The Pennsylvania State University

The Graduate School

College of Engineering

POLYNOMIAL OPTIMIZATION BASED APPROACHES TO SYSTEM DESIGN, ANALYSIS AND IDENTIFICATION

A Dissertation in
Electrical Engineering
by
Chao Feng

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2013

The dissertation of Chao Feng was reviewed and approved* by the following:

Constantino Lagoa
Professor of Electrical Engineering
Dissertation Adviser, Chair of Committee

Ji-Woong Lee
Assistant Professor of Electrical Engineering

David Miller
Professor of Electrical Engineering

Qian Wang
Professor of Mechanical Engineering

Victor Pasko
Professor of Electrical Engineering
Graduate Program Coordinator

*Signatures are on file in the Graduate School.

Abstract

In recent developments of system and control theory, a large effort has been devoted to finding equivalent convex formulation of the problems of interest. A successful example is the wide application of linear matrix inequalities (LMIs) in formulating system design and analysis problems. From a theoretic point of view, such problems can be considered *solved*, as convex optimization can be solved reliably and efficient using interior-point methods or other methods available in the literature and/or commercial software. On the other hand, however, many challenging problems in system and control theory have been proven to be NP-Complete or NP-hard. Therefore, unless proven P=NP, the best way to tackle these problems is to find *approximate* solutions using limited computational resources. Recent developments in polynomial optimization, which include moment-based approach and its dual sum-of-square method, shed some light on solving some of those challenging problems, as it provides systematic approaches to build *asymptotically convergent convex* relaxations to a general polynomial optimization problem. In this dissertation, we use this as the main optimization tool to address various important yet difficult problems in system and control theory. The problems addressed are categorized into four topics: 1) chance-constraint optimization, 2) distributional robustness, 3) hybrid system identification, and 4) generalized fixed order interpolation. The first two topics is closely related to the probabilistic framework developed in recent years. In the first topic, we design special polynomial functions and use them to develop deterministic approaches to address the probabilistic constraints. Comparing the scenario approach in the literature of probabilistic control, which give *soft* bounds on probability, our approaches provide *hard* bounds. The second topic is connected to system analysis with uncertainty under probabilistic framework, in a *distributional-free*

manner. Instead of assuming some fixed distribution on the uncertainty, it aims at finding the *worst-case* expected performance of the system, assuming the distribution of uncertainty is unknown but obey some loose conditions. The last two topics addressed concern hybrid system identification and generalized interpolation. We first show that these problems can be equivalently reformulated as polynomial optimization problems. While the recent developed polynomial optimization tools can construct convex relaxations to these problems, the required computational cost is prohibitively large. It is not surprising as a polynomial problem is NP-hard in general. In this dissertation, we exploit the very specific structure of these problems and provide numerically efficient algorithms to solve these problems.

Table of Contents

List of Figures

List of Tables

Acknowledgments

It has been my great honor and pleasure to work with my advisor, Prof. Constantino Lagoa since Fall 2006. He is an outstanding researcher, a great mentor and a considerate friend. During my PhD study, his advice helped me working in the right direction and he is always there with great ideas when I got stuck in research. I always enjoy the discussion with him (sometime with beer), from where many of my research ideas were inspired. I would like to express my sincerest thanks to him for all his support and guidance through my entire graduate life in US.

I would like to thank my committee members: Prof. Miller David, Prof. Ji-Woong Lee and Prof. Qian Wang, for their time to read my dissertation and giving me valuable suggestions.

I would also like to thank Prof. Mario Sznaier, Dr. Fabrizio Dabbene and Dr. Necmiye Ozay, from whom I benefit a lot in research with their discussion and interaction.

I would like to thank my previous lab-mates, Abdullah Ashoor, Wenjing Su, Wenjing Ma and Tao Ding, for their help and friendship.

Last but not the least, I would like to thank my parents for their unconditional love and support. I would also like to thank my wife for her encouragement and support in finishing this dissertation.

Chapter 1

Introduction

Classical system and control theory has been very successful in developing deterministic and analytic solutions to many important control problems. Classical examples include Linear Quadratic Gaussian (LQG) control and Kalman filtering. In recent developments, however, many efforts have been devoted to constructing computationally tractable optimization problems to (approximately) solve challenging problems in the field. One of the main reasons is that, while difficult to find analytic solutions, many challenging and important problems can be reformulated as convex optimization problems. And all convex optimization problems, at least from a theoretical point of view, can be solved reliably and efficiently using interior-point methods or other special methods for convex optimization [1]. In principle, they can be solved in polynomial time, and thus are generally treated as computationally tractable problems. In this sense, the reformulated problems as well as the original problems from system and control theory are considered as solved. One successful example is the wide application of linear matrix inequalities (LMIs) in formulating problems in system and control theory [2].

