STATISTICAL ANALYSIS OF MEMBERSHIP-SET BASED ESTIMATORS

Nuray AT

Doctor of Philosophy Dissertation

Electrical and Electronics Engineering Program

August 2005

ABSTRACT

Doctor of Philosophy Dissertation

STATISTICAL ANALYSIS OF MEMBERSHIP-SET BASED ESTIMATORS

NURAY AT

Anadolu University
Graduate School of Sciences
Electrical and Electronics Engineering Program

Adviser: Prof. Dr. Hüseyin Akçay
2005, 64 pages

In this thesis, statistical analysis of membership-set based estimators is studied in two different settings. In the first setting, periodic input signals and orthonormal regressors are considered. First, upper and lower probability bounds on the diameter of the membership-set are obtained. Then, explicit formulae for the central algorithm and the diameter of the membership-set are derived, and the set of all projection algorithms is characterized for the special case of overparameterized model structures where the number of the unknown parameters equals the input period. In the second setting, one-dimensional parameter space and arbitrary scalar regressors with magnitude constraints are considered. Non-asymptotic, order-tight, upper and lower bounds on the convergence rate of the parameter estimate variance for the central and the minimax algorithms are derived.

Keywords: Membership-set; Central algorithms; Projection algorithms; Minimax algorithms; Convergence analysis.

ÖZET

Doktora Tezi

ÜYELİK-SETİ BAZLI KESTİRİM ALGORİTMALARININ İSTATİKSEL ANALİZİ

NURAY AT

Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Hüseyin Akçay
2005, 64 sayfa

Bu tezde üyelik-seti bazlı kestirim algoritmalarının istatiksel analizi iki farklı durum için çalışıldı. İlk durumda periyodik girdi ve birimdik regresorlar kullanıldı. Bu durum için ilk önce üyelik-seti çapına ait alt ve üst olasılık sınırları elde edildi. Daha sonra bilinmeyen parametre sayısının girdi periyoduna eşit olduğu durumda merkezi algoritmaya ve üyelik-seti çapına ait kestirim değerleri verildi, ve projeksiyon algoritmalarının oluşturduğu set karakterize edildi. İkinci durumda tek boyutlu parametre uzayı göz önüne alınarak gelişigüzel skalar ve büyüklük kısıtlı sistem girdileri kullanıldı. Bu durum için merkezi algoritma ve en küçük-en büyük algoritma varyanslarının asimptotik olmayan ve gevşek olmayan yakınsama hızlarına ait alt ve üst sınırlar hesaplandı.

Anahtar Kelimeler: Üyelik-seti; Merkezi algoritmalar; Projeksiyon algoritmaları; En küçük-en büyük algoritmaları; Yakınsama analizi.

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

System identification deals with the problem of building mathematical models of systems based on observed data from the systems [1]. However, the observed data are always associated with some uncertainty and the solution of the problem depends on the type of assumptions made on uncertainty.

The classical approaches assume a probabilistic description of uncertainty [1 – 3]. Along this direction, a variety of theoretically sound and practically efficient schemes have been developed such as Maximum Likelihood, Least-Squares, and Bayesian techniques.

An alternative approach, referred to as set membership or bounded error approach, assumes a deterministic description of uncertainty which is known only to have given bounds [4 – 16]. In the following, the literature on deterministic worst case estimation, or robust estimation, is reviewed.

The problem of estimation with set membership uncertainty has been studied since the late 1960s [4 – 7]. In these studies, particularly, the state estimation of a linear dynamic system is considered. In [4], a recursive algorithm for calculating a bounding ellipsoid which always contains the true state of the system was developed. In [5], the characterization of the set of all possible states was provided using the concept of support functions. Bertsekas and Rhodes [6] considered two distinct types of constraints on the uncertain quantities: energy-type constraints and individual instantaneous constraints; in the former case, the set of all possible states which is an ellipsoid is characterized, while in the latter case, a bounding ellipsoid to the set of all possible states is derived. Schlaepfer and Schweppe [7] extended the work in [4] to the case of continuous-time linear dynamic systems, and presented an approach for obtaining a bounding ellipsoid to the set of all possible states.

In [8], a different perspective has been taken within the context of state estimation of a linear dynamic system, the so-called independent time-varying parametric uncertainties. In this setting, the set of all possible states might not be convex due to multiplication of sets of uncertainties; hence, a recursive characterization of the set of all possible states is difficult to obtain. On the other hand, it was shown in the same work that the convex hull of the set of all possible states can be recursively propagated forward in time.

Some first steps in the development of a theory for parameter estimation and in the evaluation of uncertainty intervals have been taken by Milanese and Belforte [10]. In this study, for linear families of models and estimators, very simple and computationally feasible algorithms were derived for evaluating uncertainty intervals, and the solution to the problem of minimum uncertainty interval estimator was given by means of a linear programming problem.

In [11], the general theory of optimal algorithms [17, 18] was applied to solve problems in the fields of parameter estimation and time series prediction. In particular, for linear parameterizations, the results of [10] were extended and the derivation of computationally simple optimal algorithms for these two problems was examined.

The investigation of the optimal algorithms and of the optimal information for linear problems was carried on by Milanese et al. [12] where a particular emphasis was given to constructive aspects. Two main results were [12]: first, the Y-strong optimality (see Chapter 2 for the definition) of the central algorithm [11] is proved; second, a simple solution to a particular case of optimal information problem, called optimal sampling problem, is given.

Optimality properties of the central and the projection algorithms (see Chapter 2 for the definition) for linear problems with set membership uncertainty were investigated in [13]. Particular attention was devoted to least-squares algorithms, and the noise was assumed to be unknown but bounded in the ℓ_2-norm (Hilbert norm). Least-squares algorithms are based on the minimization of the ℓ_2-norm of the residual. In this setting, it was shown that least-squares algorithms are Y-strongly, and therefore, globally optimal. On the other hand, it was noted in the same work that algorithms based on least-absolute values or minimax criteria, in general, are neither Y-strongly optimal nor globally optimal.

Maximum Likelihood Estimators (MLEs) are commonly used in the stochastic setting due to their asymptotic optimality (see, for example [19]) properties. On the other hand, it is important to know their optimality properties in the worst case setting. To this effect, Tempo and Wasilkowski [14] studied the following problem: Under what circumstances are MLEs optimal in the worst case? This problem was attacked by analyzing worst case optimality of MLEs for a number of probability density functions (pdfs) of noise such as uniform, normal, and Laplace. The findings were [14]: For uniform distribution, in general, MLEs are neither optimal nor robust-interpolatory (see Chapter 2 for the definition), though there exist two MLEs which are optimal or robust-interpolatory, if the noise is bounded in the ℓ_∞-norm. For normal distribution, the unique MLE is robust-interpolatory and optimal, if the noise is bounded in the ℓ_2-norm. For Laplace distribution, the unique MLE is robust-interpolatory, if the noise is bounded in the ℓ_1-norm, but is not optimal.

In [15], optimal algorithms for robust estimation and filtering in the presence of bounded noise were constructed, and the previous results [11] obtained for complete (one-to-one) and approximate information were extended to the case of partial (not necessarily one-to-one) and approximate information. The use of partial and approximate information seems useful 1) for problems in which two different sources of noise should be considered, for example, process and measurement noise, and/or 2) for problems in which the parameters are not completely identifiable from the measurements.

An outline of the main results in the area of estimation theory for set membership uncertainty up to the early 1990s can be found in [16].

In many areas of signals, systems, and control theory, orthogonal functions play an important role in subjects of analysis and design. The use of orthonormal basis functions for system identification and approximation has been studied to a great extent in the literature, see [20 – 30] and the references therein.

In the context of system identification, the transfer function of the system is represented as a series expansion in terms of orthonormal basis functions, and the identification is performed by estimating the expansion coefficients. The main motivation for using orthonormal basis functions to represent the system is that the resulting model structure leads to a linear regressor form (see Chapter 2). Another factor in using orthonormal basis functions is due to their numerical robustness property compared to non-orthonormal ones.

The most commonly used orthonormal basis, the trigonometric basis, yields a Finite Impulse Response (FIR) model structure. However, the use of FIR model structures to represent a system with a slow mode has the disadvantage that the number of terms in the series expansion to provide an acceptable approximation of the system can be prohibitively high. To this effect, model structures allowing for the encoding of prior knowledge of pole positions have been introduced, the so-called Laguerre and Kautz models [20 – 24]. Laguerre functions involve a scalar design parameter which is chosen in a range that matches the dominant dynamics of the system to increase the rate of convergence of the Laguerre series expansion. On the other hand, if prior knowledge of a resonant mode exists, it is more appropriate to use the two-parameter Kautz basis.

A generalization of the Laguerre and Kautz models has been considered in [24 – 30]. A state-space approach was taken in [24, 25] where use is made of balanced realizations of inner functions. In this setting, it was shown that every stable system gives rise to a complete set of orthonormal functions which form a basis for the signal space of interest.

A unifying construction of orthonormal bases for system identification and approximation was given in [28]; the well-known FIR, Laguerre, and Kautz models are all special cases of this construction. Moreover, the unifying construction allows prior knowledge of an arbitrary number of modes to be incorporated.

Recently, the rational wavelet basis has been suggested [26, 27]. The wavelet basis enjoys the advantage of generalizing the FIR, Laguerre, and Kautz models, and further the generalized orthonormal bases in [24, 25, 28 – 30]; hence, it allows the utilization of much more prior knowledge about the system.

In the context of robust estimation, perhaps a more important question is that whether the linear span of a basis can arbitrarily well approximate any given element in the space in question [29]. This property is referred to as a set A being *fundamental* in a space X, i.e., the closure of the linear span of A under the norm on X is equal to X. A sufficient condition for the wavelet basis to be fundamental in the disc algebra $A(\mathbf{D})$ was stated in [27]. This result was improved in [29] and a much milder necessary and sufficient condition for the wavelet basis to be fundamental in $A(\mathbf{D})$, and in the Hardy space $H_p(\mathbf{D})$ for all $1 \leq p < \infty$ was given; moreover, several sufficient conditions for the wavelet basis to be fundamental in ℓ_1 were provided. To this end, re-parameterization of the linear space spanned by the wavelet basis as the linear space spanned by the generalized orthonormal basis of [24, 25] is utilized.

In the context of bounded error estimation, there are basically two directions. The first one aims to characterize the feasible parameter set, also referred to as the membership-set (see Chapter 2); and the second one aims to compute a specific estimate within the membership-set that enjoys some optimality properties.

For linear models, the feasible parameter set is a polytope. Even then the computational complexity of an exact representation of this set can be quite high, and often approximate descriptions are used. Typical choices are boxes or ellipsoids. In [31], optimal inner bounding of the feasible parameter set by means of balls in the ℓ_∞-norm (boxes), the ℓ_2-norm (ellipsoids), and the ℓ_1-norm (diamonds) was investigated.

Although the characterization of the membership-set is a deterministic procedure, one may still wish to make probabilistic assumptions about its context, as in this thesis, so as to investigate its average behavior. This line of research, set membership identification in a probabilistic framework, has been conducted by [32 – 41] and is reviewed next.

Some fundamental properties of the membership-set were studied by Bai et al. [33]. Specifically, the three main problems were considered: the size of the membership-set; optimal inputs and complexity of the membership-set based estimators; and relations with least-squares estimate. The followings were found [33]: First, the size of the membership-set is derived if the noise is bounded by ε but otherwise unknown, for a given $\varepsilon > 0$; and the probability distribution of the size of the membership-set is obtained if the noise is a sequence of independent identically distributed random variables with a pdf supported in $[-\varepsilon, \varepsilon]$. Second, optimality conditions on the input in order to minimize the size of the membership-set are derived. Finally, the relations between least-squares algorithm and membership-set based estimators are studied, and necessary and sufficient conditions under which the least-squares estimate lies in the membership-set are obtained. In [34], an analytic center approach was proposed for bounded error parameter estimation. The analytic center minimizes the logarithmic average output error among all the estimates within the membership-set; is an MLE for a certain class of noise density functions; and allows an easy-to-compute sequential algorithm. The convergence of this sequential algorithm was achieved and, more significantly, it was shown that the number of Newton iterations required to compute a sequence of analytic centers is linear in the number of observed data points [34]. Lastly, it was remarked that the analytic center is almost-optimal as it lies in the membership-set. In [35], the asymptotic performance of the analytic center approach was established and it was shown that the analytic center converges to the true parameter under mild conditions.

The problem of robust estimation for families of noise distributions with bounded support was examined in [36] where a specific attention was given to minimax algorithms. The minimax algorithm has some nice robust convergence properties [42] and also some near-optimality properties with respect to bounded disturbances. In [36], it was shown that the minimax estimate outperforms the least-squares estimate for certain distributions; more precisely, if the noise distribution does not vanish at its endpoints, the variance of the minimax estimate converges to zero much faster than that of the least-squares estimate. On the other hand, it was also noted in the same work that there are noise distributions which make the convergence of the minimax algorithm extremely slow.

In [37], the size of the membership-set was studied in a probabilistic framework; in particular, upper and lower non-asymptotic probability bounds on the diameter of the membership-set were given assuming that the regressors are persistently exciting and the measurement noise is a sequence of independent identically distributed bounded random variables. These bounds were then used in the derivation of confidence intervals for interpolatory estimators (see Chapter 2 for the definition) providing a criterion whether a given estimator is likely to lie in the membership-set or not.

We remark that the diameter is more useful than the volume to quantify the size of the membership-set since volume of a set could be zero while its diameter is infinity at the same time.

Finite sample properties of system identification methods have been studied both in deterministic and in stochastic settings. The deterministic setting has been considered in [43 – 45]. In this framework, under the assumptions of unknown but bounded noise and the system being in the model set, it was shown [43, 44] that the required number of data points increases exponentially with the model order in order to keep the diameter of the membership-set below a certain value (measured in the ℓ_1-norm of the impulse response). In [45], the finite sample properties of worst case deterministic identification in ℓ_1 and in H_∞ were studied using n-widths and metric complexity.

The stochastic setting has been considered in [46 – 48]. In [46], the finite sample properties of prediction error methods using a quadratic criterion function were studied. Specifically, the following problem was considered: How many data points are required to guarantee with high probability that the expected value of the quadratic identification criterion is close to its empirical mean value? This problem was resolved using risk minimization theory; more precisely, uniform probabilistic bounds on the difference between the expected value of the squared prediction error and its empirical mean evaluated on a finite number of data points are derived, and therefore, the sample sizes are obtained. Further, it was shown that the number of data points required to maintain a given bound on the deviation grows no faster than quadratically with the number of parameters. Similar results, concerning probabilistic bounds on the difference between the expected value of the identification criterion evaluated at the estimated parameters and at the optimal parameters, were obtained for a general linear model class in [47]; nevertheless, derived bounds are known to be not tight. In [48], the finite sample results using frequency domain measurements corrupted by Gaussian noise were established.

1.1 Organization of the Thesis

The thesis is organized as follows:

In Chapter 2, the basic concepts in set membership identification theory are introduced. Also, orthonormal basis functions are reviewed.

In Chapter 3, periodic input signals and orthonormal regressors are considered. First, upper and lower probability bounds on the diameter of the membership-set are obtained. Then, explicit formulae for the central algorithm and the diameter of the membership-set are derived, and the set of all projection algorithms is characterized for the special case of overparameterized model structures. A preliminary version of this chapter appeared in [38] and the present version is submitted for publication [39].

In Chapter 4, one-dimensional parameter space and arbitrary scalar regressors with magnitude constraints are considered. Non-asymptotic, order-tight, upper and lower bounds on the convergence rate of the parameter estimate variance for the central and the minimax algorithms are derived. The results pertaining to statistical analysis of the central algorithm appeared in [40] and the present version of this chapter is accepted for publication [41].

2. PRELIMINARIES

In this chapter, we will introduce the basic concepts in set membership identification theory and review orthonormal basis functions. The material presented here is fairly standard and can be found, for example, in [12, 16 – 18, 22 – 24, 28, 49].

2.1 The Setting

Let X, Y, and Z be normed linear spaces:

- X: problem element space,
- Y: measurement space,
- Z: solution space.

Suppose K to be a bounded subset of X. Consider a given operator S, called a *solution operator*, which maps X into Z

$$S : X \to Z.$$

The goal is to approximate $S(x)$ for $x \in K \subseteq X$ using only partial information about x. Let H be the *information operator*

$$H : X \to Y.$$

In general, the exact information $H(x)$ about x is not available and only the perturbed information $y \in Y$ is known. Under the additive noise assumption, the perturbed information y is given by

$$y = H(x) + \eta,$$

where η is assumed to be unknown-but-bounded, that is,

$$\|\eta\|_\infty \leq \varepsilon,$$

for a given $\varepsilon > 0$. An *algorithm* A is a mapping (not necessarily a linear one) from Y into Z

$$A : Y \to Z$$

which provides an approximation $A(y)$ to $S(x)$. Such an algorithm is also called an *estimator*.

We now introduce the following sets [16]:

- $FPS(y,\varepsilon)$: feasible problem element set

$$FPS(y,\varepsilon) = \{x \in K : \|H(x) - y\|_\infty \leq \varepsilon\},$$

- $FSS(y,\varepsilon)$: feasible solution set

$$FSS(y,\varepsilon) = S(FPS(y,\varepsilon)),$$

- $MUS(y,\varepsilon)$: measurement uncertainty set

$$MUS(y,\varepsilon) = \{\bar{y} \in Y : \|\bar{y} - y\|_\infty \leq \varepsilon\},$$

- $EUS_A(y,\varepsilon)$: estimate uncertainty set for a given algorithm A

$$EUS_A(y,\varepsilon) = A(MUS(y,\varepsilon)).$$

The process is illustrated in Figure 2.1 where K is taken to be the space X. Note that the sets $FSS(y,\varepsilon) \subseteq Z$ and $EUS_A(y,\varepsilon) \subseteq Z$ are usually different since the latter depends on the particular algorithm used.

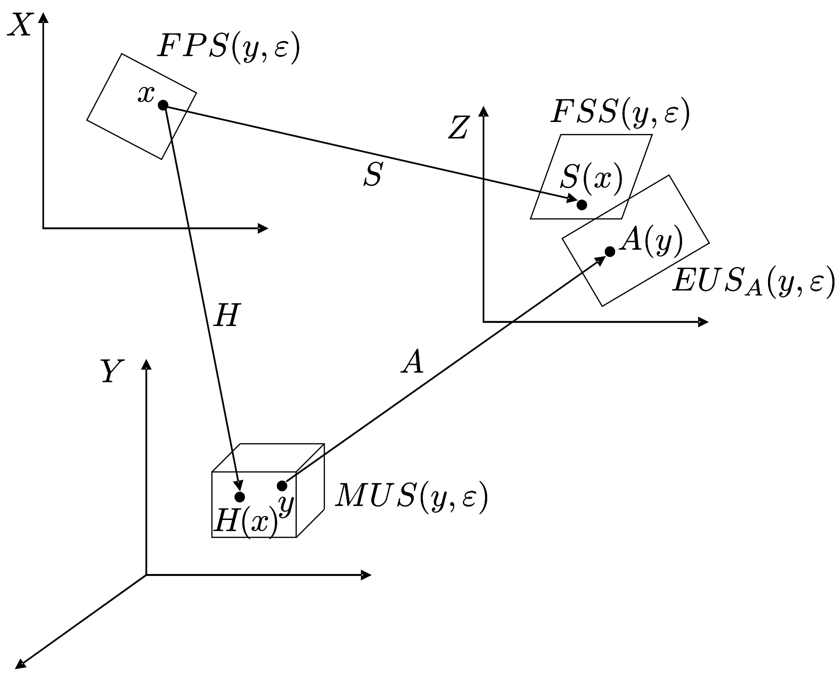

Figure 2.1: General estimation problem

2.2 Measure of Information

In this section, we review some set-theoretic concepts such as the radius and the diameter of a set.

