, 1992.

[50] L. Y. Wang and L. Lin. On the Metric Dimensions of Discrete-Time Systems. *Systems and Control Letters*, 1992. to appear.

[51] L. Y. Wang and G. Zames. Local-Global Algebras for Slow $\mathbf{H}^\infty$ Adaptation: Part II–Optimization of Stable Plants. *IEEE*, AC-36(2):143–151, February 1991.

[52] N. Wiener. *The Fourier Integral and Certain of its Applications*. Dover Publications, Inc., 1933.

[53] G. Zames. Feedback and Complexity, Plenary Lecture. In *Proc. 1976 CDC*, pages 1–2, 1976.

[54] G. Zames. On the Metric Complexity of Causal Linear Systems: $\epsilon$-Entropy and $\epsilon$-Dimension for Continuous Time. *IEEE*, AC-24(2):222–230, April 1979.

[55] G. Zames. Feedback and Optimal Sensitivity: Model Reference Transformations, Seminorms, and Approximate Inverses. *IEEE*, AC-26(2):301–320, April 1981.

[56] G. Zames. Feedback Organizations, Learning and Complexity in $\mathbf{H}^\infty$ . In *Plenary Lecture, Proc. 1988 ACC*, Atlanta, 1988.

[57] G. Zames. Adaptive Feedback, Identification and Complexity: An Overview. In *Proc. 1993 CDC*, 1993.

[58] G. Zames, L. Lin, and L. Y. Wang. Fast Identification n-Widths and Uncertainty Principles for LTI and Slowly Varying Systems. To appear in the IEEE Trans. on Auto. Control.

[59] G. Zames and J. G. Owen. A Note on Metric Dimension and Feedback in Discrete Time. *IEEE*, AC-38(4):664–667, April 1993.

[60] G. Zames and L. Y. Wang. Local-Global Algebras for Slow $\mathbf{H}^\infty$ Adaptation: Part I–Inversion and Stability. *IEEE*, AC-36(2):130–142, February 1991.

[61] Z. Zang, R. R. Bitmead, and M. Gevers. $\mathbf{H}^2$ Iterative Model Refinement and Cotrol Robustness Enhancement. In *Proceedings of the 30th Conference on Decision and Control*, 1991.

[62] C. Zerovs, P. R. Belanger, and G. A. Dumont. On PID Controller Tuning Using Orthonormal Series Identification. *Automatica*, 24:165–176, 1988.