On the other hand, many challenging problems raised in system and control theory are so difficult that it is not possible to have equivalent convex optimization problems. In [3, 4], it is proven that a selection of interesting control problems are NP-Complete or NP-hard. Therefore, unless proven "p=NP" which is very unlikely (at least in the foreseeable future), our best bet to tackle these problems is to find "approximate" solutions using

limited computational resource.

A "good" approximation for a problem in system and control needs to satisfy 1) it is computationally tractable (in the context of this dissertation, it means convex); and 2) its solution is " very close" to the optimal solution. There are various measures to define how "close" is the solution of the approximation to the optimal solution. One common way to define the "closeness" is to find the gap between the values of the cost function for the two solutions. One of the main optimization tools used in this dissertation is approximating algorithms for polynomial optimization. These recently developed algorithms provide a systematic approach to construct asymptotically convergent approximations. These algorithms construct semi-definite programs (SDPs) approximations for polynomial optimization problems. The solution of these relaxations can be made arbitrary "close" to the optimal solution, in the sense the gap between the values of the cost function can be made arbitrary small.

It should be noted that polynomial optimization problems are in general non-convex and difficult. Indeed, they are NP-hard problems [5]. Thus, it is elegant to find that these problems can be relaxed as a hierarchy of converging convex optimization programs. In [6], the so-called Sum-of-Squares (SOS) technique is developed to construct such convex relaxations to find a positive polynomial in a semi-algebraic set. In [7], a moment based optimization approach, which is dual of a SOS, is proposed to find the global minimum of a constrained polynomial optimization problem. These approaches are very powerful because 1) they provide convex relaxations that are efficiently solvable; and 2) the relaxations are convergent, i.e., the gap of the cost function between the solutions of the relaxations and the global optimum can be arbitrarily small.

Since many of the challenging problems in system and control theory can be converted to polynomial optimization problems, it is not surprising that the aforementioned polynomial optimization approaches have attracted considerable attention among the researchers in system and control theory. For example, for a linear system affected by real parametric uncertainty, the robust stability verification can be recast as finding a number of matrix polynomials satisfying certain conditions [8], or equivalently, as seeking a polynomial param-

eter dependent Lyapunov function of given degree [9]. In nonlinear systems, a polynomial approach can be used to find a polynomial Lyapunov function satisfying certain conditions [10].

In this dissertation, effort have been put on exploiting applications of polynomial approaches to important and challenging problems in system and control theory. In particular, four topics are covered including chance-constraint optimization, distributional robustness, hybrid system identification, and generalized fixed order interpolation. The first two topics are closely related to the probabilistic framework developed in recent years in control. Contrary to the well known scenario approaches, in which Monte Carlo simulations are used, we propose several deterministic approaches without sampling to address probabilistic constraints in this dissertation. These approaches aim at developing efficient numerical algorithms for robust and chance-constraint optimization problems, with applications specifically to robust control design and worst-case system performance analysis for systems with polynomial dependence on uncertainties. One of the major contributions is that these approaches build interesting connections between the conventional robust control and chance-constraint control problems.

The last two topics are in the field of system identification and interpolation. Those are fairly mature research areas, yet still have important problems that have not been addressed in the literature. In this dissertation, numerically efficient algorithms are developed for several complex control problems including input-output switched system identification and generalized mixed-data interpolation. While it is shown these problems can be reformulated as polynomial optimization problems, it should be emphasized that solving their convex relaxations directly still suffers significant computational difficulty in general. This is a consequence of the fact that the relaxations have to be *tight* enough to have meaningful solutions, which means the relaxed optimization problem is of large size and, hence, very difficult to solve. This should not be surprising as a polynomial problem is NP-hard in general. However, one important contribution in the dissertation is to show that those control problems have very specific structure which can be used to significantly simplify the relaxed convex programs and hence, to greatly reduce the required computational cost.

The remainder of the introduction briefly reviews the background and the related results in the literature of the topics covered in this dissertation.

1.1 Robust Control, Probabilistic Control and Chance-constraint Optimization

Many robust control and analysis problems subject to parametric uncertainty can be recast as *robust convex optimization problems*; see, for instance, [2, 11]. When a simple uncertainty structure is considered, this deterministic approach has been proven to be very successful in solving such problems. For example, when the system coefficients are linear or affine functions of independent variables, extreme point results can be easily applied; e.g., see [12]. For more general linear fractional dependence on norm-bounded *unstructured* uncertainty, efficient methods based on robust linear matrix inequalities (LMI) have been successfully developed; e.g., see [13].