Let Z be a linear normed space. Consider a set $M \subseteq Z$. The *radius* of M, denoted by $R(M)$, is defined as

$$R(M) = \inf_{z \in Z} \sup_{a \in M} \|z - a\|_Z . \tag{2.1}$$

Similarly, the *diameter* of M, denoted by $D(M)$, is defined by

$$D(M) = \sup_{z,a \in M} \|z - a\|_Z . \tag{2.2}$$

Note that one may view $R(M)$ as the radius of the smallest ball that contains the set M, and $D(M)$ as the largest distance between any two elements of M. Hence, the following relationship holds

$$R(M) \leq D(M) \leq 2R(M).$$

Furthermore, for any $a \in M$,

$$R(M) \leq \sup_{z \in M} \|z - a\|_Z \leq D(M) \leq 2R(M).$$

There are several ways of measuring the size of a set including the radius, the diameter, and the volume. In this thesis, we use the diameter of a set as a measure of its size. The local diameter of information at y is given by $D(FSS(y, \varepsilon))$, whereas the global diameter of information is defined as

$$D(\varepsilon) = \sup_{y \in \mathcal{Y}} D(FSS(y, \varepsilon)),$$

where

$$\mathcal{Y} = \{y \in Y : FPS(y, \varepsilon) \neq \emptyset\}. \tag{2.3}$$

Note that the local diameter of information is computed for a specific perturbed information (or, measurement) y. On the other hand, the global diameter of information is obtained using the set $\mathcal{Y}$ of measurements.

Similarly, the local radius of information and the radius of information are given, respectively, by

$$R(FSS(y, \varepsilon))$$

and

$$R(\varepsilon) = \sup_{y \in \mathcal{Y}} R(FSS(y, \varepsilon)). \tag{2.4}$$

2.3 Optimality Concepts

The performance of a given algorithm A can be specified according to the following errors:

- $E(A, y)$: Y-local error

$$E(A, y) = \sup_{x \in FPS(y,\varepsilon)} \|S(x) - A(y)\|_Z \,,$$

- $E(A)$: global error

$$E(A) = \sup_{y \in \mathcal{Y}} E(A, y),$$

  where the set $\mathcal{Y}$ is as in (2.3).

An algorithm $A^\sharp$ is *Y-strongly optimal* if

$$E(A^\sharp, y) \leq E(A, y), \qquad \forall A, \forall y \in \mathcal{Y}.$$

Note that Y-strong optimality is a particularly important property in system identification, where a set of measurements is available and one wants to minimize the error related to the worst case problem element $x \in FPS(y, \varepsilon)$ for each $y \in \mathcal{Y}$.

Similarly, an algorithm $A^\star$ is called *globally optimal* if

$$E(A^\star) \leq E(A), \qquad \forall A. \tag{2.5}$$

It is worthwhile to observe that Y-strong optimality is stronger property than global optimality in the sense that a Y-strongly optimal algorithm is also globally optimal while the converse is not necessarily true.

The minimum global error (also known as the *intrinsic error*) is given by the radius of information in (2.4), that is,

$$R(\varepsilon) = \min_A E(A).$$

Moreover, from (2.5) we have

$$E(A^\star) = R(\varepsilon).$$

It may be difficult in some cases to implement optimal algorithms. On the other hand, almost-optimal algorithms are generally easy to implement. An algorithm $A^\dagger$ is called *Y-strongly almost-optimal* (or Y-strongly optimal within a factor of k, $k > 1$) if

$$E(A^\dagger, y) \leq kE(A^\sharp, y).$$

Likewise, an algorithm $A^\ddagger$ is called *globally almost-optimal* if

$$E(A^\ddagger) \leq kE(A^\star) = kR(\varepsilon).$$

An important class of almost-optimal algorithms, *interpolatory estimators*, are studied in the following section.

2.4 Interpolatory Estimators

In system identification, it is usually desired for an algorithm to provide estimates within the feasible solution set. Interpolatory estimators form one such a case. An algorithm A_i is called interpolatory if

$$A_i \in FSS(y, \varepsilon).$$

Some of the well-known interpolatory estimators are described in the following.

The most popular one is the *central* algorithm which picks the so-called Chebyshev center of the feasible solution set. More precisely,

$$A_c = \arg \inf_{z' \in Z} \sup_{z'' \in FSS(y,\varepsilon)} \|z' - z''\|_Z . \tag{2.6}$$

This is the best worst case estimate of the solution. Note that the Chebyshev center of $FSS(y, \varepsilon)$ needs not be unique. It is a well-known fact that any central algorithm is Y-strongly optimal in the class of all algorithms. The problems with central algorithms: they are known to be very sensitive to outliers and are not robust with respect to the bound ε.

To overcome the latter, robust-interpolatory algorithms are proposed. An algorithm A_r is called *robust-interpolatory* if and only if it is interpolatory for all $\varepsilon > 0$. The robust-interpolatory algorithms coincide with the *projection* algorithms [14] defined by

$$A_p = S(x_p), \qquad x_p = \arg \min_{x \in K} \|H(x) - y\|_Y . \tag{2.7}$$

This is the best estimate of the solution in the sense that the Y-norm of the output error is minimized.

A recently introduced algorithm [34, 35] picks the *analytic center* of the feasible solution set. More precisely,

$$A_a = S(x_a), \qquad x_a = \arg \max_{x \in FPS(y,\varepsilon)} \sum_{t=1}^{N} \ln \left(\varepsilon^2 - \{[H(x) - y](t)\}^2 \right), \tag{2.8}$$

where N denotes the number of measurements.

Note that by definition, interpolatory estimators are Y-strongly optimal within a factor of two. Hence,

$$E(A_p, y) \leq 2E(A_c, y), \qquad \forall y \in \mathcal{Y}$$

and

$$E(A_a, y) \leq 2E(A_c, y), \qquad \forall y \in \mathcal{Y}.$$

In this thesis, we mainly consider the central and the projection algorithms.

2.5 Application to Parameter Estimation

In this section, we apply the general framework discussed earlier to the problem of parameter estimation. Here, the solution operator S is identity, and therefore, the feasible solution set coincides with the feasible parameter set which will be referred to as, hereinafter, the *membership-set*.

Consider a discrete-time scalar system represented by the output-error model:

$$y(t) = \phi^T(t)\theta + \eta(t), \qquad t = 1, 2, \cdots, N \tag{2.9}$$

where $y(t)$ is the system output, $\phi(t) \in \mathbb{R}^n$ is the measurable regression vector, $\theta \in \mathbb{R}^n$ is the unknown parameter vector, and $\eta(t)$ is the additive measurement noise. Let $u(t)$ be the applied input. We assume that $\eta(t)$ is a sequence of independent identically distributed random variables satisfying

$$\|\eta\|_\infty = \max_{1 \leq t \leq N} |\eta(t)| \leq \varepsilon, \qquad \varepsilon > 0. \tag{2.10}$$

The membership-set, the set of all possible parameters that are consistent with the model (2.9), the available data $\{\phi(t), y(t)\}_{t=1}^N$, and the noise assumption (3.3), is defined as

$$S_N(y, u, \varepsilon) = \bigcap_{t=1}^{N} \{\widehat{\theta} \in \mathbb{R}^n : \ |y(t) - \phi^T(t)\widehat{\theta}| \leq \varepsilon\}. \tag{2.11}$$

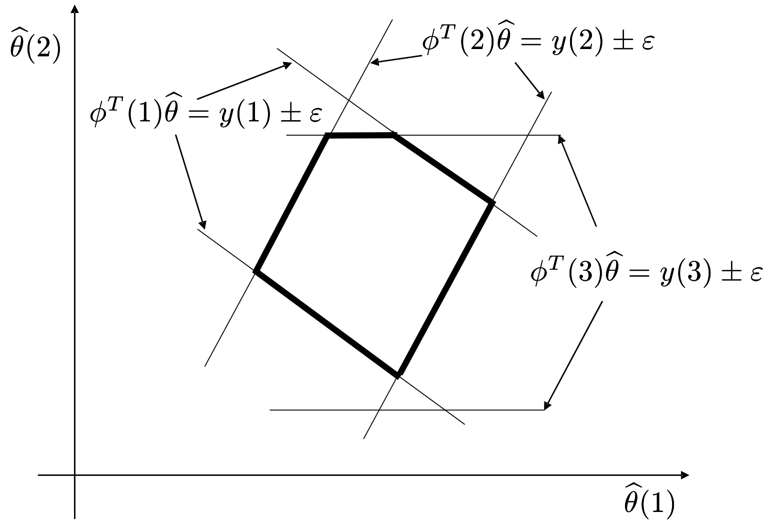

Figure 2.2: A graphical illustration of the membership-set

Figure 2.2 shows a graphical illustration of $S_N(y, u, \varepsilon)$.

From (2.1) and (2.2), the radius and the diameter of $S_N(y, u, \varepsilon)$ are given respectively by

$$R(S_N(y,u,\varepsilon)) = \min_{\theta' \in S_N(y,u,\varepsilon)} \max_{\theta'' \in S_N(y,u,\varepsilon)} ||\theta' - \theta''||_\infty \tag{2.12}$$

and

$$D(S_N(y,u,\varepsilon)) = \max_{\theta',\theta'' \in S_N(y,u,\varepsilon)} ||\theta' - \theta''||_\infty. \tag{2.13}$$

We now study the central and the projection algorithms in this framework. From (2.6), the central estimate is given by

$$\widehat{\theta}_c = \arg \min_{\theta' \in S_N(y,u,\varepsilon)} \max_{\theta'' \in S_N(y,u,\varepsilon)} ||\theta' - \theta''||_\infty. \tag{2.14}$$

Due to the ℓ_∞-norm imposed on the parameter space, from [11] we have the following simple formula for the central algorithm for $k = 1, \cdots, n$,

$$\widehat{\theta}_c(k) = \frac{1}{2}\left[\max_{\widehat{\theta} \in S_N(y,u,\varepsilon)} \widehat{\theta}(k) + \min_{\widehat{\theta} \in S_N(y,u,\varepsilon)} \widehat{\theta}(k)\right]. \tag{2.15}$$

Figure 2.3 shows a graphical illustration of $\widehat{\theta}_c$. Note that this estimator is also a maximum likelihood estimator when $\eta(t)$ is uniformly distributed random variable in $[-\varepsilon, \varepsilon]$ [14].

From (2.7), the projection algorithms are given by

$$\widehat{\theta}_p = \arg \min_{\widehat{\theta} \in S_N(y,u,\varepsilon)} \max_{1 \le t \le N} |y(t) - \phi^T(t)\widehat{\theta}|. \tag{2.16}$$

Robustness properties of the projection algorithms have been studied extensively in both set membership identification [13, 14] and worst case identification contexts [42, 50, 29].

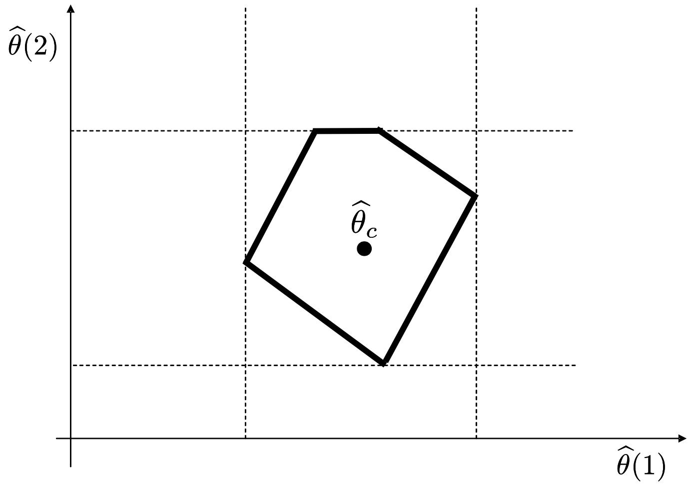

Figure 2.3: A graphical illustration of the central algorithm

2.6 Orthonormal Bases for System Identification

The main motivation for using orthonormal basis functions in system identification and approximation is that the resulting model structure leads to a linear regressor form. In the following, we outline this procedure.

2.6.1 A linear regression form

Consider a discrete-time scalar system represented by

$$y(t) = g(t) * u(t). \tag{2.17}$$

Here, $y(t)$ is the output, $g(t)$ is the impulse response, $u(t)$ is the input, and the convolution operator is

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

Suppose the transfer function $G(z)$ is represented as

$$G(z) = \sum_{k=0}^{\infty} \theta(k)G_k(z),$$

where $G_k(z)$ are the orthonormal basis functions and $\theta(k)$ are the expansion coefficients. Then, (2.17) can be written as the following linear regressor form

$$y(t) = \phi^T(t)\theta$$

and

$$\phi^T(t)\theta = \sum_{k=0}^{\infty} \theta(k)[g_k(t) * u(t)],$$

where $g_k(t)$ are the impulse responses of the orthonormal basis functions.

Next, we examine some of the well-known orthonormal basis functions.

2.6.2 Orthonormal bases

FIR Basis

The simplest choice for the basis functions

$$G_k(z) = z^{-k}, \qquad k \geq 0$$

leads to the FIR model structures. However, the representation of a system with poles close to the unit circle in terms of this basis is not appropriate since the number of terms in the series expansion to provide an acceptable approximation of the system can be prohibitively high.

On the other hand, a prior knowledge of the underlying dynamics can be incorporated into the model structure leading to the popular use of the so-called Laguerre model.

Laguerre Basis

The Laguerre basis is given by [20, 22]

$$G_k(z) = \left(\frac{\sqrt{1-\xi^2}}{z-\xi}\right)\left(\frac{1-\xi z}{z-\xi}\right)^k, \qquad k \geq 0,$$

where $\xi \in \mathbb{R}$, $|\xi| < 1$ is a free parameter. The design parameter ξ is chosen in a range that matches the dominant dynamics of the system. Note that the FIR model is a special case of the Laguerre structure when $\xi = 0$.

The Laguerre basis allows only the incorporation of prior knowledge of non-resonant dynamics. When prior knowledge of a resonant mode exists it is more appropriate to use the Kautz basis.

Kautz Basis

The two-parameter Kautz basis is given by [23]

$$G_k(z) = \begin{cases} \frac{\sqrt{(1-a^2)(1-c^2)}}{z^2-a(c+1)z+c}\left(\frac{cz^2-a(c+1)z+1}{z^2-a(c+1)z+c}\right)^{\frac{k-1}{2}}, & k \text{ odd} \\ \frac{\sqrt{(1-c^2)}(z-a)}{z^2-a(c+1)z+c}\left(\frac{cz^2-a(c+1)z+1}{z^2-a(c+1)z+c}\right)^{\frac{k}{2}}, & k \text{ even} \end{cases}$$

where $|a| < 1$, $|c| < 1$, and $k \geq 0$. Note that the Laguerre model is a special case of the Kautz structure when the poles are real and equal.

For systems with several resonant/non-resonant dynamics, model structures allowing for the encoding of prior knowledge of several modes would be more desirable.

Generalized Orthonormal Bases

These basis functions are generated from a balanced realization of a given inner function [24, 25]. They allow the incorporation of prior information about a set of the poles of the underlying dynamics. In [28], the simple construction

$$G_k(z) = \left(\frac{\sqrt{1-|z_k|^2}}{z - z_k} \right) \prod_{i=0}^{k-1} \left(\frac{1-\bar{z}_i z}{z - z_i} \right), \qquad k \geq 0 \tag{2.18}$$

is proposed. This structure allows prior knowledge about an arbitrary number, say n, of modes $\{z_0, z_1, \cdots, z_{n-1}\}$ to be incorporated. Moreover, the well-known FIR, Laguerre, and Kautz models are all special cases of this construction.

3. SET MEMBERSHIP IDENTIFICATION WITH PERIODIC INPUTS AND ORTHONORMAL BASIS FUNCTIONS

In this chapter, we will consider periodic input signals and the orthonormal regressors. First, we will derive upper and lower probability bounds on the diameter of the membership-set. Then, we will study the central and the projection algorithms for the special case that the number of the unknown parameters equals the input period, in other words, the overparameterized model structures. We will derive explicit formulae for the central algorithm and the diameter of the membership-set, and characterize the set of all projection algorithms.

3.1 Problem Formulation

Consider the discrete-time scalar system in (2.9). For convenience, we repeat here:

$$y(t) = \phi^T(t)\theta + \eta(t), \qquad t = 1, 2, \cdots, N \tag{3.1}$$

where $y(t)$ is the system output, $\phi(t) \in \mathbb{R}^n$ is the measurable regression vector, $\theta \in \mathbb{R}^n$ is the unknown parameter vector, and $\eta(t)$ is the additive measurement noise.

We assume a linear regression model:

$$\phi^T(t)\theta = \sum_{k=1}^{n} \theta(k)[g_k(t) * u(t)], \tag{3.2}$$

where $u(t)$ is the applied input, and $g_k(t)$ are the impulse responses of the orthonormal basis functions [24, 28, 29]. Note that the orthonormal basis functions can be constructed, for example, as in (2.18).

The use of the basis functions defined by (2.18) is especially fruitful when some amount of data about the (main) time constants of the system at hand is available. This prior knowledge is reflected in the choice of the poles of the basis functions and the accuracy of this information has a strong influence on the number of significant parameters to be estimated.

In the model (3.1), we assume that $\eta(t)$ is a sequence of independent identically distributed random variables satisfying (2.10), that is,

$$\max_{1 \le t \le N} |\eta(t)| \le \varepsilon, \qquad \varepsilon > 0. \tag{3.3}$$

In [33], the volume of the membership-set and the optimal input design for FIR models and periodic inputs for the special case that the number of parameters equals the input period was studied in a statistical framework.

In this chapter, we study a generalization of this problem to the linear regression model (3.2). We derive exact expressions for the central algorithm (2.15) and the diameter of the membership-set (2.13), and characterize the solution set of the projection algorithms (2.16) denoted by $P_N(y,u)$, under the assumption that $u(t)$ is an m-periodic sequence:

$$u(t+im) = u(t), \qquad \text{for all integers } \; i \tag{3.4}$$

and m equals n. This restriction amounts to overparameterizing the linear regression model (3.2) where m represents an upper bound on the number of parameters to be estimated for a given periodic input signal.

In practice, the periodicity and the model order constraints are not essential for the application of the central and the projection algorithms. However, the statistical analysis of the resulting estimators is very difficult since in general it is impossible to give closed form expressions for them that are valid even asymptotically. In [36], the simple case of a single constant regressor is considered. We also note that if the chosen basis poles are close to the system poles, a small number of the basis functions is sufficient to accurately model the dynamics. Thus, from a practical standpoint (3.4) is not very restrictive as it looks.

In throughout the chapter, inequalities between vectors, such as $x \leq y$, are to be interpreted componentwise, and $\mathbf{1}$ denotes vectors with entries equal one.