However, the problem is significantly harder to solve when it involves structured uncertainty. In fact, it has been proven that many linear robust control problems are NP-hard [3]. Thus, computable solutions can be in general obtained only at the expense of introducing conservatism. For example, in [13] the authors provide sufficient conditions for the case when the uncertainty enters the data in a linear fractional form, while in [14] the case of polytopic uncertainty is considered and a numerically tractable relaxation of the problem is provided, together with an a-priori bound on the degree of conservatism of the approximation. In [3], an approximate solution is found by projecting the "optimizer" in an augmented feasible set of higher dimension.

In many practical situations, however, the introduced conservatism is not desired. In fact, the gap between the true robust performance and its estimates given by those conservative methods might be too large to provide any useful information. In order to overcome these difficulties, a different paradigm called *probabilistic robustness* has been proposed in the literature. From a robustness perspective, there are various motivations for the study of this topic, ranging from philosophical to computational; e.g., see [15, 16]. In this frame-

work, uncertain feedback systems are studied from a probabilistic point of view; see, e.g., [17, 18, 15, 16] and reference therein. A similar methodology can also be applied to control design problems, which is usually referred to as *scenario approach* [19, 20, 21].

In this probabilistic framework, randomized algorithms are used. More specifically, by randomly generating a large number of samples over the uncertainty set, we force the constraints to be satisfied for these uncertainty samples instead of being satisfied "robustly" for the whole uncertainty set. In this way, the robust constraints are "relaxed" to be satisfied for a finite member of uncertainty samples (they are fixed after sampling), rather than to be satisfied for all the infinitely many points in the uncertainty set. In analysis problems, the nominal system with an uncertainty sample is analyzed and the overall performance of the uncertain system is estimated by using all the samples. In robust design problems, by substituting every uncertainty sample into the robust condition, the robust constraint can be converted to a conventional constraint. Hence, the robust constraint can be approximated by a set of conventional constraints. By doing this, computationally efficient (polynomial-time) algorithms can be developed to address the robust control problems in the probabilistic control, which avoid two main drawbacks of the conventional robust control methods: NP-hardness and conservatism. The cost is that the results obtained by these randomized techniques are only shown to hold with a given degree of confidence (which is determined by the number of samples). In other words, given the stochastic nature of the algorithms, the solutions are always *soft* solutions, in that they always entail a strictly positive (though arbitrary small) probability of being "far away" from the optimal one.

The above discussed method is known as Monte-Carlo simulation in stochastic optimization in the literature. All results concluded from this method are known as two-level-probabilistic conditioned results. In contrast, the chance-constraints are one-level-probabilistic. To illustrate the difference, let us consider the following three constraints:

robust constraint (1.1), chance-constraint (1.2), and two-level-probabilistic constraint (1.3),

$$f(x,q) \leq 0 \text{ for all } q \in \mathcal{Q}, \tag{1.1}$$

$$\text{Prob}\left\{f(x,q) \leq 0\right\} \geq 1-\delta, \tag{1.2}$$

$$\text{Prob}\left\{\text{Prob}\left\{f(x,q) \leq 0\right\} \geq 1-\delta\right\} \geq 1-\epsilon, \tag{1.3}$$

where x is the design parameter to be determined, q is uncertainty support on the set $\mathcal{Q}$ and $f(x,q)$ is the desired constrain function.

In certain practical applications, it is desirable to handle the probability constraint (1.2) directly rather than using the two-level-probability version (1.3). This motivates the work in the first part of this dissertation. A different framework is proposed to address a robust design problem, which enjoys the following interesting properties:

- It is computationally efficient, i.e., it is solvable in *polynomial time* respect to the number of uncertainty parameters;
- It has a provable probability of constraint violation;
- Its solution can be made arbitrarily close to the solution of the robust optimization problem.

In general, the constraint (1.2) is non-convex and hard to address. To overcome this difficulty, we choose a so-called "kinship" function $\kappa(\cdot)$, which has the following key property:

$$\int_{\mathcal{Q}} \kappa[f(x,q)]\,\mu(q)\,\mathrm{d}q \geq \text{Prob}\left\{q \in \mathcal{Q} \ : \ f(x,q) > 0\right\}$$

where μ represents the probability density function of q. Moreover, we choose the kinship function such that

$$\int_{\mathcal{Q}} \kappa[f(x,q)]\,\mu(q)\,\mathrm{d}q$$

is convex in design parameter x. Thus, the constraint (1.2) can be relaxed to a convex constraint and can be addressed as a convex optimization. In this way, a novel approach based on kinship function design is developed in this dissertation to address the probability

constraint (1.2), which satisfies: 1) computationally efficient; 2) preserving hard probability bounds and 3) convergent.