3.2 Membership-Set with Periodic Inputs and Orthonormal Regressors

In this section, we show that $S_N(y,u,\varepsilon)$ is reduced to a convex polytope supported by $2m$ hyperplanes in $\mathbb{R}^n$ when $u(t)$ is an m-periodic sequence. Let U denote the circulant matrix

$$U = \begin{bmatrix} u(1) & u(m) & \cdots & u(2) \\ u(2) & u(1) & \cdots & u(3) \\ \vdots & \vdots & \ddots & \vdots \\ u(m) & u(m-1) & \cdots & u(1) \end{bmatrix}. \tag{3.5}$$

Define $\widehat{\mathcal{G}} \in \mathbb{R}^{m\times n}$ by

$$\widehat{\mathcal{G}}= [\widehat{g}_1 \;\; \widehat{g}_2 \;\; \cdots \;\; \widehat{g}_n] \,, \tag{3.6}$$

where $\widehat{g}_k$ denote the estimates of the impulse response coefficients of $G_k(z)$ defined by

$$\widehat{g}_k(i)=\frac{1}{m}\sum_{l=0}^{m-1} G_k(e^{j\frac{2\pi}{m}l})\, e^{-j\frac{2\pi}{m}li}, \qquad 0 \le i < m. \tag{3.7}$$

Lemma 3.2.1 *Consider the set $S_N(y,u,\varepsilon)$ defined by (2.11). Let*

$$\underline{y}(t)=\min_{i\ge 0} y(t+im), \qquad \overline{y}(t)=\max_{i\ge 0} y(t+im), \qquad t=1,2,\cdots,m. \tag{3.8}$$

Let $u(t)$, U, and $\widehat{\mathcal{G}}$ be as in (3.4), (3.5), and (3.6). Then,

$$S_N(y,u,\varepsilon) = \{\widehat{\theta} \in \mathbb{R}^n : \, \overline{y} - \varepsilon\mathbf{1} \le U\,\widehat{\mathcal{G}}\,\widehat{\theta} \le \underline{y} + \varepsilon\mathbf{1}\}. \tag{3.9}$$

Proof. The membership-set is described by the following inequalities

$$y(t) - \varepsilon \le \sum_{k=1}^{n} [g_k(t) * u(t)]\,\widehat{\theta}(k) \le y(t)+\varepsilon, \qquad t=1,2,\cdots,N. \tag{3.10}$$

Under the m-periodicity assumption on $u(t)$ note that

$$\begin{aligned}
\sum_{k=1}^{n}[g_k(t)*u(t)]\,\widehat{\theta}(k) &= \sum_{k=1}^{n}\sum_{\zeta=0}^{\infty} g_k(\zeta)u(t-\zeta)\,\widehat{\theta}(k) \\
&= \sum_{i=0}^{m-1} u(t-i)\sum_{k=1}^{n}\sum_{\zeta=0}^{\infty} g_k(i+m\zeta)\,\widehat{\theta}(k) \\
&= \kappa^T(t)\,\Xi\,\widehat{\theta}
\end{aligned}$$

where

$$\begin{aligned}
\kappa^T(t) &= [u(t) \;\; \cdots \;\; u(t-m+1)], \\
\Xi(i,k) &= \sum_{\zeta=0}^{\infty} g_k(i+m\zeta), \qquad 0\le i<m, \qquad 1\le k\le n.
\end{aligned}$$

Thus,

$$\max_{i\ge 0} y(t+im) - \varepsilon \le \kappa^T(t)\,\Xi\,\widehat{\theta} \le \min_{i\ge 0} y(t+im) + \varepsilon. \tag{3.11}$$

Now, evaluate

$$G_k(e^{j\theta}) = \sum_{\zeta=0}^{\infty} g_k(\zeta)\, e^{j\zeta\theta}$$

at the m roots of 1

$$G_k(e^{j\frac{2\pi}{m}l}) = \sum_{\zeta=0}^{\infty} g_k(\zeta)\, e^{j\frac{2\pi}{m}l\zeta}, \qquad l=0,\cdots,m-1 \tag{3.12}$$

and multiply (3.12) with $\frac{1}{m}\, e^{-j\frac{2\pi}{m}li}$ and sum over l to get

$$\begin{aligned}
\widehat{g}_k(i) &= \frac{1}{m}\sum_{l=0}^{m-1} G_k(e^{j\frac{2\pi}{m}l})\, e^{-j\frac{2\pi}{m}li} \\
&= \sum_{\zeta=0}^{\infty} g_k(\zeta)\frac{1}{m}\sum_{l=0}^{m-1} e^{j\frac{2\pi}{m}l(\zeta-i)} \\
&= \Xi(i,k).
\end{aligned} \tag{3.13}$$

Hence, (3.9) follows from (3.5)–(3.8), (3.11), and (3.13). ∎

Letting for $t = 1, 2, \cdots, m$,

$$\underline{\eta}(t) = \min_{i\geq 0} \eta(t+im), \qquad \overline{\eta}(t) = \max_{i\geq 0} \eta(t+im), \tag{3.14}$$

we can write $\underline{y}$ and $\overline{y}$ as

$$\underline{y} = U\,\widehat{\mathcal{G}}\,\theta + \underline{\eta}, \qquad \overline{y} = U\,\widehat{\mathcal{G}}\,\theta + \overline{\eta}. \tag{3.15}$$

Then, from Lemma 3.2.1

$$S_N(y,u,\varepsilon) = \theta \oplus \{\widetilde{\theta} \in \mathbb{R}^n : \overline{\eta} - \varepsilon\mathbf{1} \leq U\,\widehat{\mathcal{G}}\,\widetilde{\theta} \leq \underline{\eta} + \varepsilon\mathbf{1}\}. \tag{3.16}$$

Note that if $\eta(t)$ is a sequence of independent identically distributed random variables, then the marginal and the joint probability distribution functions of the order statistics $\underline{\eta}(t)$ and $\overline{\eta}(t)$ can be computed from a knowledge about the distribution function of $\eta(t)$, see for example Appendix A or [51]. In addition, $(\underline{\eta}(k), \overline{\eta}(k))$ and $(\underline{\eta}(l), \overline{\eta}(l))$ are independent pairs if $k \neq l$. This greatly simplifies the stochastic analysis of the membership-set, and therefore, the interpolatory estimators.

3.3 The Size of the Membership-Set

In this section, we obtain upper and lower probability bounds on the diameter of the membership-set. We remark that in the following $D(S_N(y,u,\varepsilon);\ell_2)$ denotes the membership-set diameter in the ℓ_2-norm.

Lemma 3.3.1 *Consider the system represented by (3.1). Let $u(t)$ be an m-periodic sequence; and $\underline{\eta}$ and $\overline{\eta}$ be as in (3.14). Let*

$$\mathcal{U} = U\widehat{\mathcal{G}}, \tag{3.17}$$

where U and $\widehat{\mathcal{G}}$ are as in (3.5) and (3.6), respectively. Let $\overline{\sigma}$ and $\underline{\sigma}$ denote the largest and the smallest singular values of $\mathcal{U}$. Then,

$$\frac{1}{\overline{\sigma}}\,(\varepsilon - \max_{1\leq t\leq m} e''(t)) \leq D(S_N(y,u,\varepsilon);\ell_2) \leq \frac{2\sqrt{m}}{\underline{\sigma}}\,(\varepsilon - \min_{1\leq t\leq m} e'(t)), \tag{3.18}$$

where for t=1,...,m,

$$e'(t) = \min\{-\underline{\eta}(t), \overline{\eta}(t)\}, \qquad e''(t) = \max\{-\underline{\eta}(t), \overline{\eta}(t)\}. \tag{3.19}$$

Proof. Let

$$\widetilde{S}_N(y,u,\varepsilon) = \{\widetilde{\theta} \in \mathbb{R}^n : \overline{\eta} - \varepsilon\mathbf{1} \leq \mathcal{U}\,\widetilde{\theta} \leq \underline{\eta} + \varepsilon\mathbf{1}\}. \tag{3.20}$$

From 3.16,

$$S_N(y,u,\varepsilon) = \theta \oplus \widetilde{S}_N(y,u,\varepsilon).$$

Thus,

$$D(S_N(y,u,\varepsilon);\ell_2) = D(\widetilde{S}_N(y,u,\varepsilon);\ell_2)$$

and $0 \in \widetilde{S}_N(y,u,\varepsilon)$. Let $x \in \widetilde{S}_N(y,u,\varepsilon)$. From the following inequalities

$$\|\mathcal{U}x\|_\infty \geq \frac{1}{\sqrt{m}}\,\|\mathcal{U}x\|_2 \geq \frac{\underline{\sigma}}{\sqrt{m}}\|x\|_2,$$

observe that there exists an index $1 \leq k^{(x)} \leq m$ which depends on x and satisfies

$$|\sum_{l=1}^{n} \mathcal{U}(k^{(x)},l)x_l| \geq \frac{\underline{\sigma}}{\sqrt{m}}||x||_2.$$

Equivalently,

$$\sum_{l=1}^{n} \mathcal{U}(k^{(x)},l)x_l \geq \frac{\underline{\sigma}}{\sqrt{m}}\|x\|_2 \quad \text{or} \quad \sum_{l=1}^{n} \mathcal{U}(k^{(x)},l)x_l \leq -\frac{\underline{\sigma}}{\sqrt{m}}\|x\|_2. \tag{3.21}$$

Moreover, from (3.20) the following inequalities

$$\overline{\eta}(k^{(x)}) - \varepsilon \leq \sum_{l=1}^{n} \mathcal{U}(k^{(x)},l)x_l \leq \underline{\eta}(k^{(x)}) + \varepsilon \tag{3.22}$$

are satisfied. Hence, we have from (3.22) when the first inequality in (3.21) holds

$$\|x\|_2 \leq \frac{\sqrt{m}}{\underline{\sigma}}\,(\varepsilon + \underline{\eta}(k^{(x)}) \leq \frac{\sqrt{m}}{\underline{\sigma}}\,\max_{1\leq t\leq m}(\varepsilon + \underline{\eta}(t))$$

and when the second inequality in (3.21) holds

$$\|x\|_2 \leq \frac{\sqrt{m}}{\underline{\sigma}}\,(\varepsilon - \overline{\eta}(k^{(x)}) \leq \frac{\sqrt{m}}{\underline{\sigma}}\,\max_{1\leq t\leq m}(\varepsilon - \overline{\eta}(t)).$$

Thus,

$$\|x\|_2 \leq \frac{\sqrt{m}}{\underline{\sigma}} \max_{1\leq t\leq m} \{\varepsilon + \underline{\eta}(t), \varepsilon - \overline{\eta}(t)\}, \quad \text{for all } x \in \widetilde{S}_N(y,u,\varepsilon). \tag{3.23}$$

Using the triangle inequality with (3.23) yields the upper bound in the lemma as follows

$$D(\widetilde{S}_N(y,u,\varepsilon);\ell_2) \leq \frac{2\sqrt{m}}{\underline{\sigma}} \max_{1\leq t\leq m} \{\varepsilon + \underline{\eta}(t), \varepsilon - \overline{\eta}(t)\}. \tag{3.24}$$

To establish the lower bound, we define a subset of $\widetilde{S}_N(y,u,\varepsilon)$ by

$$\widehat{S}_N(y,u,\varepsilon) = \{x \in \mathbb{R}^n : (\max_{1\leq t\leq m} \overline{\eta}(t) - \varepsilon)\mathbf{1} \leq \mathcal{U}x \leq (\min_{1\leq t\leq m} \underline{\eta}(t) + \varepsilon)\mathbf{1}\}$$

which contains the origin. Thus,

$$D(\widehat{S}_N(y,u,\varepsilon);\ell_2) \leq D(\widetilde{S}_N(y,u,\varepsilon);\ell_2). \tag{3.25}$$

Let $\tilde{x} \in \mathbb{R}^n$ be any vector such that

$$\|\tilde{x}\|_2 = \frac{1}{\overline{\sigma}}(\varepsilon - \max_{1\leq t\leq m}\{-\underline{\eta}(t), \overline{\eta}(t)\}). \tag{3.26}$$

Then, $\tilde{x} \in \widehat{S}_N(y,u,\varepsilon)$ and

$$D(\widehat{S}_N(y,u,\varepsilon);\ell_2) \geq \|\tilde{x}\|_2. \tag{3.27}$$

The inequalities (3.25) and (3.27) with (3.26) yield the lower bound in the lemma which completes the proof. ■

Observe from (3.18) that the upper and lower bounds on the diameter of the membership-set weakly depend on the inputs. In particular, they involve only the smallest and the largest singular values of $\mathcal{U}$.

Probability Bounds on the Diameter of the Membership-Set

Suppose that $\eta(t)$ is a sequence of independent identically distributed random variables in $[-\varepsilon, \varepsilon]$. Let $e'(t)$ and $e''(t)$ be as in (3.19); and $N = Mm$ for some positive integer M. Let

$$\hat{\nu} = \varepsilon - \frac{\nu\underline{\sigma}}{2\sqrt{m}}. \tag{3.28}$$

Taking the probabilities of the right-hand side inequality of (3.18), we get

$$\begin{aligned} \text{Prob}(D(S_N(y,u,\varepsilon);\ell_2) > \nu) &\leq \text{Prob}(\min_{1\leq t\leq m} e'(t) < \hat{\nu}) \\ &= 1 - \prod_{t=1}^{m} \text{Prob}(e'(t) \geq \hat{\nu}) \\ &= 1 - \prod_{t=1}^{m} \text{Prob}(\underline{\eta}(t) \leq -\hat{\nu},\, \overline{\eta}(t) \geq \hat{\nu}). \end{aligned} \tag{3.29}$$

In the following, we further assume that $\eta(t)$ has a symmetric distribution which does not contain any point masses except possibly at the end points of its support. Since

$$\begin{aligned}
\text{Prob}(\underline{\eta}(t) \leq -\widehat{\nu}, \overline{\eta}(t) \geq \widehat{\nu}) &= 1 - \text{Prob}(\underline{\eta}(t) > -\widehat{\nu}) - \text{Prob}(\overline{\eta}(t) < \widehat{\nu}) \\
&\quad + \text{Prob}(-\widehat{\nu} < \underline{\eta}(t) \leq \overline{\eta}(t) < \widehat{\nu}) \\
&= 1 - (1 - F_\eta(-\widehat{\nu}))^M - F_\eta^M(\widehat{\nu}) \\
&\quad + (F_\eta(\widehat{\nu}) - F_\eta(-\widehat{\nu}))^M \\
&= 1 - 2F_\eta^M(\widehat{\nu}) + (2F_\eta(\widehat{\nu}) - 1)^M,
\end{aligned}$$

the upper probability bound on the membership-set diameter is given by

$$\text{Prob}(D(S_N(y,u,\varepsilon);\ell_2) > \nu) \leq 1 - [1 - 2F_\eta^M(\widehat{\nu}) + (2F_\eta(\widehat{\nu}) - 1)^M]^m. \tag{3.30}$$

Note that if $0 < F_\eta(\widehat{\nu}) < 1$, then $\text{Prob}(D(S_N(y,u,\varepsilon);\ell_2) > \nu)$ converges to zero as $N \to \infty$.

Definition 3.3.1 *A bound ε on $\eta(t)$ is said to be tight bound, if for each $\mu > 0$ satisfying $\varepsilon > \mu$,*

$$\textit{Prob}(\eta(t) \geq \mu) > 0 \qquad \textit{and} \qquad \textit{Prob}(\eta(t) \leq -\mu) > 0.$$

Thus, if ε is a tight bound on $\eta(t)$, we have

$$\lim_{N\to\infty} \text{Prob}(D(S_N(y,u,\varepsilon);\ell_2) > \nu) = 0, \qquad \nu > 0.$$

Let

$$\nu^* = \varepsilon - \overline{\sigma}\nu. \tag{3.31}$$

Taking the probabilities of the left-hand side inequality of (3.18), we obtain the lower probability bound on the membership-set diameter as follows

$$\begin{aligned}
\text{Prob}(D(S_N(y,u,\varepsilon);\ell_2) > \nu) &\geq \text{Prob}(\max_{1\leq t\leq m} e''(t) < \nu^*) \\
&= [2F_\eta(\nu^*) - 1]^N. \qquad (3.32)
\end{aligned}$$

Note that if $F_\eta(\nu^*) = 0$ and N is even, then $\text{Prob}(D(S_N(y,u,\varepsilon);\ell_2) > \nu) = 1$. Thus, if ε is not a tight bound on $\eta(t)$, the membership-set can not converge to a singleton as noted in [33, 37].

3.4 An Overparameterized Model Structure

In the context of system identification, typically the model order is chosen on a trial-and-error basis after the input is chosen and the data are collected. With periodic inputs, it suffices to let $m \geq n$ for the parameter identifiability [1]. Then, we may interpret m as an upper bound on the number of parameters to be identified, and the constraint n equals m is equivalent to employing an overparameterized model structure. However, this choice greatly simplifies the ensuing analysis as shown in the following.

Assuming $m = n$, we let

$$\mathcal{V} = \widehat{\mathcal{G}}^{-1} U^{-1}. \tag{3.33}$$

Thus, from (3.16)

$$S_N(y, u, \varepsilon) = \theta \oplus \mathcal{V}\{\widetilde{x} \in \mathbb{R}^n : \overline{\eta} - \varepsilon \mathbf{1} \leq \widetilde{x} \leq \underline{\eta} + \varepsilon \mathbf{1}\}. \tag{3.34}$$

If the basis functions are uniformly exponentially stable, then uniformly in k,

$$\lim_{n\to\infty} \|\widehat{g}_k - g_k\|_\infty = 0. \tag{3.35}$$

In fact, the convergence rate is geometric. The basis functions in (2.18) are uniformly exponentially stable if and only if there exists an $r < 1$ such that for all n,

$$\max_{1\leq k\leq n} |z_k| < r. \tag{3.36}$$

Hence, provided that the basis functions are orthonormal and satisfy (3.36), the following holds

$$\lim_{n\to\infty} \widehat{\mathcal{G}}^{-1} = \lim_{n\to\infty} \widehat{\mathcal{G}}^T \tag{3.37}$$

where $\widehat{\mathcal{G}}^T$ denotes the transpose of $\widehat{\mathcal{G}}$.

It is a well-known fact that the inverse of a circulant matrix is also a circulant matrix [52]. Thus,

$$V = U^{-1}$$

is also a circulant matrix with the first row vector denoted by $[v(1)\ v(n)\ \cdots v(2)]$.

Ordinarily, the central algorithm is not robust-interpolatory if $m > n$ [13]. On the other hand, the following result establishes that for the special case considered in this section the central algorithm is also robust-interpolatory.

Theorem 3.4.1 *Consider the central algorithm in (2.15). Suppose $u(t)$ is an n-periodic sequence. Let $\underline{y}$ and $\overline{y}$ be as in (3.8); $\underline{\eta}$ and $\overline{\eta}$ be as in (3.14); and $\mathcal{V}$ be as in (3.33). Then,*

$$\widehat{\theta}_c = \mathcal{V}\,\frac{\overline{y}+\underline{y}}{2}. \tag{3.38}$$

Equivalently,

$$\widehat{\theta}_c = \theta + \mathcal{V}\frac{\overline{\eta}+\underline{\eta}}{2}. \tag{3.39}$$

Proof. From (3.34), for $k = 1, 2, \cdots, n$,

$$\max_{\widehat{\theta}\in S_N(y,u,\varepsilon)} \widehat{\theta}(k) = \theta(k) + \sum_{i=1}^{n} \mathcal{V}(k,i)\, x^{(k)}(i),$$

where $x^{(k)}$ is defined componentwise by

$$x^{(k)}(i) \triangleq \begin{cases} \underline{\eta}(i)+\varepsilon, & \text{if } \mathcal{V}(k,i) \ge 0 \\ \overline{\eta}(i)-\varepsilon, & \text{if } \mathcal{V}(k,i) < 0. \end{cases} \tag{3.40}$$

Notice that $\underline{\eta}(i)+\varepsilon \ge 0$ and $\overline{\eta}(i)-\varepsilon \le 0$ for all i. Likewise, for $k = 1, \cdots, n$,

$$\min_{\widehat{\theta}\in S_N(y,u,\varepsilon)} \widehat{\theta}(k) = \theta(k) + \sum_{i=1}^{n} \mathcal{V}(k,i)\, z^{(k)}(i),$$

where $z^{(k)}$ is defined componentwise by

$$z^{(k)}(i) \triangleq \begin{cases} \overline{\eta}(i)-\varepsilon, & \text{if } \mathcal{V}(k,i) \ge 0 \\ \underline{\eta}(i)+\varepsilon, & \text{if } \mathcal{V}(k,i) < 0. \end{cases} \tag{3.41}$$

Then,

$$\begin{aligned} \widehat{\theta}_c(k) &= \theta(k) + \frac{1}{2}\sum_{i=1}^{n} \mathcal{V}(k,i)[\underline{\eta}(i)+\overline{\eta}(i)] && (3.42)\\ &= \sum_{i=1}^{n}\sum_{l=1}^{n} \mathcal{V}(k,i)\,[U\widehat{\mathcal{G}}](i,l)\theta(l) + \frac{1}{2}\sum_{i=1}^{n} \mathcal{V}(k,i)\,[\underline{\eta}(i)+\overline{\eta}(i)] \\ &= \frac{1}{2}\sum_{i=1}^{n} \mathcal{V}(k,i)\,\left\{\underline{\eta}(i) + \sum_{l=1}^{n}[U\widehat{\mathcal{G}}](i,l)\theta(l) + \overline{\eta}(i) + \sum_{l=1}^{n}[U\widehat{\mathcal{G}}](i,l)\theta(l)\right\} \\ &= \frac{1}{2}\sum_{i=1}^{n} \mathcal{V}(k,i)\,[\underline{y}(i)+\overline{y}(i)] && (3.43) \end{aligned}$$

■

In Theorem 3.4.1, the central algorithm $\widehat{\theta}_c$ is expressed as a linear transformation of midrange estimators, which are independent identically distributed random variables if $\eta(t)$ is a sequence of independent identically distributed random variables. Thus once the distribution of $\eta(t)$ is given, the distribution function of $\widehat{\theta}_c$ can be computed without difficulty.

To illustrate the above result, suppose that $\eta(t)$ is a sequence of independent identically and symmetrically distributed random variables and we wish to compute the quantity

$$\sigma_c^2 = \mathbf{E}\{[\widehat{\theta}_c - \mathbf{E}(\widehat{\theta}_c)]^T[\widehat{\theta}_c - \mathbf{E}(\widehat{\theta}_c)]\} \tag{3.44}$$

where $\mathbf{E}\{x\}$ denotes the expected value of x. Then, $\widehat{\theta}_c$ is unbiased and from Theorem 3.4.1

$$\sigma_c^2 = \|\mathcal{V}\|_F^2 \, \mathbf{E}\left[\frac{\underline{\eta}(1) + \overline{\eta}(1)}{2}\right]^2, \tag{3.45}$$

where $\|\mathcal{V}\|_F$ is the Frobenius norm of $\mathcal{V}$ defined by

$$\|\mathcal{V}\|_F = \left(\sum_{k,l=1}^{n} [\mathcal{V}(k,l)]^2\right)^{1/2}.$$

It follows that the sum of the variances of $\widehat{\theta}_c(k)$ tends to zero in proportion to the variances of the midrange estimators. Interestingly, from (3.37) and the fact that the Frobenius norm is unitarily invariant we have

$$\lim_{n\to\infty} \|\mathcal{V}\|_F \to \|V\|_F$$

provided that the orthonormal basis functions used in the regression model (3.2) satisfy (3.36). This means σ_c^2 is independent from the chosen basis poles as $n \to \infty$. If the inputs are bounded as $|u(t)| \leq 1$ for all t, then $\|V\|_F$ is minimized by letting $U = I_n$ where I_n is the n by n identity matrix.

Remark 3.4.1 *Let*

$$\mu = \varepsilon\mathbf{1} - \frac{\overline{y} - \underline{y}}{2}. \tag{3.46}$$

From (3.34) and Theorem 3.4.1,

$$S_N(y,u,\varepsilon) = \widehat{\theta}_c \oplus \mathcal{V}\{\widetilde{\theta} \in \mathbb{R}^n : -\mu \leq \widetilde{\theta} \leq \mu\}.$$

Thus, the Chebyshev center of the membership-set does not depend on the norm used in measuring the size of $S_N(y,u,\varepsilon)$. This is due to the fact that the box $[-\mu,\mu]$ is convex balanced and $\mathcal{V}$ is a linear transformation.

Next, we derive an explicit formula for the diameter of the membership-set.

Theorem 3.4.2 *Consider the diameter of* $S_N(y,u,\varepsilon)$ *defined by (2.13). Suppose* $u(t)$ *is an* n*-periodic sequence. Let* $\underline{y}$ *and* $\overline{y}$ *be as in (3.8), and* $\mathcal{V}$ *be as in (3.33). Then,*

$$D(S_N(y,u,\varepsilon)) = 2 \max_{1\le k\le n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|\, \mu(i) \tag{3.47}$$

where μ *is as in (3.46).*

Proof. Let $\widehat{\theta}_1, \widehat{\theta}_2 \in S_N(y,u,\varepsilon)$. From (3.34), we have $\widehat{\theta}_1 = \theta + \mathcal{V}x$ and $\widehat{\theta}_2 = \theta + \mathcal{V}z$ for some x, z in the box $[\overline{\eta} - \varepsilon\mathbf{1}, \underline{\eta} + \varepsilon\mathbf{1}]$. Then, for $k = 1, \cdots, n$,

$$\begin{aligned} |\widehat{\theta}_1(k) - \widehat{\theta}_2(k)| &= |\sum_{i=1}^{n} \mathcal{V}(k,i)\, [x(i) - z(i)]| \\ &\le \sum_{i=1}^{n} |\mathcal{V}(k,i)|\, [2\varepsilon + \underline{\eta}(i) - \overline{\eta}(i)]. \end{aligned}$$

Hence,

$$D(S_N(y,u,\varepsilon)) \le \max_{1\le k\le n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|\, [2\varepsilon + \underline{\eta}(i) - \overline{\eta}(i)]. \tag{3.48}$$

For the reverse inequality, suppose that the maximum on the right-hand side of (3.48) is attained for $k = k^*$. Let $x^{(k)}$ and $z^{(k)}$ be as in (3.40) and (3.41). Put $\widehat{\theta}_1 = \theta + \mathcal{V}x^{(k^*)}$ and $\widehat{\theta}_2 = \theta + \mathcal{V}z^{(k^*)}$. Then,

$$\begin{aligned} \widehat{\theta}_1(k^*) - \widehat{\theta}_2(k^*) &= \sum_{i=1}^{n} \mathcal{V}(k^*,i)\, [x^{(k^*)}(i) - z^{(k^*)}(i)] \\ &= \sum_{i=1}^{n} |\mathcal{V}(k^*,i)|\, [2\varepsilon + \underline{\eta}(i) - \overline{\eta}(i)]. \end{aligned}$$

Hence,

$$\begin{aligned} \|\widehat{\theta}_1 - \widehat{\theta}_2\|_\infty &\ge \sum_{i=1}^{n} |\mathcal{V}(k^*,i)|\, [2\varepsilon + \underline{\eta}(i) - \overline{\eta}(i)] \\ &= \max_{1\le k\le n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|\, [2\varepsilon + \underline{\eta}(i) - \overline{\eta}(i)] \end{aligned}$$

and therefore,

$$D(S_N(y,u,\varepsilon)) \ge \max_{1\le k\le n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|\, [2\varepsilon + \underline{\eta}(i) - \overline{\eta}(i)]. \tag{3.49}$$

The inequalities (3.48), (3.49), and the equalities in (3.15) yield (3.47). ■

Next, we study the characterization of the set of all projection algorithms.

Corollary 3.4.1 *Consider the projection algorithms defined by (2.16). Suppose* $u(t)$ *is an* n*-periodic sequence. Let* $\underline{y}$ *and* $\overline{y}$ *be as in (3.8), and* $\mathcal{V}$ *be as in (3.33). Let*

$$e_p = \max_{1\le k\le n} \frac{\overline{y}(k) - \underline{y}(k)}{2} \tag{3.50}$$

and $\widehat{\theta}_c$ be as in (2.15). Then, $P_N(y,u) = S_N(y,u,e_p)$ and its Chebyshev center is $\widehat{\theta}_c$. Furthermore,

$$D(P_N(y,u)) = \max_{1\le k\le n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|\,[2e_p + \underline{y}(i) - \overline{y}(i)]. \tag{3.51}$$

Proof. From (2.16), $\widehat{\theta}_p \in P_N(y,u)$ if and only if $\widehat{\theta}_p$ is a solution of the following linear programming problem:

$$\begin{cases} \quad\text{min} & e \\ \text{subject to} & \widehat{\theta}_p \in S_N(y,u,e) \end{cases} \tag{3.52}$$

From (3.34), we have

$$S_N(y,u,e) = \mathcal{V}\{\tilde{x} \in \mathbb{R}^n : \overline{y} - e\mathbf{1} \le \tilde{x} + U\widehat{\mathcal{G}}\theta \le \underline{y} + e\mathbf{1}\}. \tag{3.53}$$

A necessary condition for non-empty $S_N(y,u,e)$ is derived from (3.53) as follows

$$e \ge \max_{1\le k\le n} \frac{\overline{y}(k) - \underline{y}(k)}{2}. \tag{3.54}$$

Observe that $S_N(y,u,e_p)$ is non-empty, and from (3.54) the minimum in (3.52) is achieved for e_p. Hence, $S_N(y,u,e_p)$ is the set of all possible solutions of (2.16). From Theorem 3.4.1 recall that $\widehat{\theta}_c$ is the Chebyshev center of $S_N(y,u,\varepsilon)$ for all $\varepsilon > 0$. Thus, it is the Chebyshev center of $P_N(y,u) = S_N(y,u,e_p)$. The last assertion follows from (3.47) with the substitution $\varepsilon = e_p$. ■

From (3.50) and (3.34), notice that $P_N(y,u)$ is the image of an $n-1$ dimensional box under the linear transformation $\mathcal{V}$. Thus, $P_N(y,u)$ is contained in a hyperplane. However, its diameter relative to $D(S_N(y,u,\varepsilon))$ may be rather large, in particular for large n as will be demonstrated next.

Suppose $\eta(t)$ is a sequence of independent identically and symmetrically distributed random variables in $[-\varepsilon,\varepsilon]$. Let $N = Mn$. Observe that if ε is a tight bound on $\eta(t)$, $\mu(k)$ converges to zero almost everywhere for all k as $M \to \infty$.

Proposition 3.4.1 *Consider the central and the projection algorithms in (2.15) and (2.16). Let $\mathcal{V}$ be as in (3.33) and $N = Mn$. Consider the basis functions in (2.18). Assume that there exists a constant $0 < c < \infty$ such that for all n,*

$$\max_{1\le k\le n} \sum_{i=1}^{n} |\mathcal{V}(k,i)| \le c \max_{1\le k,i\le n} |\mathcal{V}(k,i)|. \tag{3.55}$$

Let $\eta(t)$ be a sequence of independent identically and symmetrically distributed random variables in $[-\varepsilon, \varepsilon]$. Suppose $u(t)$ is an n-periodic sequence and ε is a tight bound on $\eta(t)$. Then, for each fixed $M < \infty$

$$D(P_N(y,u)) \to D(S_N(y,u,\varepsilon)) \qquad \text{a.e.} \qquad (n \to \infty).$$

Proof. From (3.47) and (3.51),

$$D(P_N) \geq D(S_N) - 2(\varepsilon - e_p) \max_{1\leq k\leq n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|.$$

Thus, from (3.50) we get

$$\frac{D(P_N)}{D(S_N)} \geq 1 - \frac{\max\limits_{1\leq k\leq n} \sum\limits_{i=1}^{n} |\mathcal{V}(k,i)| \min\limits_{1\leq k\leq n} \mu(k)}{\max\limits_{1\leq k\leq n} \sum\limits_{i=1}^{n} |\mathcal{V}(k,i)|\,\mu(i)}. \tag{3.56}$$

Choose k^* and i^* such that

$$|\mathcal{V}(k^*, i^*)| = \max_{1\leq k,i\leq n} |\mathcal{V}(k,i)|.$$

Then,

$$\max_{1\leq k,i\leq n} |\mathcal{V}(k,i)|\,\mu(i*) \quad\leq\quad \max_{1\leq k\leq n} \sum_{i=1}^{n} |\mathcal{V}(k,i)|\,\mu(i).$$

Hence, from (3.56) and the hypothesis we have

$$\frac{D(P_N)}{D(S_N)} \geq 1 - c \min_{1\leq k\leq n} \frac{\mu(k)}{\mu(i*)}$$

for some absolute positive constant $c < \infty$.

Let δ be a given number satisfying $0 < \delta < 1$. Then,

$$\begin{aligned} &\text{Prob}(\mu(i^*) > (c/\delta) \min_{1\leq k\leq n} \mu(k)) \\ &\leq \text{Prob}\left(\frac{D(P_N)}{D(S_N)} > 1-\delta\right). \end{aligned} \tag{3.57}$$

Since $\mu(k)$, $k \geq 1$ are independent identically distributed random variables,

$$\begin{aligned} &\text{Prob}(\mu(i^*) > (c/\delta) \min_{1\leq k\leq n} \mu(k)|\, \mu(i^*) = x) \\ &\qquad = 1 - [\text{Prob}\,(\mu(1) \geq \delta x/c)]^{n-1}. \end{aligned}$$

Integrating the above conditional probability, we get

$$
\begin{aligned}
& \mathrm{Prob}(\mu(i^*) > (c/\delta) \min_{1\le k\le n} \mu(k)) \\
= & \int_0^{\varepsilon} \mathrm{Prob}(\mu(i^*) > \frac{c}{\delta} \min_k \mu(k) |\, \mu(i^*) = x) f_{\mu(i^*)}(x)\, dx \\
\ge & \int_{x^*}^{\varepsilon} \mathrm{Prob}(\mu(1) > \frac{c}{\delta} \min_k \mu(k) |\, \mu(1) = x) f_{\mu(1)}(x)\, dx \\
\ge & \mathrm{Prob}(x^* > (c/\delta) \min_{k\ne 1} \mu(k))\, \mathrm{Prob}(\mu(1) \ge x^*) \\
= & \{1 - [\mathrm{Prob}\,(\mu(1) > \delta x^*/c)]^{n-1}\}\, \mathrm{Prob}(\mu(1) \ge x^*)
\end{aligned}
$$

where $0 < x^* < \varepsilon$ is a number to be fixed next.

Now, choose x^* such that

$$1 - \delta \le \mathrm{Prob}\,(\mu(1) > x^*) < 1$$

which is possible since the probability in the middle tends to one as x^* approaches to zero that is a tight lower bound on $\mu(1)$. Hence,

$$\lim_{n\to\infty} \inf \mathrm{Prob}(\mu(i^*) > (c/\delta) \min_{1\le k\le n} \mu(k)) \ge 1 - \delta.$$

Therefore, from (3.57) we have

$$\mathrm{Prob}\left(\frac{D(P_N)}{D(S_N)} > 1 - \delta \right) \ge 1 - \delta.$$

Since δ is arbitrary, it follows that $D(P_N) \to D(S_N)$ in probability as $n \to \infty$. Recall that for bounded sequences of random variables, convergence in probability, convergence in mean, and convergence almost everywhere are all equivalent, for convergence of random variables see Appendix B. ■

As a special case of the regression model in (3.2), let us consider the following FIR model:

$$\phi^T(t)\theta = \sum_{k=1}^{n} u(t-k)\theta(k).$$

Then, the condition (3.55) is satisfied if there exists some constant $0 < c < \infty$ such that for all n,

$$\sum_{k=1}^{n} |v(k)| \le c \max_{1\le k\le n} |v(k)|.$$

This condition is satisfied, for example, by all n-periodic pulse inputs:

$$u(t) = \begin{cases} 1, & t = n, \\ 0, & t = 1, \cdots, n-1. \end{cases} \tag{3.58}$$

Example Suppose $\eta(t)$ is a sequence of independent identically and uniformly distributed random variables in $[-0.1, 0.1]$, $u(t)$ is as in (3.58), and for all k, z_k satisfies $z_k = \frac{2}{3k}$. For $M = 100$, and $n = 5, 15, 25, 35, 45$, we computed the ratio $D(P_N(y,u))/D(S_N(y,u,\varepsilon))$ as 0.56, 0.95, 0.95, 0.86, 0.90, respectively. The purpose of this simulation example was to demonstrate that $P_N(y,u)$ could be a rather large subset of $S_N(y,u,\varepsilon)$ even for modest values of n; and the convergence in Proposition 3.4.1 may take place slowly, as evidenced by the fluctuations in the values of $D(P_N(y,u))/D(S_N(y,u,\varepsilon))$.

Next, with the same set of z_k and $n = 5$, we computed $D(S_N(y,u,\varepsilon)) = 0.0088$, 0.0012, 0.000041, respectively for $M = 100$, 1000, 10000. Notice that the diameter of the membership-set shrinks to zero by about an order of magnitude of ten. This is due to the fact that the minimax estimates $\underline{y}$ and $\overline{y}$ defined in (3.8), which describe the membership-set (3.9), have standard deviations tending to zero as fast as $O(\frac{1}{M})$ for the uniformly distributed noise.

The mean-squared convergence properties of the estimators $\widehat{\theta}_c$ and $\widehat{\theta}_p$ can be deduced from Proposition 3.4.1 and Corollary 3.4.1 when $\eta(t)$ is a sequence of independent identically distributed random variables. The properties pertaining to fast convergence of $\widehat{\theta}_c$ and $\widehat{\theta}_p$ have already been investigated in [36, 37, 33] in more general settings. Then, by utilizing Proposition 3.4.1 and Corollary 3.4.1, one can complement the results in [36, 37, 33].

In the rest of this chapter, we will briefly study the least-squares algorithm for the special case considered in this chapter. Though not necessary, for a comparison with the estimators $\widehat{\theta}_c$ and $\widehat{\theta}_p$, we let $N = Mm$. It is fairly easy to show that the least-squares estimator of θ denoted by $\widehat{\theta}_{ls}$ is given by

$$\widehat{\theta}_{ls} = \mathcal{V}\widetilde{y} = \theta + \widetilde{\eta} \tag{3.59}$$

where for $k = 1, 2, \cdots, m$, $\widetilde{y}$ and $\widetilde{\eta}$ are defined by

$$\widetilde{y}(k) = \frac{1}{M}\sum_{i=0}^{M-1} y(k+im), \qquad \widetilde{\eta}(k) = \frac{1}{M}\sum_{i=0}^{M-1} \eta(k+im).$$

The least-squares estimator (3.59) obtained by minimizing the quadratic norm of the prediction errors [1] is unbiased and has a variance decaying to zero as $O(\frac{1}{M})$ when $\eta(t)$ is a sequence of zero-mean, independent identically distributed random variables.