1.2 Distributional Robustness Analysis

Continuing to the discussion in the previous section, it should be noted that in the aforementioned probabilistic framework, it is essential to assign a probability density function (PDF) on the uncertainty. In chance-constraint optimization problems, the probability constraints (1.2) are only well-defined with respect to some given PDFs. Similarly, in Monte-Carlo simulation based approaches, the uncertainty samples have to be randomly generated according to a specific PDF as well. Therefore, the very first question to be addressed is which distribution should be selected if no a priori information is available.

In many situations, the uniform distribution is adopted for a couple of reasons. First, using the uniform distribution is straightforward and makes the procedure of generating uncertainty samples much easier in many cases. Second, from an information theory point of view, it is natural to assign the uniform distribution if no a prior information is available on uncertainty distribution, as the uniform distribution is the maximum entropy distribution on any interval. Third, in certain cases, the uniform distribution leads to the worst-case estimation [22], which is naturally preferable in robust control.

However, it is not always the case that uniform distribution leads to the worst-case situation. Thus, from a robustness point of view, it is more desirable to find or analyze the worst-case situation regardless of any specific system uncertainty distribution. This is usually referred to as *distribution-free* approaches; e.g., see [22, 23, 24, 25, 26]. More specifically, in these approaches, the uncertainty parameters are assumed to be random without specific statistical description. Instead of assigning a given PDF, we assume made the "shape" of the probability density functions (PDFs) is known. For example, in [22] the PDFs are assumed to be symmetric and non-increasing with respect to the absolute value of the parameters. In this set-up, the questions aimed to answer are: i) what is the *worst-case* distribution of the uncertainty parameters in the pre-defined class that leads to the worst-case expected performance of the system? and ii) how to determine the worst-case

expected performance?

The distribution-free framework was first introduced in [22]. In that paper, the uniformity principle, which can be regarded as a theoretical justification for the use of the uniform distribution in Monte Carlo simulations, is proposed and applied to probabilistic stability analysis of a linear system with parametric uncertainty. The worst-case properties of the uniform distribution were further developed in [23], showing that the uniform distribution requires the minimum number of samples to attain a specific performance level for a class of Lipschitz continuous functions. In [27], it has been also proven that the worst-case expected value of multi-linearly parameterized H_∞ norm is attained by the uniform distribution. However, it turns out that this principle is not universally valid. For example, in the problem of determining Hurwitz stability of a polynomial, if its coefficients involve nonlinear uncertainty structures, only the so-called *truncated uniformity principle* holds in general. That is, the worst-case distribution is a so-called *truncated uniform distribution*. This truncation phenomenon is considered in [28, 29].

In terms of applications, this framework has been adopted to many robustness or performance analysis problems, especially in circuit networks. In [24], a new Monte Carlo paradigm was proposed for the analysis of resistive networks based on *distributional robustness*. In [26], an iterative algorithm was given to compute hard bounds on the probability of performance in circuit analysis, under the assumption that a property oracle exists. The worst-case probability of stability for polynomials with multi-linear uncertainty is studied in [25].

In this dissertation, we consider a general worst-case performance problem:

$$\sup_{f \in \mathcal{F}} \quad \mathcal{E}_f[g(q)], \tag{1.4}$$

where $\mathcal{F}$ is a given class of distribution and g is a continuous function. First we aim at finding sufficient conditions that guarantee the uniform principle, i.e. the uniform distribution leads to the supremum. Moreover, if this is not the case, algorithms are proposed to estimate the worst-case expected performance by solving convex optimization problems. One important property is that the corresponding estimates are *hard bounds* on the "true" value. That is,

unlike the ones estimated by using Monte Carlo simulation based approaches, which always entail a certain *risk of failure*, the hard bounds provided here give a *deterministic* estimate.

In summary, the contribution of the dissertation under this topic lies in three parts. First, a sufficient condition is provided to verify that if the uniform distribution is indeed one of the optimal solutions for problem (4.1), i.e.

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)] = \mathcal{E}_{\mathcal{U}}[g(q)]$$

This condition extends the applicability of the uniformity principle introduced in [22] to a particular set of semi-algebraic functions (which include polynomials). Second, it is shown that problem (4.1) can be recast as a infinite dimensional convex optimization problem. In addition, a hierarchy of converging linear matrix inequality (LMI) relaxations, which are finite dimensional convex programs, is provided. Finally, by introducing the concept of *bounding approximations* of indicator functions, a computationally efficient algorithm is provided to compute hard bounds on the probability of a semi-algebraic function being negative.

1.3 Hybrid System Identification

Under this topic, we consider a specific set of problems of hybrid system identification: identifying a switched affine system with minimum number of sub-models that is compatible with input/output data and a priori information. A switched affine system is a system whose behavior is determined by switching dynamics. These systems arise in many different contexts: for example, circuit network, biological systems, systems with interaction with logic devices and continuous processes. In addition, they can be used to approximate nonlinear dynamics. Thus, due to the potential application to a vast set of practical problems, the problem of identifying input/output hybrid models has attracted considerable attention, and several approaches have been developed.

There are many results available in the literature for the identification of switched affine systems. One may refer to a thorough review [30] for a summary of recent developments. In

the case where measurements are noise-free, an algebraic procedure, known as Generalized Principal Component Analysis (GPCA), has been proposed in [31, 32] to efficiently solve the problem. The problem can be also formulated as a mixed linear integer optimization problem [33] or in terms of linear complementary inequalities [34], leading to generically NP-hard problems. More recently, a greedy algorithm has been proposed to identify the system while minimizing the number of switches [35]. For robust identification of switched affine systems subject to process noise, an efficient moment-based convex approach using convex relaxations on rank minimization has been proposed in [36] . A similar approach was also pursued in [37] to solve a different problem: segmenting a collection of noisy measurements into subspaces.

In this dissertation, we consider the case of measurement noise, which has not been addressed in the literature, to the best of our knowledge. To tackle the problem, we first provide an equivalent polynomial optimization problem. It is then shown that the polynomial optimization problem inherently has a sparse structure and satisfies the so-called *running intersection property*. This sparse structure can be used to significantly reduce computational complexity as explained in [38, 39, 40]. Two algorithms are proposed based on this sparse optimization to further reduce computational cost. The first one is a randomized algorithm based on a hit-and-run type approach. The second one is deterministic and can be solved via an equivalent *fixed size* semi-definite program (SDP). A rank minimization relaxation algorithm is utilized to notably reduce the computational cost.

1.4 Generalized Fixed Order Interpolation

Before introducing the generalized fixed order interpolation problem, let us first briefly review the existing results for the classical interpolation problems. Conventional interpolation problems without degree constraint have long been studied with time-domain data and/or frequency-domain data available. For example, in frequency domain, a classical Nevalinna-Pick interpolation problem (without a degree constraint) is addressed in [41, 42], where it is shown a solution exists if and only if the so-called Pick matrix is positive semidefinite. Given $N+1$ interpolation points, it is known that an interpolant of degree at most N

always exists. Moreover, a constructive algorithm is given based on convex optimization for finding all such interpolants in [43]. Alternatively, given the interpolation conditions on time-domain data, the problem of finding all positive real functions is usually referred as the Caratheodory extension problem and a complete parameterization is given in [44]. A so-called *maximum entropy* solution is given in [45] to parameterize all rational covariance extensions of degree at most N, given $N+1$ interpolation points. Interpolatory identification algorithms based on linear matrix inequalities (LMIs) have been proposed in [46, 47, 48] to solve mixed time/frequency robust identification problems, resulting in interpolants of degree at most N given $N+1$ mixed domain interpolation points. Based on interpolation theory developed in [43], the problem of stabilizing controller synthesis is discussed in [49, 50]. On the other hand, in the area of system identification that closely relates to interpolation, polynomial based approaches have been proposed and discussed in the literature; e.g., see [51] for set-membership error-in-variables identification and see [52] for fixed-order identification with stability constraints.

In this dissertation, we go beyond the results available in the literature and provide a comprehensive framework for fixed-order interpolation which enables one to address a wide range of complex problems such as fixed order controller design, system identification and spectral estimation. Moreover, by relying on results on polynomial optimization and rank minimization, we provide efficient algorithms to address this complex problem.

1.5 The Sequel

The rest of the dissertation is organized as follows. Next chapter presents some notation and some preliminary results in polynomial optimization, which are important to understand the main contribution of this dissertation. Then, in the next four chapters, we introduce and discuss in details on our results and algorithms about the four research topics as introduced in Section 1.1 to Section 1.4: 1) robust and probabilistic optimization, 2) distributional robustness, 3) hybrid system identification, and 4) generalized fixed order interpolation. Chapter 7 gives concluding remarks of the dissertation. Most proofs, especially the lengthy ones, are put in the appendices to keep the flow of the dissertation smooth.