The choice between the estimators $\widehat{\theta}_c$ (or $\widehat{\theta}_p$) and $\widehat{\theta}_{ls}$ depends on the distribution of $\eta(t)$. If $\eta(t)$ has a heavy-tailed distribution, $\widehat{\theta}_c$ and $\widehat{\theta}_p$ have better convergence properties than $\widehat{\theta}_{ls}$ [37] whereas for a noise distribution with thin tails, (3.44) implies a slower converge rate for the variance of $\widehat{\theta}_c$ than that of $\widehat{\theta}_{ls}$. In the latter case, (3.47) and (3.51) imply large mean values for the diameters of $S_N(y, u, \varepsilon)$ and $P_N(y, u)$.

For scalar estimation problems with bounded disturbances, the statistical robustness of the quadratic norm was established in [36]. The results of this chapter can be used to extend this conclusion from a single constant regressor to the regression model (3.2) when m equals n and the periodic inputs as in (3.4) are used. A somewhat more general statistical robustness problem was formulated and solved in [53].

3.5 Summary

In this chapter, we studied the central and the projection identification algorithms with periodic input signals and orthonormal basis functions and derived an explicit formula for the former for a special case. Also derived were the expressions for the diameters of the membership-set and the set of all projection algorithms. These results should be useful in analyzing statistical properties of these estimators when it is possible to employ overparameterized model structures with arbitrary basis functions and periodic inputs.

4. STATISTICAL ANALYSIS OF CENTRAL AND MINIMAX ALGORITHMS

In this chapter, we will study the statistical properties of the central and the minimax algorithms in a one-dimensional parameter space setting assuming that the regressor signal and its inverse are magnitude bounded. We will derive non-asymptotic, order-tight, lower and upper bounds on the convergence rate of the parameter estimate variance for the central and the minimax algorithms. This presents an extension of the previous work for constant scalar regressors to arbitrary scalar regressors save for magnitude constraints.

4.1 Problem Formulation

Consider the problem of estimating an unknown scalar parameter θ in the model

$$y(t) = u(t)\theta + \eta(t), \qquad t = 1, 2, \cdots, N \tag{4.1}$$

where $y(t)$ is the corrupted measurement, $u(t)$ is the measurable regressor, and $\eta(t)$ is the measurement noise. We assume that $u(t)$ is a deterministic signal and $\eta(t)$ is a sequence of independent identically distributed random variables satisfying (2.10).

In this simplified setup, the membership-set in (2.11) is given by

$$S_N(y, u, \varepsilon) = \bigcap_{t=1}^{N} \{\widehat{\theta} \in \mathbb{R} : |y(t) - u(t)\widehat{\theta}| \leq \varepsilon\}. \tag{4.2}$$

Moreover, from (2.12), (2.13), and (2.14), the radius, the diameter, and the Chebyshev center of $S_N(y, u, \varepsilon)$ are given respectively by

$$R(S_N(y, u, \varepsilon)) = \min_{\theta' \in S_N(y,u,\varepsilon)} \max_{\theta'' \in S_N(y,u,\varepsilon)} |\theta' - \theta''|, \tag{4.3}$$

$$D(S_N(y, u, \varepsilon)) = \max_{\theta', \theta'' \in S_N(y,u,\varepsilon)} |\theta' - \theta''|, \tag{4.4}$$

and

$$\widehat{\theta}_c = \arg \min_{\theta' \in S_N(y,u,\varepsilon)} \max_{\theta'' \in S_N(y,u,\varepsilon)} |\theta' - \theta''|. \tag{4.5}$$

A quite related approach to (4.5) is the *minimax* estimator defined by

$$\widehat{\theta}_m = \arg \min_{\widehat{\theta} \in S_N(y,u,\varepsilon)} \max_{1 \leq t \leq N} |y(t) - u(t)\widehat{\theta}|. \tag{4.6}$$

The minimax estimator (4.6) has some robustness and near-optimality properties [42, 50, 29] with respect to unknown but bounded disturbances.

In [36], formal conditions for the distribution of noise for which the minimax algorithm parameter estimate variance converges to zero faster than that of the least squares algorithm $O(N^{-1})$ were developed when the regressor signal equals to one all the time. This result was extended in [37] to multi-dimensional persistently exciting regressors under the same noise assumptions.

The purpose of this chapter is to study the statistical properties of the central and the minimax algorithms in the simplified setup (4.1) assuming that the regressor and its inverse are magnitude bounded. In [37], non-asymptotic, tight-in-the (convergence) rate, upper and lower probability bounds on the membership-set diameter were derived. These results are applicable only in one direction: if the probability density function of the noise does not vanish or has point masses at the boundary of its support, then the upper probability bound on $D(S_N(y, u, \varepsilon))$ provides also an upper bound on the convergence rate of the parameter estimate variance, which is faster than $O(N^{-1})$. Furthermore, this result applies not only to the central and the minimax algorithms but also to all interpolatory algorithms. However, a lower probability bound on $D(S_N(y, u, \varepsilon))$ does not necessarily lead to a lower bound on the convergence rate of the parameter estimate variance as pointed out in [37].

In this chapter, we consider finite sample properties of the estimators (4.5) and (4.6). In real life, only a finite number of data is available. On the other hand, almost all results in mainstream identification apply to asymptotic properties [1]. Our work is quite different than [46, 47, 54] in which finite sample properties of the quadratic prediction error criterion are investigated. For other criteria, in general, there is no direct link between the criterion and the quality of the estimate, which is certainly true for the central algorithm (4.5) and the minimax estimate (4.6).

4.2 Statistical Analysis of the Central Algorithm

In this section, we analyze the statistical properties of the central algorithm for the model (4.1) assuming that the regressor and its inverse are magnitude bounded. In the analysis, we also impose certain restrictions on the pdf of the measurement noise denoted by f_η. For a scalar-valued function $x(t)$, we define its signature function by

$$\operatorname{sgn}[x(t)]=\begin{cases} x(t)/|x(t)|, & x(t)\neq 0 \\ 0, & x(t)=0. \end{cases}$$

We first show that the membership-set is a closed interval when $\theta \in \mathbb{R}$.

Lemma 4.2.1 *Consider the set $S_N(y,u,\varepsilon)$ defined by (4.2). Let*

$$x_1=\max_{u(t)\neq 0} \frac{y(t)\operatorname{sgn}[u(t)]-\varepsilon}{|u(t)|}, \qquad x_N=\min_{u(t)\neq 0} \frac{y(t)\operatorname{sgn}[u(t)]+\varepsilon}{|u(t)|}. \tag{4.7}$$

Then, $S_N(y,u,\varepsilon)$ is a closed interval with the endpoints x_1 and x_N.

Proof. The membership-set is described by the inequalities

$$y(t)-\varepsilon \leq u(t)\widehat{\theta} \leq y(t)+\varepsilon, \qquad t=1,2,\cdots,N. \tag{4.8}$$

Since the intervals in (4.8) coincide with the entire real line whenever $u(t)$ equals 0, we may disregard such intervals. Then, from (4.8) for each $\widehat{\theta}\in S_N(y,u,\varepsilon)$ we have

$$\max_{u(t)>0} \frac{y(t)-\varepsilon}{u(t)} \leq \widehat{\theta} \leq \min_{u(t)>0} \frac{y(t)+\varepsilon}{u(t)}$$

and

$$\max_{u(t)<0} \frac{y(t)+\varepsilon}{u(t)} \leq \widehat{\theta} \leq \min_{u(t)<0} \frac{y(t)-\varepsilon}{u(t)}.$$

Thus,

$$\begin{aligned} \widehat{\theta} &\leq \min\left\{ \min_{u(t)>0} \frac{y(t)+\varepsilon}{u(t)}, \min_{u(t)<0} \frac{y(t)-\varepsilon}{u(t)} \right\} \\ &= \min_{u(t)\neq 0} \frac{y(t)\operatorname{sgn}[u(t)]+\varepsilon}{|u(t)|}. \end{aligned}$$

Likewise,

$$\widehat{\theta} \geq \max_{u(t)\neq 0} \frac{y(t)\operatorname{sgn}[u(t)]-\varepsilon}{|u(t)|}.$$

It follows that $\widehat{\theta}\in[x_1,x_N]$. The reverse inclusion follows also from the above inequalities. ∎

Thus, from Lemma 4.2.1, we have the following closed form expressions for the Chebyshev center and the radius of $S_N(y,u,\varepsilon)$:

$$\widehat{\theta}_c = \frac{x_1+x_N}{2}, \qquad R\left(S_N(y,u,\varepsilon)\right) = \frac{x_N-x_1}{2}. \tag{4.9}$$

The extreme order statistics x_1 and x_N can be written as

$$x_1 = \theta + \widetilde{x}_1, \qquad x_N = \theta + \widetilde{x}_N, \tag{4.10}$$

where

$$\widetilde{x}_1 = \max_{u(t)\neq 0} \frac{\widetilde{\eta}(t) - \varepsilon}{|u(t)|}, \qquad \widetilde{x}_N = \min_{u(t)\neq 0} \frac{\widetilde{\eta}(t) + \varepsilon}{|u(t)|}, \tag{4.11}$$

and

$$\widetilde{\eta}(t) = \eta(t)\,\mathrm{sgn}[u(t)], \qquad t = 1, 2, \cdots, N$$

is a sequence of independent random variables provided that $\eta(t)$ is a sequence of independent random variables. Note that even if $\eta(t)$ is a sequence of independent identically distributed random variables, $\widetilde{\eta}(t)$ is not necessarily an identically distributed sequence unless $\mathrm{sgn}[u(t)]$ is constant or f_η has a symmetry with respect to the origin, i.e., $f_\eta(s) = f_\eta(-s)$ for all s. It is a known fact that the statistical analysis of the central algorithm becomes much simpler when f_η does not vanish at the endpoints of its support denoted by $\pm\varepsilon$ [36]. Hereinafter, we shall therefore assume that $\eta(t)$ has a symmetrical pdf f_η that vanishes outside the open interval $(-\varepsilon, \varepsilon)$.

We define the distribution function of $\eta(t)$ by

$$F_\eta(s) = \mathrm{Prob}(\eta(t) \le s) = \int_{-\infty}^{s} f_\eta(v)\, dv.$$

Our analysis starts with a computation of the probabilities of some rare events generated by the random variables $\widetilde{x}_1$ and $\widetilde{x}_N$ in the following.

Lemma 4.2.2 *Let $\widetilde{x}_1$ and $\widetilde{x}_N$ be as in (4.11). Suppose that $\eta(t)$ is a sequence of independent identically and symmetrically distributed random variables with a pdf supported in $[-\varepsilon, \varepsilon]$. Then, for all $\xi \le 0$ and $\zeta \ge 0$,*

$$\mathrm{Prob}(\widetilde{x}_1 \le \xi) = \prod_{t=1}^{N} [1 - F_\eta(-\varepsilon - \xi|u(t)|)], \tag{4.12}$$

$$\mathrm{Prob}(\widetilde{x}_N > \zeta) = \prod_{t=1}^{N} [1 - F_\eta(-\varepsilon + \zeta|u(t)|)], \tag{4.13}$$

$$\mathrm{Prob}(\widetilde{x}_1 \le \xi, \widetilde{x}_N > \zeta) = \prod_{t=1}^{N} [1 - F_\eta(-\varepsilon - \xi|u(t)|) - F_\eta(-\varepsilon + \zeta|u(t)|)]. \tag{4.14}$$

Proof. From (4.11), the independence of the random variables $\widetilde{\eta}(t)$, and the fact that $\widetilde{\eta}(t)$ and $\eta(t)$ have the same distribution due to the symmetry assumption, we have

$$\begin{aligned}
\text{Prob}(\widetilde{x}_1 \le \xi) &= \text{Prob}(\widetilde{\eta}(t) \le \varepsilon + \xi|u(t)|,\ t = 1, \cdots, N) \\
&= \prod_{t=1}^{N} \text{Prob}(\widetilde{\eta}(t) \le \varepsilon + \xi|u(t)|) \\
&= \prod_{t=1}^{N} [1 - F_{\widetilde{\eta}}(-\varepsilon - \xi|u(t)|)].
\end{aligned}$$

The proofs of the remaining identities are similar. ■

The following lemma enables us to express the first two moments of $\widetilde{x}_1$ and $\widetilde{x}_N$ as the integrals of the probabilities in (4.12)–(4.14).

Lemma 4.2.3 *Let x and z be random variables supported in the compact interval $[a, b]$. Then,*

$$\mathbf{E}(x) = a + \int_a^b \text{Prob}(x > x_0)\, dx_0, \tag{4.15}$$

$$\mathbf{E}(x^2) = a^2 + 2\int_a^b \text{Prob}(x > x_0) x_0\, dx_0, \tag{4.16}$$

$$\mathbf{E}(xz) = b\mathbf{E}(x) + a\mathbf{E}(z) - ab - \int_a^b \int_a^b \text{Prob}(x > x_0, z \le z_0)\, dx_0 dz_0, \tag{4.17}$$

$$\mathbf{E}(x)\mathbf{E}(z) = b\mathbf{E}(x) + a\mathbf{E}(z) - ab - \int_a^b \int_a^b \text{Prob}(x > x_0)\text{Prob}(z \le z_0) dx_0 dz_0. \tag{4.18}$$

Proof. We remark that (4.15) and (4.16) might appear in some standard probability textbooks. However, their proofs are included in the following for the sake of completeness. Using the formula for integration by parts, we obtain the identity in (4.15) as follows

$$\begin{aligned}
\mathbf{E}(x) &= \int_a^b x_0 f_x(x_0)\, dx_0 = b\int_a^b f_x(s)\, ds - \int_a^b \int_a^{x_0} f_x(s)\, ds\, dx_0 \\
&= b - \int_a^b \text{Prob}(x \le x_0)\, dx_0 \\
&= a + \int_a^b \text{Prob}(x > x_0)\, dx_0.
\end{aligned} \tag{4.19}$$

Similarly,

$$\mathbf{E}(x^2) = \int_a^b x_0^2 f_x(x_0)\, dx_0 = b^2 \int_a^b f_x(s)\, ds - 2\int_a^b x_0 \int_a^{x_0} f_x(s)\, ds\, dx_0$$

$$
\begin{aligned}
&= b^2 - 2\int_a^b \text{Prob}(x \leq x_0) x_0\, dx_0 \\
&= a^2 + 2\int_a^b \text{Prob}(x > x_0) x_0\, dx_0.
\end{aligned} \tag{4.20}
$$

Next, from several applications of the formula for integration by parts

$$
\begin{aligned}
\mathbf{E}(xz) &= \int_a^b \int_a^b x_0 z_0 f_{x,z}(x_0, z_0)\, dx_0\, dz_0 \\
&= \int_a^b \left\{ b\int_a^b f_{x,z}(s, z_0)\, ds - \int_a^b \int_a^{x_0} f_{x,z}(s, z_0)\, ds\, dx_0 \right\} z_0\, dz_0 \\
&= b\int_a^b z_0 f_z(z_0)\, dz_0 - b\int_a^b \int_a^b \int_a^{x_0} f_{x,z}(s, l)\, ds\, dx_0\, dl \\
&\quad + \int_a^b \int_a^{z_0} \int_a^b \int_a^{x_0} f_{x,z}(s, l)\, ds\, dx_0\, dl\, dz_0 \\
&= b\,\mathbf{E}(z) - b\int_a^b \int_a^b \int_a^{x_0} f_{x,z}(s, l)\, ds\, dl\, dx_0 \\
&\quad + \int_a^b \int_a^b \int_a^{z_0} \int_a^{x_0} f_{x,z}(s, l)\, ds\, dl\, dx_0\, dz_0 \\
&= b\,\mathbf{E}(z) - b\int_a^b \text{Prob}(x \leq x_0)\, dx_0 \\
&\quad + \int_a^b \int_a^b \text{Prob}(x \leq x_0, z \leq z_0)\, dx_0\, dz_0
\end{aligned} \tag{4.21}
$$

with

$$
\begin{aligned}
\int_a^b \int_a^b \text{Prob}(x \leq x_0, z \leq z_0)\, dx_0\, dz_0 &= \int_a^b \int_a^b \text{Prob}(z \leq z_0)\, dx_0\, dz_0 \\
&\quad - \int_a^b \int_a^b \text{Prob}(x > x_0, z \leq z_0)\, dx_0\, dz_0
\end{aligned}
$$

and several uses of (4.19) yield the identity in (4.17). The identity in (4.18) follows similarly from several applications of (4.19). ■

The variance of $\widehat{\theta}_c$ denoted by $\sigma^2_{\widehat{\theta}_c}$ is given by

$$
\sigma^2_{\widehat{\theta}_c} = (1/4)\,(\sigma^2_{\widetilde{x}_1} + \sigma^2_{\widetilde{x}_N}) + (1/2)\,\sigma_{\widetilde{x}_1}\sigma_{\widetilde{x}_N}\, r_{\widetilde{x}_1\widetilde{x}_N} \tag{4.22}
$$

where $r_{\widetilde{x}_1\widetilde{x}_N}$ is the correlation coefficient of $\widetilde{x}_1$ and $\widetilde{x}_N$ defined by

$$
r_{\widetilde{x}_1\widetilde{x}_N} = \frac{\mathbf{E}(\widetilde{x}_1\widetilde{x}_N) - \mathbf{E}(\widetilde{x}_1)\mathbf{E}(\widetilde{x}_N)}{\sigma_{\widetilde{x}_1}\sigma_{\widetilde{x}_N}}. \tag{4.23}
$$

Similarly, the variance of $R(S_N(y, u, \varepsilon))$ denoted by $\sigma^2_{R(S_N)}$ can be written as

$$
\sigma^2_{R(S_N)} = (1/4)(\sigma^2_{\widetilde{x}_1} + \sigma^2_{\widetilde{x}_N}) - (1/2)\sigma_{\widetilde{x}_1}\sigma_{\widetilde{x}_N} r_{\widetilde{x}_1\widetilde{x}_N}. \tag{4.24}
$$

Recall that magnitude of a correlation coefficient is not greater than unity. However, for the extreme order statistics defined in (4.11), we have the following more precise result.

Proposition 4.2.1 *Let $\widetilde{x}_1$ and $\widetilde{x}_N$ be as in (4.11). Consider the correlation coefficient of the extreme order statistics $r_{\widetilde{x}_1\widetilde{x}_N}$ defined by (4.23). Suppose that $\eta(t)$ is a sequence of independent identically and symmetrically distributed random variables with a pdf supported in $[-\varepsilon, \varepsilon]$. Assume that for all t, and for some $\alpha, \beta > 0$,*

$$\alpha \leq |u(t)| \leq \beta. \tag{4.25}$$

Then, $0 \leq r_{\widetilde{x}_1\widetilde{x}_N} \leq 1$.

Proof. From (4.11) and (4.25), observe that $\widetilde{x}_1$ and $\widetilde{x}_N$ are bounded random variables:

$$-2\varepsilon/\alpha \leq \widetilde{x}_1 \leq 0, \qquad 0 \leq \widetilde{x}_N \leq 2\varepsilon/\alpha.$$

Thus, from Lemma 4.2.3, the numerator of $r_{\widetilde{x}_1\widetilde{x}_N}$ can be written as

$$\begin{aligned}\mathbf{E}(\widetilde{x}_1\widetilde{x}_N) - \mathbf{E}(\widetilde{x}_1)\mathbf{E}(\widetilde{x}_N) &= \int_0^{2\varepsilon/\alpha}\int_{-2\varepsilon/\alpha}^{0} [\text{Prob}(\widetilde{x}_1 \leq \xi)\,\text{Prob}(\widetilde{x}_N > \zeta) \\ &\qquad -\text{Prob}(\widetilde{x}_1 \leq \xi, \widetilde{x}_N > \zeta)]\; d\xi d\zeta.\end{aligned} \tag{4.26}$$

Next, from Lemma 4.2.2,

$$\begin{aligned}\text{Prob}(\widetilde{x}_1 \leq \xi)\,\text{Prob}(\widetilde{x}_N > \zeta) &= \prod_{t=1}^{N}[1 - F_\eta(-\varepsilon - \xi|u(t)|) - F_\eta(-\varepsilon + \zeta|u(t)|) \\ &\qquad + F_\eta(-\varepsilon - \xi|u(t)|)F_\eta(-\varepsilon + \zeta|u(t)|)] \\ &\geq \text{Prob}(\widetilde{x}_1 \leq \xi, \widetilde{x}_N > \zeta).\end{aligned} \tag{4.27}$$

Therefore, the integrand in (4.26) is nonnegative. Hence, the conclusion follows.

■

The random variable $\widetilde{x}_1$ defined by (4.11) can be written as

$$\widetilde{x}_1 = -\min_{u(t)\neq 0} \frac{-\widetilde{\eta}(t) + \varepsilon}{|u(t)|}.$$

Due to the symmetry assumption, $f_{-\widetilde{\eta}}(s) = f_\eta(s)$ for all $|s| \leq \varepsilon$. Thus,

$$\mathbf{E}(\widetilde{x}_1) = -\mathbf{E}(\widetilde{x}_N).$$

Therefore, $\widehat{\theta}_c$ is an unbiased estimator of θ for all N. Moreover, for all N

$$\mathbf{E}(\widetilde{x}_1^2) = \mathbf{E}(\widetilde{x}_N^2).$$

Hence, from Proposition 4.2.1, we have the following important inequalities

$$(1/2)\,\sigma^2_{\widetilde{x}_N} \le \sigma^2_{\widehat{\theta}_c} \le \sigma^2_{\widetilde{x}_N}. \tag{4.28}$$

Note that the variance of $R(S_N(y,u,\varepsilon))$ can be written as

$$\sigma^2_{R(S_N)} = \sigma^2_{\widehat{\theta}_c} - \sigma^2_{\widetilde{x}_N}\, r_{\widetilde{x}_1\widetilde{x}_N}. \tag{4.29}$$

In general, exact computation of $\sigma^2_{\widehat{\theta}_c}$ is difficult since it involves the correlation between $\widetilde{x}_1$ and $\widetilde{x}_N$. On the other hand, exactly computing or tightly bounding $\sigma^2_{\widetilde{x}_N}$ is easier. Thus, given a pdf for $\eta(t)$, from (4.28) we can determine how fast the variance of the central algorithm decays to zero as the number of data increases to infinity. This will be illustrated for the following pdf:

$$f_{\eta^*}(x) = \frac{p+1}{2\varepsilon}\left(1 - \frac{|x|}{\varepsilon}\right)^p, \quad |x| \le \varepsilon, \qquad p \ge 0. \tag{4.30}$$

Figure 4.1 shows the pdf in (4.30) for different values of p = 0, 0.4, 1, 2, and $\varepsilon = 0.5$.

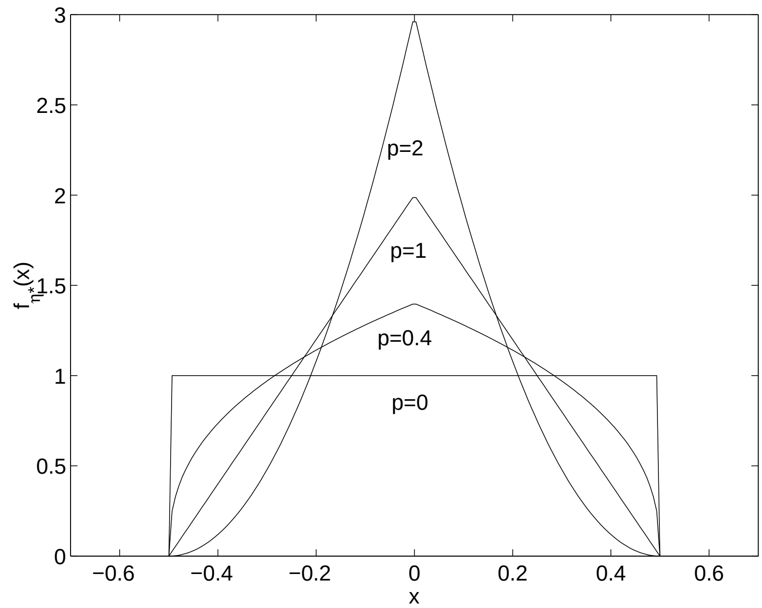

Figure 4.1: The pdf in (4.30) with $p = 0$, 0.4, 1, 2 and $\varepsilon = 0.5$

Recall that the gamma function, denoted by $\Gamma(z)$, is defined by [55]:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt.$$

Alternatively,

$$\frac{1}{\Gamma(z)} = ze^{\gamma z}\prod_{n=1}^{\infty}\left[\left(1+\frac{z}{n}\right)e^{-z/n}\right], \tag{4.31}$$

where γ is the Euler's constant

$$\gamma = \lim_{n\to\infty}\left[1+\frac{1}{2}+\frac{1}{3}+\cdots\frac{1}{n}-\ln n\right] = 0.57721566\cdots.$$

Lemma 4.2.4 *Let $\widetilde{x}_N$ be as in (4.11). Assume that the disturbances $\eta(t)$ are independent and have the pdf in (4.30). Let $u(t)$ be as in (4.25) with $\beta < C_p\,\alpha$, where*

$$C_p = \Gamma^{1/2}\left(1+\frac{2}{p+1}\right)\Gamma^{-1}\left(1+\frac{1}{p+1}\right). \tag{4.32}$$

Then,

$$\liminf_{N\to\infty}\sigma^2_{\widetilde{x}_N}\left(\frac{2}{N}\right)^{-2/(p+1)} \geq \frac{\varepsilon^2}{\beta^2}\Gamma^2\left(1+\frac{1}{p+1}\right)\left[C_p^2-\left(\frac{\beta}{\alpha}\right)^2\right] \tag{4.33}$$

$$\limsup_{N\to\infty}\mathbf{E}(\widetilde{x}_N^2)\left(\frac{2}{N}\right)^{-2/(p+1)} \leq \frac{\varepsilon^2}{\alpha^2}\Gamma\left(1+\frac{2}{p+1}\right). \tag{4.34}$$

Proof. From Lemma 4.2.3, (4.13), and (4.25),

$$\begin{aligned}
\mathbf{E}(\widetilde{x}_N) &= \int_0^{2\varepsilon/\alpha}\text{Prob}(\widetilde{x}_N > x)dx \\
&\leq \int_0^{2\varepsilon/\alpha}[1-F_{\eta^*}(-\varepsilon+\alpha x)]^N dx \\
&= \int_0^N\left(1-\frac{s}{N}\right)^N\left(\frac{2}{N}\right)^{1/(p+1)}\frac{\varepsilon}{\alpha(p+1)}s^{-p/(p+1)}ds \\
&= \left(\frac{2}{N}\right)^{1/(p+1)}\frac{\varepsilon}{\alpha(p+1)}\gamma_1(N)
\end{aligned}$$

where

$$\gamma_1(N) = \int_0^N\left(1-\frac{s}{N}\right)^N s^{-p/(p+1)}ds.$$

Similarly,

$$\begin{aligned}
\mathbf{E}(\widetilde{x}_N^2) &= 2\int_0^{2\varepsilon/\alpha}\text{Prob}(\widetilde{x}_N > x)xdx \\
&\geq 2\int_0^{2\varepsilon/\alpha}[1-F_{\eta^*}(-\varepsilon+\beta x)]^N xdx \\
&= 2\int_0^N\left(1-\frac{s}{N}\right)^N\left(\frac{2}{N}\right)^{2/(p+1)}\frac{\varepsilon^2}{\beta^2(p+1)}s^{-(p-1)/(p+1)}ds \\
&= 2\left(\frac{2}{N}\right)^{2/(p+1)}\frac{\varepsilon^2}{\beta^2(p+1)}\gamma_2(N)
\end{aligned}$$

where

$$\gamma_2(N) = \int_0^N \left(1 - \frac{s}{N}\right)^N s^{-(p-1)/(p+1)} ds. \qquad (4.35)$$

Thus,

$$\begin{aligned} \sigma^2_{\tilde{x}_N} &= \mathbf{E}(\tilde{x}^2_N) - \mathbf{E}^2(\tilde{x}_N) \\ &\geq \left(\frac{2}{N}\right)^{2/(p+1)} \frac{\varepsilon^2}{\beta^2} \left[\frac{2}{p+1}\gamma_2(N) - \left(\frac{\beta}{\alpha}\right)^2 \frac{1}{(p+1)^2}\gamma_1^2(N)\right]. \qquad (4.36) \end{aligned}$$

We define the indicator function of a set A, denoted by $\chi_A(s)$, as

$$\chi_A(s) = \begin{cases} 1, & \text{if } s \in A \\ 0, & \text{otherwise.} \end{cases}$$

From the dominated convergence theorem, note that

$$\begin{aligned} \lim_{N\to\infty} \gamma_1(N) &= \lim_{N\to\infty} \int_0^\infty \chi_{[0,N]}(s) \left(1 - \frac{s}{N}\right)^N s^{-p/(p+1)} ds \\ &= \int_0^\infty \lim_{N\to\infty} \left\{\chi_{[0,N]}(s) \left(1 - \frac{s}{N}\right)^N\right\} s^{-p/(p+1)} ds \\ &= \int_0^\infty e^{-s} s^{-p/(p+1)} ds \\ &= \Gamma\left(\frac{1}{p+1}\right). \end{aligned}$$

Likewise,

$$\begin{aligned} \lim_{N\to\infty} \gamma_2(N) &= \int_0^\infty e^{-s} s^{-(p-1)/(p+1)} ds \\ &= \Gamma\left(\frac{2}{p+1}\right). \qquad (4.37) \end{aligned}$$

Then,

$$\begin{aligned} &\lim_{N\to\infty} \left[\frac{2}{p+1}\gamma_2(N) - \left(\frac{\beta}{\alpha}\right)^2 \frac{1}{(p+1)^2}\gamma_1^2(N)\right] = \\ &= \frac{2}{p+1}\Gamma\left(\frac{2}{p+1}\right) - \left(\frac{\beta}{\alpha}\right)^2 \frac{1}{(p+1)^2}\Gamma^2\left(\frac{1}{p+1}\right) \\ &= \Gamma\left(1 + \frac{2}{p+1}\right) - \left(\frac{\beta}{\alpha}\right)^2 \Gamma^2\left(1 + \frac{1}{p+1}\right), \qquad (4.38) \end{aligned}$$

where the last equality above follows from the recurrence formula [55]:

$$\Gamma(z+1) = z\Gamma(z).$$

From (4.31), observe that

$$\left(\frac{1}{z\Gamma(z)}\right)^2 2z\Gamma(2z) = \frac{\prod_{n=1}^\infty \left(1 + \frac{z}{n}\right)^2}{\prod_{n=1}^\infty \left(1 + \frac{2z}{n}\right)} \geq 1.$$

With $z = \frac{1}{p+1}$, we have

$$C_p^2 \geq 1,$$

and hence, $C_p \geq 1$ for all $p \geq 0$. This can also be seen from the Figure 4.2. Thus for all sufficiently large N, (4.33) follows from (4.36) and (4.38).

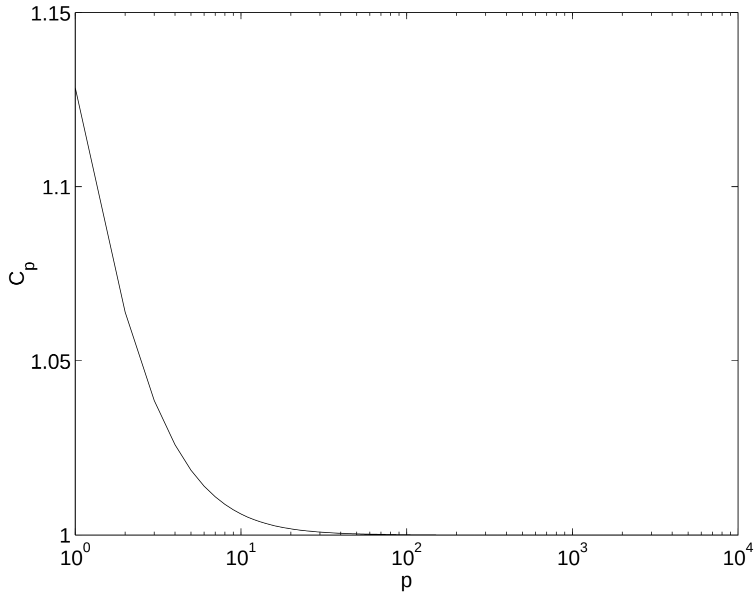

Figure 4.2: Plot of C_p in (4.32) as a function of p

An upper bound on $\sigma^2_{\tilde{x}_N}$ is obtained as follows. From Lemma 4.2.3, (4.13), and (4.25),

$$\begin{aligned}
\mathbf{E}(\tilde{x}_N^2) &= 2\int_0^{2\varepsilon/\alpha} \text{Prob}(\tilde{x}_N > x)xdx \\
&\leq 2\int_0^{2\varepsilon/\alpha} [1 - F_{\eta^*}(-\varepsilon + \alpha x)]^N xdx \\
&= 2\int_0^N \left(1 - \frac{s}{N}\right)^N \left(\frac{2}{N}\right)^{2/(p+1)} \frac{\varepsilon^2}{\alpha^2(p+1)} s^{-(p-1)/(p+1)} ds \\
&= 2\left(\frac{2}{N}\right)^{2/(p+1)} \frac{\varepsilon^2}{\alpha^2(p+1)} \gamma_2(N),
\end{aligned}$$

where $\gamma_2(N)$ is as in (4.35). Hence, for all sufficiently large N, (4.34) follows from (4.37) and the recurrence formula.

■

Combining (4.28) with Lemma 4.2.4, we obtain the following result.

Theorem 4.2.1 *Consider the Chebyshev center $\widehat{\theta}_c$ for the model (4.1). Assume that the disturbances $\eta(t)$ are independent and have the pdf in (4.30). Let $u(t)$ be as in (4.25) with $\beta < C_p\,\alpha$, where C_p is defined by (4.32). Then, there exist positive constants c_1 and c_2*

$$c_1 \leq \liminf_{N\to\infty} \sigma^2_{\widehat{\theta}_c} N^{2/(p+1)} \leq \limsup_{N\to\infty} \sigma^2_{\widehat{\theta}_c} N^{2/(p+1)} \leq c_2. \tag{4.39}$$

The above results can be summarized as follows.

- The Chebyshev center is the midrange of the extreme order statistics x_1 and x_N, which are the extreme points of $S_N(y, u, \varepsilon)$:

$$\widehat{\theta}_c = (1/2)x_1 + (1/2)x_N.$$

- Provided that the disturbances are independent identically and symmetrically distributed random variables, and the regressor signal and its inverse are magnitude bounded, the variance of the Chebyshev center is bounded below by

$$\sigma^2_{\widehat{\theta}_c} \geq (1/4)\,(\sigma^2_{x_1} + \sigma^2_{x_N}).$$

- If, in addition, the pdf of $\eta(t)$ is chosen as in (4.30), then the variance of the Chebyshev center converges to zero at the rate $O(N^{-2/(p+1)})$.

4.2.1 Correlation analysis of the extreme order statistics

As a final result in this section, let us show that the gap between the upper and lower bounds in (4.28) closes asymptotically when $\eta(t)$ have the pdf in (4.30). To this end, we have the following result.

Proposition 4.2.2 *Let $\widetilde{x}_1$ and $\widetilde{x}_N$ be as in (4.11). Assume that the disturbances $\eta(t)$ are independent and have the pdf in (4.30). Let $u(t)$ be as in (4.25). Then,*

$$\mathbf{E}(\widetilde{x}_1\widetilde{x}_N) - \mathbf{E}(\widetilde{x}_1)\mathbf{E}(\widetilde{x}_N) = O\left(\frac{(\ln N)^2}{N}\right). \tag{4.40}$$

Proof. Let $\mathcal{D}_N$ be the set of all pairs (ξ, ζ) satisfying $\xi \leq 0$, $\zeta \geq 0$,

$$\text{Prob}(\widetilde{x}_1 \leq \xi) > 1/N, \qquad \text{and} \qquad \text{Prob}(\widetilde{x}_N > \zeta) > 1/N.$$

From (4.12),

$$\text{Prob}(\widetilde{x}_1 \leq \xi) \leq [1 - F_{\eta^*}(-\varepsilon - \xi\alpha)]^N.$$

Thus,

$$\ln[1 - F_{\eta^*}(-\varepsilon - \xi\alpha)] > -\frac{\ln N}{N}. \tag{4.41}$$

Hence, from (4.41) and the inequalities [55]

$$\frac{x}{1+x} < \ln(1+x) < x \qquad (x > -1,\ x \neq 0),$$

we have

$$F_{\eta^*}(-\varepsilon - \xi\alpha) < \frac{\ln N}{N}. \tag{4.42}$$

Likewise,

$$F_{\eta^*}(-\varepsilon + \zeta\alpha) < \frac{\ln N}{N}. \tag{4.43}$$

From (4.30) and (4.42),

$$F_{\eta^*}(-\varepsilon - \xi\beta) < (\beta/\alpha)^{p+1} \frac{\ln N}{N}. \tag{4.44}$$

Similarly,

$$F_{\eta^*}(-\varepsilon + \zeta\beta) < (\beta/\alpha)^{p+1} \frac{\ln N}{N}. \tag{4.45}$$

Next, from Lemma 4.2.2

$$\text{Prob}(\widetilde{x}_1 \leq \xi, \widetilde{x}_N > \zeta) = \text{Prob}(\widetilde{x}_1 \leq \xi)\, \text{Prob}(\widetilde{x}_N > \zeta)\, \Delta_N(\xi, \zeta)$$

where

$$\Delta_N(\xi,\zeta) = \prod_{t=1}^{N} (1 - \delta_t(\xi,\zeta))$$

and

$$\delta_t(\xi,\zeta) = \frac{F_{\eta^*}(-\varepsilon - \xi|u(t)|)}{1 - F_{\eta^*}(-\varepsilon - \xi|u(t)|)}\, \frac{F_{\eta^*}(-\varepsilon + \zeta|u(t)|)}{1 - F_{\eta^*}(-\varepsilon + \zeta|u(t)|)}.$$