Chapter 2

Polynomial Optimization and Convex Relaxations

In this chapter, we summarize some results in the literature about polynomial optimization. The notation and results introduced thereafter are essential to understand the contributions, including theorems and algorithms, presented in this dissertation.

2.1 Notation

x^i	$x^i \doteq x_1^{i_1} \cdots x_d^{i_d}$ where $x = (x_1, \ldots, x_d) \in \mathcal{R}^d$ and $i = (i_1, \ldots, i_d) \in \mathcal{N}^d$ whenever x and i are multi-dimensional
$\mathcal{E}_\mu[p(x)]$	the mean value of $p(x)$ w.r.t the probability measure μ on x
$m = \{m_i\}_0^N$	the moment sequence where $m_i = \mathcal{E}_\mu x^i$ for some probability measure μ and $0 \leq i_1 + \ldots + i_d \leq N$; for example, if $d = 2$, $\{m_i\}_0^2 = \{m_{00}, m_{10}, m_{01}, m_{20}, m_{11}, m_{02}\}$
$M_N(m)$	the moment matrix in $\mathcal{R}^{\binom{N+d}{N} \times \binom{N+d}{N}}$ constructed by $m \doteq \{m_i\}_0^{2N}$
$M \succeq 0$	the matrix M is symmetric and positive semi-definite
$\|x\|_p$	the ℓ_p norm of the vector x, $p = 2$ or ∞

2.2 General Polynomial Optimization

Consider the following general constrained polynomial optimization problem:

$$p^*_{\mathcal{K}} := \min_{x \in \mathcal{K}} p_0(x) \tag{2.1}$$

where $\mathcal{K} \subset \mathbb{R}^d$ is a compact semi-algebraic set defined as

$$\mathcal{K} \doteq \{x \colon p_i(x) \geq 0, i = 1, \ldots, L\}, \tag{2.2}$$

where $p_i(x)$ are polynomial in x. This problem is usually not convex, and hence, hard to solve in general. Yet, let's consider a related problem in the probability measure space:

$$\tilde{p}^*_{\mathcal{K}} := \min_{\mu \in \mathcal{P}(\mathcal{K})} \int p_0(x)\mu(x)dx := \min_{\mu \in \mathcal{P}(\mathcal{K})} \mathbf{E}_\mu \left[p_0(x)\right] \tag{2.3}$$

where $\mathcal{P}(\mathcal{K})$ is the space of finite Borel probability measures on $\mathcal{K}$. Although (2.3) is an infinite dimensional problem, it is, in contrast to (2.1), convex. The following result, taken from [38], establishes the equivalence between the two problems:

Theorem 2.1 *Problems* (2.1) *and* (2.3) *are equivalent; that is:*

- $\tilde{p}^*_{\mathcal{K}} = p^*_{\mathcal{K}}$.
- *If* x^* *is a global minimizer of* (2.1)*, then the Dirac distribution* $\mu^* = \delta_{x^*}$ *with support on the point* x^* *is a global minimizer of* (2.3)*.*
- *For every optimal solution* μ^* *of* (2.3)*,* $p_0(x) = p^*_{\mathcal{K}}$-$\mu^*$ *almost everywhere.*

□

One direct consequence of this theorem is that, it is possible to develop a *convergent* sequence of LMI based convex relaxations to problem (2.1), where the optimization variables are $m_\alpha \doteq \mathbf{E}_\mu x^\alpha$, the moments of the unknown distribution μ.

2.2.1 Moment Sets

The moment set for all probability measures supported on the compact semi-algebraic set $\mathcal{K}$ is defined as

$$\mathcal{M}_{\mathcal{K}} \doteq \left\{ \{m_\alpha\}_{\alpha \in \mathcal{N}^d} : \begin{array}{l} m_\alpha = \int_{\mathcal{Q}} q^\alpha \mu(q) dq \text{ for some} \\ \text{measure } \mu \text{ supported on } \mathcal{K}, \text{for all } \alpha \end{array} \right\}, \tag{2.4}$$

where the vector $\{m_\alpha\}$ is called the *moment sequence* supported on $\mathcal{K}$. The *N-truncated* moment set, denoted by $\mathcal{M}_{\mathcal{K}}^N$, is a projection of $\mathcal{M}_{\mathcal{K}}$. More specifically, $\{m_\alpha\}^N \in \mathcal{M}_{\mathcal{K}}^N$ if there exists a moment sequence $\{\tilde{m}_\alpha\} \in \mathcal{M}_{\mathcal{K}}$ such that

$$m_\alpha = \tilde{m}_\alpha, \text{ for all } \alpha \in \mathcal{N}^d \text{ with } \alpha = (\alpha_1, \ldots, \alpha_d) \text{ and } 0 \leq \alpha_1 + \cdots + \alpha_d \leq N,$$

where $\{m_\alpha\}^N$ is called the *N-truncated moment sequence* supported on $\mathcal{K}$. Note the number of elements in $\{m_\alpha\}^N$ is $\mathcal{L}(N) = \binom{d+N}{d}$.