If

$$(\beta/\alpha)^{p+1} \frac{\ln N}{N} \leq \frac{1}{2},$$

then from (4.25), we have

$$\begin{aligned}
\delta_t(\xi,\zeta) &\leq 4F_{\eta^*}(-\varepsilon - \xi|u(t)|) F_{\eta^*}(-\varepsilon + \zeta|u(t)|) \\
&\leq 4F_{\eta^*}(-\varepsilon - \xi\beta) F_{\eta^*}(-\varepsilon + \zeta\beta) \\
&= \bar{\delta}(\xi,\zeta).
\end{aligned}$$

Thus,

$$\Delta_N(\xi,\zeta) \geq [1-\bar{\delta}(\xi,\zeta)]^N. \tag{4.46}$$

From (4.44) and (4.45),

$$\bar{\delta}(\xi,\zeta) < 4(\beta/\alpha)^{2(p+1)}\left(\frac{\ln N}{N}\right)^2.$$

Hence, from (4.46) and the inequalities

$$e^{-x/(1-x)} < 1-x < e^{-x} \qquad (0<x<1),$$

we have

$$\begin{aligned}\Delta_N(\xi,\zeta) &> \left[1-4(\beta/\alpha)^{2(p+1)}\left(\frac{\ln N}{N}\right)^2\right]^N \\ &> \exp\left(-\frac{4(\beta/\alpha)^{2(p+1)}\frac{(\ln N)^2}{N}}{1-4(\beta/\alpha)^{2(p+1)}\left(\frac{\ln N}{N}\right)^2}\right) \\ &> \exp\left(-8(\beta/\alpha)^{2(p+1)}\frac{(\ln N)^2}{N}\right). \end{aligned} \tag{4.47}$$

Let $\mathcal{Q}=\{(\xi,\zeta)\in\mathbb{R}^2 : -2\varepsilon/\alpha \leq \xi \leq 0,\ 0\leq\zeta\leq 2\varepsilon/\alpha\}$. Then,

$$\mathbf{E}(\widetilde{x}_1\widetilde{x}_N)-\mathbf{E}(\widetilde{x}_1)\mathbf{E}(\widetilde{x}_N) = \int\int_{\mathcal{Q}} \text{Prob}(\widetilde{x}_1\leq\xi)\,\text{Prob}(\widetilde{x}_N>\zeta)[1-\Delta_N(\xi,\zeta)]\,d\xi d\zeta. \tag{4.48}$$

Moreover, from (4.47)

$$\begin{aligned} 0 &\leq \int\int_{\mathcal{D}_N} \text{Prob}(\widetilde{x}_1\leq\xi)\,\text{Prob}(\widetilde{x}_N>\zeta)[1-\Delta_N(\xi,\zeta)]\,d\xi d\zeta \\ &\leq 32(\varepsilon/\alpha)^2(\beta/\alpha)^{2(p+1)}\frac{(\ln N)^2}{N} \end{aligned}$$

and

$$0 \leq \int\int_{\mathcal{Q}-\mathcal{D}_N} \text{Prob}(\widetilde{x}_1\leq\xi)\,\text{Prob}(\widetilde{x}_N>\zeta)[1-\Delta_N(\xi,\zeta)]\,d\xi d\zeta \leq \frac{4\varepsilon^2}{\alpha^2 N}.$$

From the above inequalities and (4.48), we have

$$\mathbf{E}(\widetilde{x}_1\widetilde{x}_N)-\mathbf{E}(\widetilde{x}_1)\mathbf{E}(\widetilde{x}_N) \leq 32(\varepsilon/\alpha)^2(\beta/\alpha)^{2(p+1)}\frac{(\ln N)^2}{N}+\frac{4\varepsilon^2}{\alpha^2 N} \tag{4.49}$$

which is the desired result. ■

When $u(t)$ is constant, it is a well-known fact that $\widetilde{x}_1$ and $\widetilde{x}_N$ are asymptotically uncorrelated [51]. The essence of Proposition 4.2.2 is the upper bound estimate on the right hand side of (4.40), which turns out to be quite important in the asymptotic analysis of $r_{\widetilde{x}_1\widetilde{x}_N}$.

Corollary 4.2.1 *Let* $\widetilde{x}_1$, $\widetilde{x}_N$, *and* $r_{\widetilde{x}_1\widetilde{x}_N}$ *be as in (4.11) and (4.23), respectively. Assume that the disturbances* $\eta(t)$ *are independent and have the pdf in (4.30). Let* $u(t)$ *be as in (4.25). Then,*

$$\lim_{N\to\infty} r_{\widetilde{x}_1\widetilde{x}_N} = 0.$$

Our final result in this section is captured in the following.

Theorem 4.2.2 *Consider the Chebyshev center* $\widehat{\theta}_c$ *of the membership-set* $S_N(y, u, \varepsilon)$ *for the model (4.1). Assume that the disturbances* $\eta(t)$ *are independent and have the pdf in (4.30). Let* $u(t)$ *be as in (4.25). Then,* $S_N(y, u, \varepsilon)$ *is the convex hull of the extreme points* x_1 *and* x_N *defined in (4.7) and*

$$\lim_{N\to\infty} \frac{\sigma^2_{\widehat{\theta}_c}}{\sigma^2_{x_1} + \sigma^2_{x_N}} = \lim_{N\to\infty} \frac{\sigma^2_{R(S_N)}}{\sigma^2_{x_1} + \sigma^2_{x_N}} = \frac{1}{4}.$$

The last result tells us that asymptotically, the variance of the Chebyshev center is as large as the variance of the radius of the membership-set. Whether the conclusions of Theorem 4.2.1 and Theorem 4.2.2 hold for multi-dimensional parameter case remain open problems for the future work.

4.3 Statistical Analysis of the Minimax Algorithm

In this section, we study the statistical properties of the minimax estimate for the model (4.1) assuming that the regressor and its inverse are magnitude bounded. In the general case, the minimax estimate is quite difficult to analyze since it is impossible to give a closed-form expression that is valid even asymptotically. Recall that $\widehat{\theta}_m$ coincides with $\widehat{\theta}_c$ when the regressor is constant. The magnitude constraint on $u(t)$ controls the distance between these estimators. More precisely, we have the following continuity type result.

Lemma 4.3.1 *Consider the Chebyshev center* $\widehat{\theta}_c$ *and the minimax estimator* $\widehat{\theta}_m$ *for the model (4.1). Let* $R(S_N(y, u, \varepsilon))$ *and* $u(t)$ *be as in (4.3) and (4.25), respectively. Then,*

$$|\widehat{\theta}_m - \widehat{\theta}_c| \leq (1/2)[(\beta/\alpha) - 1]\, R(S_N(y, u, \varepsilon)). \qquad (4.50)$$

Proof. The minimax estimator in (4.6) is the solution of the linear programming problem:

$$\begin{cases} \min & e \\ \text{subject to} & S_N(y, u, e) \text{ nonempty} \end{cases} \qquad (4.51)$$

Let e_∞ denote the optimal value of e in (4.51). Observe that $S_N(y,u,e_\infty)$ is a singleton since otherwise it is a closed interval and e_∞ can be reduced further. Thus, from the inequalities in (4.8), we have for all $u(t) \neq 0$,

$$\frac{y(t)\,\mathrm{sgn}[u(t)] - \varepsilon}{|u(t)|} + \frac{\varepsilon - e_\infty}{|u(t)|} \leq \widehat{\theta}_m \leq \frac{y(t)\,\mathrm{sgn}[u(t)] + \varepsilon}{|u(t)|} - \frac{\varepsilon - e_\infty}{|u(t)|}$$

which implies

$$x_1 + \frac{\varepsilon - e_\infty}{\beta} \leq \widehat{\theta}_m \leq x_N - \frac{\varepsilon - e_\infty}{\beta}.$$

It follows that

$$\varepsilon - e_\infty \leq \beta R(S_N(y,u,\varepsilon)). \tag{4.52}$$

Since $S_N(y,u,e_\infty)$ contains only $\widehat{\theta}_m$, from Lemma 4.2.1 we have

$$\begin{aligned}\widehat{\theta}_m &= \frac{1}{2}\left[\max_{u(t)\neq 0} \frac{y(t)\,\mathrm{sgn}[u(t)] - e_\infty}{|u(t)|} + \min_{u(t)\neq 0} \frac{y(t)\,\mathrm{sgn}[u(t)] + e_\infty}{|u(t)|}\right] \\ &\leq \widehat{\theta}_c - \frac{\varepsilon - e_\infty}{2\beta} + \frac{\varepsilon - e_\infty}{2\alpha}.\end{aligned} \tag{4.53}$$

Likewise,

$$\widehat{\theta}_m \geq \widehat{\theta}_c - \frac{\varepsilon - e_\infty}{2\alpha} + \frac{\varepsilon - e_\infty}{2\beta}. \tag{4.54}$$

Thus, from (4.53) and (4.54) we get

$$|\widehat{\theta}_m - \widehat{\theta}_c| \leq (2\beta)^{-1}[(\beta/\alpha) - 1]\,(\varepsilon - e_\infty). \tag{4.55}$$

The inequalities (4.52) and (4.55) complete the proof. ∎

If ε is a tight bound (see Definition 3.3.1) on $\eta(t)$, it is known [32, 33, 56, 37] that for all $\delta > 0$,

$$\lim_{N\to\infty} \mathrm{Prob}(R(S_N(y,u,\varepsilon)) > \delta) = 0$$

which implies that $\widehat{\theta}_m$ is a consistent estimator of θ since $\widehat{\theta}_m \in S_N(y,u,\varepsilon)$ for all N.

Confidence intervals for $\widehat{\theta}_m$ can be computed either by using the general results in [37] derived for all interpolatory estimators or resorting to the Chebyshev's inequality assuming that the variance of $\widehat{\theta}_m$ denoted by $\sigma^2_{\widehat{\theta}_m}$ is available. To this end, we have the following result.

Lemma 4.3.2 *Consider the minimax estimator $\widehat{\theta}_m$ for the model (4.1). Let $\widehat{\theta}_c$, $\widetilde{x}_N$, and $u(t)$ be as in (4.5), (4.11), and (4.25), respectively. Suppose that $\eta(t)$*

is a sequence of independent identically and symmetrically distributed random variables with a pdf supported in $[-\varepsilon, \varepsilon]$. *Then,*

$$\begin{gathered}(1/2)\sigma^2_{\widehat{\theta}_c} - (1/4)[(\beta/\alpha) - 1]^2 \{\sigma^2_{\widehat{\theta}_c} + 2\mathbf{E}(\widetilde{x}^2_N)\} \leq \sigma^2_{\widehat{\theta}_m} \\ \sigma^2_{\widehat{\theta}_m} \leq 8\sigma^2_{\widehat{\theta}_c} + 2[(\beta/\alpha) - 1]^2\{\sigma^2_{\widehat{\theta}_c} + \mathbf{E}(\widetilde{x}^2_N)\}.\end{gathered} \tag{4.56}$$

Proof. Since the variance of a random variable is invariant to translations, we have

$$\sigma^2_{\widehat{\theta}_m} = \mathbf{E}([\widehat{\theta}_m - \theta]^2) - \mathbf{E}^2(\widehat{\theta}_m - \theta). \tag{4.57}$$

By Lemma 4.3.1 and the symmetry assumption on the pdf of $\eta(t)$ note that

$$\begin{aligned}|\mathbf{E}(\widehat{\theta}_m - \theta)| &= |\mathbf{E}(\widehat{\theta}_m - \widehat{\theta}_c)| + |\mathbf{E}(\widehat{\theta}_c - \theta)| \\ &= |\mathbf{E}(\widehat{\theta}_m - \widehat{\theta}_c)| \\ &\leq (1/2)[(\beta/\alpha) - 1]\, \mathbf{E}(R(S_N(y, u, \varepsilon))) \\ &= (1/2)[(\beta/\alpha) - 1]\, \mathbf{E}(\widetilde{x}_N)\end{aligned} \tag{4.58}$$

since $\mathbf{E}(R(S_N(y, u, \varepsilon))) = \mathbf{E}(\widetilde{x}_N)$ and $\mathbf{E}(\widehat{\theta}_c) = \theta$. Again by Lemma 4.3.1, whenever $\widehat{\theta}_c - \theta > \delta$ holds, the inequality $\widehat{\theta}_m - \theta > \delta - (1/2)[(\beta/\alpha) - 1]R(S_N(y, u, \varepsilon))$ holds. Thus,

$$\text{Prob}(\widehat{\theta}_m - \theta + (1/2)[(\beta/\alpha) - 1]\, R(S_N(y, u, \varepsilon)) > \delta) \geq \text{Prob}(\widehat{\theta}_c - \theta > \delta).$$

Multiplying with 2δ and integrating both sides of this inequality with respect to δ, we get from (4.16)

$$\mathbf{E}(\{\widehat{\theta}_m - \theta + (1/2)[(\beta/\alpha) - 1]\, R(S_N(y, u, \varepsilon))\}^2) \geq \mathbf{E}([\widehat{\theta}_c - \theta]^2) = \sigma^2_{\widehat{\theta}_c}. \tag{4.59}$$

Since for any two random variables z_1 and z_2, the inequality $2z_1^2 + 2z_2^2 \geq (z_1 + z_2)^2$ implies

$$\mathbf{E}((z_1 + z_2)^2) \leq 2\mathbf{E}(z_1^2) + 2\mathbf{E}(z_2^2),$$

we have from (4.59)

$$\mathbf{E}([\widehat{\theta}_m - \theta]^2) \geq (1/2)\sigma^2_{\widehat{\theta}_c} - (1/4)[(\beta/\alpha) - 1]^2\, \mathbf{E}(R^2(S_N(y, u, \varepsilon))). \tag{4.60}$$

It follows from (4.57), (4.58), and (4.60)

$$\begin{aligned}\sigma^2_{\widehat{\theta}_m} &\geq (1/2)\sigma^2_{\widehat{\theta}_c} - (1/4)[(\beta/\alpha) - 1]^2 \{\mathbf{E}(R^2(S_N(y, u, \varepsilon))) + \mathbf{E}^2(\widetilde{x}_N)\} \\ &= (1/2)\sigma^2_{\widehat{\theta}_c} - (1/4)[(\beta/\alpha) - 1]^2 \{\sigma^2_{R(S_N)} + 2\mathbf{E}^2(\widetilde{x}_N)\} \\ &= (1/2)\sigma^2_{\widehat{\theta}_c} - (1/4)[(\beta/\alpha) - 1]^2 \{\sigma^2_{\widehat{\theta}_c} - \sigma^2_{\widetilde{x}_N} r_{\widetilde{x}_1\widetilde{x}_N} + 2\mathbf{E}^2(\widetilde{x}_N)\}\end{aligned}$$

where the last equality has followed from (4.29). Since $r_{\widetilde{x}_1\widetilde{x}_N} \geq 0$ and $\mathbf{E}^2(\widetilde{x}_N) \leq \mathbf{E}(\widetilde{x}_N^2)$, the above expression can further be simplified as follows

$$\sigma^2_{\widehat{\theta}_m} \geq (1/2)\sigma^2_{\widehat{\theta}_c} - (1/4)[(\beta/\alpha) - 1]^2 \{\sigma^2_{\widehat{\theta}_c} + 2\mathbf{E}(\widetilde{x}_N^2)\}. \tag{4.61}$$

For an upper bound on $\sigma^2_{\widehat{\theta}_m}$, by Lemma 4.3.1 we have

$$\widehat{\theta}_c - \theta > \delta - (1/2)[(\beta/\alpha) - 1]R(S_N(y, u, \varepsilon))$$

whenever $\widehat{\theta}_m - \theta > \delta$, and therefore,

$$\begin{aligned}
\text{Prob}(\widehat{\theta}_m - \theta > \delta) &\leq \text{Prob}(\widehat{\theta}_c - \theta + (1/2)[(\beta/\alpha) - 1]\, R(S_N(y, u, \varepsilon)) > \delta) \\
&\leq \text{Prob}(\widehat{\theta}_c - \theta > \delta/2) \\
&\quad + \text{Prob}([(\beta/\alpha) - 1]\, R(S_N(y, u, \varepsilon)) > \delta).
\end{aligned}$$

Multiplying with 2δ and integrating both sides of the above inequality, we get

$$\begin{aligned}
2\int_0^b \text{Prob}(\widehat{\theta}_m - \theta > \delta)\delta\, d\delta &\leq 2\int_0^b \text{Prob}(\widehat{\theta}_c - \theta > \delta/2)\delta\, d\delta \\
&\quad +2\int_0^b \text{Prob}(R(S_N(y, u, \varepsilon)) > \delta[(\beta/\alpha) - 1]^{-1})\delta\, d\delta
\end{aligned} \tag{4.62}$$

where b is an arbitrary positive number larger than the ranges of the random variables involved in the integrands. Thus, from (4.16) and (4.62),

$$\begin{aligned}
\mathbf{E}([\widehat{\theta}_m - \theta]^2) &\leq 8\mathbf{E}([\widehat{\theta}_c - \theta]^2) + 2[(\beta/\alpha) - 1]^2\, \mathbf{E}(R^2(S_N(y, u, \varepsilon))) \\
&\leq 8\sigma^2_{\widehat{\theta}_c} + 2[(\beta/\alpha) - 1]^2\{\sigma^2_{\widehat{\theta}_c} + \mathbf{E}(\widetilde{x}_N^2)\}.
\end{aligned} \tag{4.63}$$

The inequalities (4.61), (4.63), and the fact that $\sigma_x^2 \leq \mathbf{E}(x^2)$, complete the proof.

■

4.4 Simulation Example

In this section, we illustrate the results of this chapter by a simulation example. We consider the model (4.1) with $\theta = 1$ and $\varepsilon = 0.5$. As for the regressor signal, we choose a high-pass filtered sinusoid

$$u(t) = \begin{cases} |\sin(2\pi t/N)|, & \text{if } |\sin(2\pi t/N)| \geq \alpha \\ \alpha, & \text{otherwise.} \end{cases}$$

The noise is generated according to the pdf in (4.30) with $p = 1$ and $p = 10$; and the number of data points $N = 400$ is chosen. We define the relative errors of the central and the minimax algorithms in the order given as $[\widehat{\theta}_c - \theta]/R(S_N(y, u, \varepsilon))$ and $[\widehat{\theta}_m - \theta]/R(S_N(y, u, \varepsilon))$. Figures 4.3 and 4.4 show the relative errors of the central and the minimax algorithms as a function of β/α for the noise pdf in (4.30) with $p = 1$ and $p = 10$, respectively. From the figures, the following observations can be made: 1) the central and the minimax estimates coincide when $\beta/\alpha = 1$; 2) for some ranges of β/α, the performances of the central and the minimax estimates depend on β/α; and 3) after some particular β/α value, the performances of the central and the minimax algorithms are not affected by β/α.

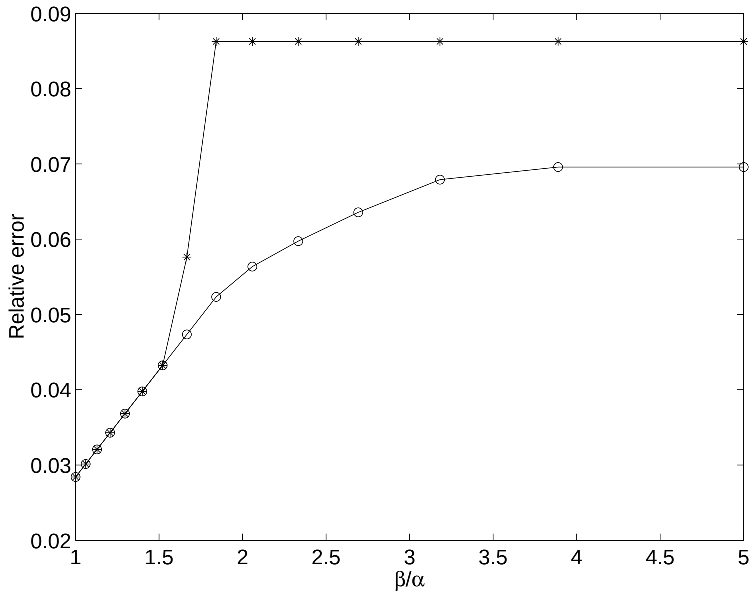

Figure 4.3: Relative errors of the central and the minimax algorithms as a function of β/α for the noise pdf in (4.30) with $p = 1$. $[\widehat{\theta}_c - \theta]/R(S_N(y, u, \varepsilon))$: $-*$; and $[\widehat{\theta}_m - \theta]/R(S_N(y, u, \varepsilon))$: $-\circ$

4.5 Summary

In this chapter, we first analyzed the statistical properties of the Chebyshev center assuming that the regressor and its inverse are magnitude bounded. This assumption led to the conclusion that the Chebyshev center is the midrange of two bounded extreme order statistics. The boundedness of these random variables facilitated integral representations for the moments of them in terms of rare event probabilities. These integrals were used to derive upper and lower bounds

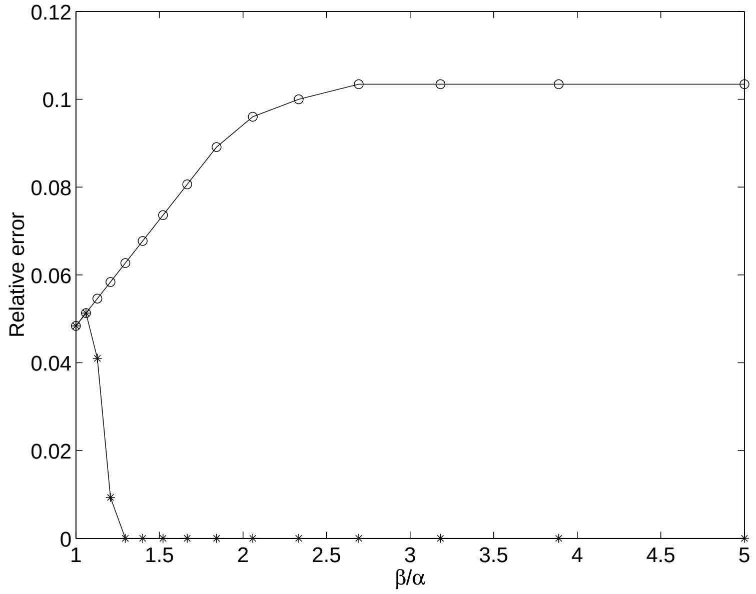

Figure 4.4: Relative errors of the central and the minimax algorithms as a function of β/α for the noise pdf in (4.30) with $p = 10$. $[\widehat{\theta}_c - \theta]/R(S_N(y, u, \varepsilon))$: $-*$; and $[\widehat{\theta}_m - \theta]/R(S_N(y, u, \varepsilon))$: $-\circ$

on the convergence rate of the variance of the central algorithm for a specific noise probability density function. We then showed that the results obtained for the central algorithm carry over to the minimax algorithm under the similar regressor constraints.

5. CONCLUDING REMARKS

5.1 Conclusions

In this thesis, statistical analysis of membership-set based estimators was studied in two different settings.

In Chapter 3, corresponding to the first case, we considered periodic input signals and the orthonormal regressors. We derived upper and lower probability bounds on the diameter of the membership-set. We then studied the central and the projection algorithms for the special case that the number of the unknown parameters equals the input period. We derived explicit formulae for the central algorithm and the diameter of the membership-set, and characterized the set of all projection algorithms.

In Chapter 4, corresponding to the second case, we studied the statistical properties of the central algorithm and the minimax algorithms in a one-dimensional parameter space setting assuming that the regressor signal and its inverse are magnitude bounded. We derived non-asymptotic, order-tight, upper and lower bounds on the convergence rate of the parameter estimate variance for the central and the minimax algorithms.

5.2 Recommendations for the Future Work

Although the parameter estimation problem considered in Chapter 4 is one-dimensional, the changing nature of the regressor signal makes the statistical analysis very difficult. For example, when the regressor magnitude is constant the central and the minimax algorithms coincide. The problem becomes even much harder when the regressor signal is vector-valued. We expect our results to provide insights in the statistical analysis of these membership-set based estimators when the unknown parameter is multi-dimensional and the probability distribution function of the noise is more general than the one considered in this thesis.

REFERENCES

[1] LJUNG, L. *System Identification: Theory for the User.* Prentice-Hall, 1987.

[2] GOODWIN, G. C. and PAYNE, R. L. *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press, 1977.

[3] SODERSTROM, T. and STOICA, P. *System Identification.* Prentice-Hall, 1989.

[4] SCHWEPPE, F. C. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Transactions on Automatic Control*, **13**:22–28, 1968.

[5] WITSENHAUSEN, H. S. Sets of possible states of linear systems given perturbed observations. *IEEE Transactions on Automatic Control*, **13**:556–558, 1968.

[6] BERTSEKAS, D. P. and RHODES, I. B. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, **16**:117–128, 1971.

[7] SCHLAEPFER, F. M. and SCHWEPPE, F. C. Continuous-time state estimation under disturbances bounded by convex sets. *IEEE Transactions on Automatic Control*, **17**:197–205, 1972.

[8] BARMISH, B. R. and SANKARAN, J. The propagation of parametric uncertainty via polytopes. *IEEE Transactions on Automatic Control*, **24**:346–349, 1979.

[9] FOGEL, E. System identification via membership set constraints with energy constrained noise. *IEEE Transactions on Automatic Control*, **24**:752–758, 1979.

[10] MILANESE, M. and BELFORTE, G. Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: linear families of models and estimators. *IEEE Transactions on Automatic Control*, **27**:408–414, 1982.

[11] MILANESE, M. and TEMPO, R. Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control*, **30**:730–738, 1985.

[12] MILANESE, M., TEMPO, R., and VICINO, A. Strongly optimal algorithms and optimal information in estimation problems. *Journal of Complexity*, **2**:78–94, 1986.

[13] KACEWICZ, B. Z., MILANESE, M., TEMPO, R., and VICINO, A. Optimality of central and projection algorithms for bounded uncertainty. *Systems and Control Letters*, **8**:161–171, 1986.

[14] TEMPO, R. and WASILKOWSKI, G. W. Maximum likelihood estimators and worst case optimal algorithms for system identification. *Systems and Control Letters*, **10**:265–270, 1988.

[15] TEMPO, R. Robust estimation and filtering in the presence of bounded noise. *IEEE Transactions on Automatic Control*, **33**:864–867, 1988.

[16] MILANESE, M. and VICINO, A. Optimal esimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, **27**:997–1009, 1991.

[17] MICCHELLI, C. A. and RIVLIN, T. J. *Optimal Estimation in Approximation Theory*, A survey of optimal recovery. Plenum Press, 1977.

[18] TRAUB,J. F. and WOZNIAKOWSKI, H. *A General Theory of Optimal Algorithms*. Academic Press, 1980.

[19] CRAMER, H. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[20] CLOWES, G. Choice of time-scaling factor for linear system approximations using orthonormal Laguerre functions. *IEEE Transactions on Automatic Control*, **10**:487–489, 1965.

[21] GU, G., KHARGONEKAR, P. P., and LEE, E. B. Approximation of infinite-dimensional systems. *IEEE Transactions on Automatic Control*, **34**:610–618, 1989.

[22] WAHLBERG, B. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, **36**:551–562, 1991.

[23] WAHLBERG, B. System identification using Kautz models. *IEEE Transactions on Automatic Control*, **39**:1276–1282, 1994.

[24] HEUBERGER, P. S. C., VAN DEN HOF, P. M. J., and BOSGRA, O. A generalized orthonormal basis for linear dynamical systems. *IEEE Transactions on Automatic Control*, **40**:451–465, 1995.

[25] VAN DEN HOF, P. M. J., HEUBERGER, P. S. C., and BOKOR, J. System identification with generalized orthonormal basis functions. *Automatica*, **31**:1821–1834, 1995.

[26] WARD, N. F. D. and PARTINGTON, J. R. Rational wavelet decompositions of transfer functions in Hardy-Sobolev classes. *Mathematics of Control, Signals, and Systems*, **8**:257–278, 1995.

[27] WARD, N. F. D. and PARTINGTON, J. R. Robust identification in the disc algebra using rational wavelets and orthonormal basis functions. *International Journal of Control*, **64**:409–423, 1996.

[28] NINNESS, B. and GUSTAFSSON, F. A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control*, **42**:515–521, 1997.

[29] AKÇAY, H. and NINNESS, B. Rational basis functions for robust identification from frequency and time-domain measurements. *Automatica*, **34**:1101–1117, 1998.

[30] NINNESS, B., HJALMARSSON, H., and GUSTAFSSON, F. The fundamental role of general orthonormal bases in system identification. *IEEE Transactions on Automatic Control*, **44**:1384–1406, 1999.

[31] VICINO, A. and MILANESE, M. Optimal inner bounds of feasible parameter set in linear estimation with bounded noise. *IEEE Transactions on Automatic Control*, **36**:759–763, 1991.

[32] VERES, S. M. and NORTON, J. P. Structure selection for bounded-parameter models: consistency conditions and selection criterion. *IEEE Transactions on Automatic Control*, **36**:474–481, 1991.

[33] BAI, E. W., TEMPO, R., and CHO, H. Membership set estimators: size, optimal inputs, complexity and relations with least squares. *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, **42**:266–277, 1995.

[34] BAI, E. W., YE, Y., and TEMPO, R. Bounded error parameter estimation: a sequential analytic center approach. *IEEE Transactions on Automatic Control*, **44**:1107–1117, 1999.

[35] BAI, E. W., FU, M., TEMPO, R., and YE, Y. Convergence results of the analytic center estimator. *IEEE Transactions on Automatic Control*, **45**:569–572, 2000.

[36] AKÇAY, H., HJALMARSSON, H., and LJUNG, L. On the choice of norms in system identification. *IEEE Transactions on Automatic Control*, **41**:1367–1372, 1996.

[37] AKÇAY, H. The size of the membership-set in a probabilistic framework. *Automatica*, **40**:253–260, 2004.

[38] AT, N. and AKÇAY, H. Central and projection algorithms in set membership identification with periodic inputs. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, pages 5188–5191, Paradise Island, Bahamas, 2004.

[39] AKÇAY, H. and AT, N. Membership identification with periodic inputs and orthonormal regressors. Submitted to *Signal Processing*.

[40] AT, N. and AKÇAY, H. Statistical analysis of the central algorithm. In *Proceedings of the 9th Canadian Workshop on Information Theory*, pages 300–303, Montreal, Canada, 2005.

[41] AKÇAY, H. and AT, N. Convergence analysis of central and minmax algorithms for models with one unknown parameter. Accepted to *Mathematics of Control, Signals, and Systems*.

[42] MAKILA, P. M. Robust identification and Galois sequences. *International Journal of Control*, **54**:1189–1200, 1991.

[43] DAHLEH, M. A., THEODOSOPOULOS, T. V., and TSITSIKLIS, J. N. The sample complexity of worst-case identification of FIR linear systems. *Systems and Control Letters*, **20**:157–166, 1993.

[44] POOLLA, K. and TIKKU, A. On the time complexity of worst-case system identification. *IEEE Transactions on Automatic Control*, **39**:944–950, 1994.

[45] ZAMES, G., LIN, L., and WANG, L. Y. Fast identification n-widths and uncertainty principles for LTI and slowly varying systems. *IEEE Transactions on Automatic Control*, **39**:1827–1838, 1994.

[46] WEYER, E., WILLIAMSON, R. C., and MAREELS, I. M. Y. Finite sample properties of linear model identification. *IEEE Transactions on Automatic Control*, **44**:1370–1383, 1999.

[47] CAMPI, M. C. and WEYER, E. Finite sample properties of system identification methods. *IEEE Transactions on Automatic Control*, **47**:1329–1334, 2002.

[48] MILANESE, M. and TARAGNA, M. H_∞ identification of "soft" uncertainty models. *Systems and Control Letters*, **37**:217–228, 1999.

[49] CHEN, J. and GU, G. *Control-Oriented System Identification: An H_∞ Approach*. John Wiley & Sons, 2000.

[50] PARTINGTON, J. R. Interpolation in normed spaces from the values of linear functionals. *Bulletin of London Mathematical Society*, **26**:165–170, 1994.

[51] DAVID, H. A. *Order Statistics*. John Wiley & Sons, 2nd edition, 1981.

[52] GRAY, R. M. Toeplitz and circulant matrices: a review. Technical report, Information Systems Laboratory, Stanford University, 1971.

[53] HJALMARSSON, H. Optimally robust system identification of systems subject to amplitude-bounded stochastic disturbances. *IEEE Transactions on Automatic Control*, **43**:947–953, 1998.

[54] CAMPI, M. C., OOI, S. K., and WEYER, E. Non-asymptotic quality assessment of generalised FIR models with periodic inputs. *Automatica*, **40**:2029–2041, 2004.

[55] ABRAMOWITZ, M. and STEGUN, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* United States Government Printing Office, 10th printing, 1972.

[56] BAI, E. W., CHO, H., and TEMPO, R. Convergence properties of the membership set. *Automatica*, 34:1245–1249, 1998.

[57] ARNOLD, B. C., BALAKRISHNAN, N., and NAGARAJA, H. N. *A First Course in Order Statistics.* John Wiley & Sons, 1992.

APPENDIX A ORDER STATISTICS

Let $X_1, X_2, ..., X_n$ denote a random sample from a continuous distribution with pdf $f(x)$. The order statistics is obtained by arranging the X_i's in nondecreasing order of magnitude so that $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$. In the following, we review the basic distribution theory of these ordered random variables and of functions involving them. The reader is referred to [51] and [57] for more detailed discussion.

Distribution of a Single Order Statistic

The pdf of the ith order statistic is denoted by $f_{i:n}(x)$, $1 \leq i \leq n$, and is defined as

$$\begin{aligned} f_{i:n}(x) &= \lim_{\delta x \to 0} \left\{ \tfrac{\text{Prob}(x < X_{i:n} \leq x + \delta x)}{\delta x} \right\} \\ &= \tfrac{n!}{(i-1)!(n-i)!} \left\{ F(x) \right\}^{i-1} \left\{ 1 - F(x) \right\}^{n-i} f(x), \quad -\infty < x < \infty. \end{aligned}$$

In particular, the pdfs of the smallest and largest order statistics are given by

$$f_{1:n}(x) = n \left\{ 1 - F(x) \right\}^{n-1} f(x), \quad -\infty < x < \infty,$$

and

$$f_{n:n}(x) = n \left\{ F(x) \right\}^{n-1} f(x), \quad -\infty < x < \infty.$$

Moreover, the expected value and the variance of the ith order statistics are given by using the standard definitions:

$$\mathbf{E}(X_{i:n}) = \int_{-\infty}^{\infty} x f_{i:n}(x) dx$$

$$\text{var}(X_{i:n}) = \int_{-\infty}^{\infty} \left(x - \mathbf{E}(X_{i:n}) \right)^2 f_{i:n}(x) dx.$$

Example Consider a uniform parent distribution in $[-\varepsilon, \varepsilon]$. The pdfs of the smallest and largest order statistics are given by

$$f_{1:n}(x) = \frac{n}{2\varepsilon} \left(\frac{1}{2} - \frac{1}{2\varepsilon} x \right)^{n-1}, \quad -\varepsilon \leq x \leq \varepsilon$$

$$f_{n:n}(x) = \frac{n}{2\varepsilon} \left(\frac{1}{2} + \frac{1}{2\varepsilon} x \right)^{n-1}, \quad -\varepsilon \leq x \leq \varepsilon.$$

Joint Distribution of Two Order Statistics

The joint pdf of $X_{i:n}$ and $X_{j:n}$, $1 \le i < j \le n$, is denoted by $f_{i,j:n}(x_i, x_j)$ and is defined as

$$\begin{aligned} f_{i,j:n}(x_i, x_j) &= \lim_{\delta x_i \to 0, \delta x_j \to 0} \left\{ \tfrac{\text{Prob}(x_i \le X_{i:n} \le x_i + \delta x_i, x_j \le X_{j:n} \le x_j + \delta x_j)}{\delta x_i \delta x_j} \right\} \\ &= \tfrac{n!}{(i-1)!(j-i-1)!(n-j)!} \\ &\quad \times \{F(x_i)\}^{i-1} \{F(x_j) - F(x_i)\}^{j-i-1} \{1 - F(x_j)\}^{n-j} f(x_i) f(x_j), \\ &\qquad -\infty < x_i < x_j < \infty. \end{aligned}$$

In particular, the joint pdf of the smallest and largest order statistics is given by

$$f_{1,n:n}(x_1, x_n) = n(n-1) \{F(x_n) - F(x_1)\}^{n-2} f(x_1) f(x_n), \quad -\infty < x_1 < x_n < \infty.$$

Example For a uniform parent distribution in $[-\varepsilon, \varepsilon]$, the joint pdf of the smallest and largest order statistics is given by

$$f_{1,n:n}(x_1, x_n) = \frac{n(n-1)}{(2\varepsilon)^n} (x_n - x_1)^{n-2}, \quad -\varepsilon \le x_1 < x_n \le \varepsilon.$$

Distribution of the Range and Midrange

From the joint pdf of two or more order statistics, we can derive the pdf of any well-behaved function of the order statistics using the standard transformation techniques. Two important ones include the range and midrange. The *range* is defined as

$$W = X_{n:n} - X_{1:n},$$

and its pdf is given by

$$f_W(w) = n(n-1) \int_{-\infty}^{\infty} \{F(w + x_1) - F(x_1)\}^{n-2} f(x_1) f(w + x_1) dx_1.$$

The *midrange* is defined as

$$M = \frac{X_{1:n} + X_{n:n}}{2},$$

and its pdf is given by

$$f_M(m) = 2n(n-1) \int_{-\infty}^{m} \{F(2m - x_1) - F(x_1)\}^{n-2} f(x_1) f(2m - x_1) dx_1.$$

APPENDIX B CONVERGENCE OF RANDOM VARIABLES

Some of the convergence concepts for a sequence of random variables:

- Almost sure convergence, $X_n \xrightarrow{a.s.} X$: A sequence of random variables $\{X_n\}$ converges almost surely, or converges with probability 1, to a random variable X if

$$\text{Prob}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

- Convergence in probability, $X_n \xrightarrow{p} X$: A sequence of random variables $\{X_n\}$ converges in probability to a random variable X if

$$\lim_{n\to\infty} \text{Prob}\left(|X_n - X| \geq \varepsilon\right) = 0$$

  for every $\varepsilon > 0$.

- Convergence in distribution, $X_n \xrightarrow{d} X$: A sequence of random variables $\{X_n\}$ with cdf $F_n(x)$ converges in distribution, or converges weakly, to a random variable X with cdf $F(x)$ if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

  for all continuity points of $F(x)$.

Note that almost sure convergence implies both convergence in probability and convergence in distribution. Likewise, convergence in probability implies the convergence in distribution.