Now we define the *N-truncated moment matrix* and its corresponding *localizing matrices*.

Definition 2.2 (Moment Matrix) *An N-truncated moment matrix $M_N(m)$ is a symmetric matrix constructed by the 2N-truncated moment sequence*

$$\{m_\alpha\}^{2N} \doteq (m_{\alpha_1}, m_{\alpha_2}, \ldots, m_{\alpha_{\mathcal{L}(2N)}}),$$

i.e.,

$$M_N(m) \doteq \begin{pmatrix} m_{\alpha_0} & m_{\alpha_1} & \cdots & m_{\alpha_{\mathcal{L}(N)}} \\ m_{\alpha_1} & m_{\alpha_1+\alpha_1} & \cdots & m_{\alpha_1+\alpha_{\mathcal{L}}(N)} \\ \vdots & \vdots & \vdots & \vdots \\ m_{\alpha_{\mathcal{L}(N)}} & m_{\alpha_{\mathcal{L}(N)}+\alpha_1} & \cdots & m_{\alpha_{\mathcal{L}(N)}+\alpha_{\mathcal{L}(N)}} \end{pmatrix} \tag{2.5}$$

$\star$

For illustration and clarity of exposition, we consider the case where $x \in R^2$, i.e. $d = 2$,

then, the moment matrix $M_2(m)$ is defined as

$$M_2(m) = \begin{pmatrix} m_{0,0} & m_{1,0} & m_{0,1} & m_{2,0} & m_{1,1} & m_{0,2} \\ m_{1,0} & m_{2,0} & m_{1,1} & m_{3,0} & m_{2,1} & m_{1,2} \\ m_{0,1} & m_{1,1} & m_{0,2} & m_{2,1} & m_{1,2} & m_{0,3} \\ m_{2,0} & m_{3,0} & m_{2,1} & m_{4,0} & m_{3,1} & m_{2,2} \\ m_{1,1} & m_{2,1} & m_{1,2} & m_{3,1} & m_{2,2} & m_{1,3} \\ m_{0,2} & m_{1,2} & m_{0,3} & m_{2,2} & m_{1,3} & m_{0,4} \end{pmatrix}$$

Theorem 2.3 *An N-truncated moment matrix $M_N(m)$ is always symmetric and positive semi-definite.* □

Definition 2.4 (Localizing Matrix) *Given an N-truncated moment matrix $M_N(m)$ and a polynomial $p_k(x) \doteq \sum_i c_{\beta_i} x^{\beta_i}$ where $\beta_i \in \mathcal{N}^d$, we define the corresponding* localizing matrix

$$M_N(p_k m) \doteq \sum_i c_{\beta_i} \begin{pmatrix} m_{\alpha_0+\beta_i} & m_{\alpha_1+\beta_i} & \cdots & m_{\alpha_{\mathcal{L}(N)}+\beta_i} \\ m_{\alpha_1+\beta_i} & m_{\alpha_1+\alpha_1+\beta_i} & \cdots & m_{\alpha_1+\alpha_{\mathcal{L}(N)}+\beta_i} \\ \vdots & \vdots & \vdots & \vdots \\ m_{\alpha_{\mathcal{L}(N)}+\beta_i} & m_{\alpha_{\mathcal{L}(N)}+\alpha_1+\beta_i} & \cdots & m_{\alpha_{\mathcal{L}(N)}+\alpha_{\mathcal{L}(N)}+\beta_i} \end{pmatrix}. \tag{2.6}$$

⋆

To illustrate, consider the case $d = 2$ and a polynomial $p_1(x) = a - bx_1x_2$, then

$$M_1(p_1 m) = \begin{pmatrix} am_{0,0} - bm_{1,1} & am_{1,0} - bm_{2,1} & am_{0,1} - bm_{1,2} \\ am_{1,0} - bm_{2,1} & am_{2,0} - bm_{3,1} & am_{1,1} - bm_{2,2} \\ am_{0,1} - bm_{1,2} & am_{1,1} - bm_{2,2} & am_{0,2} - bm_{1,3} \end{pmatrix}$$

2.2.2 Convex Relaxations

Now let us look back on the general polynomial optimization problem (2.1). With the definitions of the moment matrix and localizing matrices, it is now possible to construct LMI based convex relaxations.

Theorem 2.5 (Convex Relaxation to Polynomial Optimization) *Let*

$$\begin{aligned} p_N^* = \quad & \min_{m} \quad \sum_{\alpha} p_{0,\alpha} m_\alpha \\ & s.t. \quad M_N(m) \succeq 0, \\ & \qquad M_N(p_i m) \succeq 0, i = 1, \ldots, d, \end{aligned} \tag{2.7}$$

where $p_{0,\alpha}$ is the coefficient of x^α in $p_0(x)$, m is the truncated moment sequence that constructs the N-truncated moment matrix $M_N(m)$ is the as defined in Definition 2.2 and $M_N(p_i m)$ is the corresponding localizing matrices as defined in Definition 2.4, then, under mild conditions,

$$p_N^* \uparrow p_{\mathcal{K}}^* \text{ as } N \to \infty. \tag{2.8}$$

□

If in addition $p_0(x) - p_{\mathcal{K}}^*$ has a Sum-of-Squares (SOS) representation on $\mathcal{K}$, i.e.,

$$p_0(x) - p_{\mathcal{K}}^* = t_0^2(x) + \sum_{i=1}^{L} p_i(x) t_i^2(x) \tag{2.9}$$

for some polynomial $t_0(x)$ of degree at most N and some polynomials $t_i(x)$ of degree at most $N - d_i/2$ where d_i is the degree of $p_i(x)$, then, it is possible to construct an *equivalent* LMI based convex optimization problem for (2.1). That is,

Theorem 2.6 (Exact Convex Relaxation to Polynomial Optimization with SOS) *In problem* (2.1)*, if $p_0(x) - p_{\mathcal{K}}^*$ has the representation form* (2.9)*, then*

$$p_N^* = p_{\mathcal{K}}^*. \tag{2.10}$$

□

2.2.3 Special Case: One-dimensional Polynomial

In general, there might not exist an exact convex relaxation for the polynomial optimization problem being considered. However, if the polynomials $p_i(x)$ are one-dimensional, i.e., $x \in \mathcal{R}$, there always exists an equivalent convex formulation for problem (2.1).

Theorem 2.7 (Markov-Lukacs Theorem) *Any one-dimensional algebraic polynomial $p(x)$ of degree n, non-negative on $[a, b]$, admits the following representation:*

$$p(x) = \left(\sum_{i=0}^{k} \alpha_i x^i\right)^2 + (b-x)(x-a)\left(\sum_{i=0}^{k-1} \beta_i x^i\right)^2; \quad \text{if } n = 2k; \tag{2.11}$$

$$p(x) = (x-a)\left(\sum_{i=0}^{k} \alpha_i x^i\right)^2 + (b-x)\left(\sum_{i=0}^{k-1} \beta_i x^i\right)^2; \quad \text{if } n = 2k+1; \tag{2.12}$$

The above theorem can be rewritten in matrix forms. Here we write it in matrix form for a special case: $a = 0$ and $b = +\infty$.

Theorem 2.8 *Define two series of Hankel matrices $H_{1,k} \in \mathcal{R}^{(n_1+1)\times(n_1+1)}$ and $H_{2,k} \in \mathcal{R}^{(n_2+1)\times(n_2+1)}$, with $n_1 = \lfloor \frac{n}{2} \rfloor$ and $n_2 = \lfloor \frac{n-1}{2} \rfloor$, as*

$$H_{i,0} = \begin{pmatrix} 1 & 0 & \dots \\ 0 & 0 & \dots \\ \vdots & \vdots & \vdots \end{pmatrix}, H_{i,1} = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \dots, H_{i,2n_i} = \begin{pmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

for $i = 1, 2$; and $H_{i,k} = 0$ for $k < 0$ or $k > 2n_i, i = 1, 2$. Then, a one-dimensional polynomial $p(x)$ of degree n is non-negative on $[0, +\infty]$ if and only if there exist two positive semi-definite matrices Y_1 and Y_2 such that

$$Tr(Y_1 H_{1,k}) + Tr(Y_2 H_{2,k}) = c_k, k = 0, 1, \dots, n, \tag{2.13}$$

where c_k is the coefficient of $p(x)$, i.e., $p(x) = \sum c_i x^i$.

2.3 Sparse Polynomial Optimization

The previous section describes how to build convex LMI relaxations to solve a polynomial optimization problem. However, in terms of complexity, it might become computationally intractable if the problem size is large, i.e., the dimension of x is large and/or the relaxation degree N is large. Note that the dimension of the moment matrix $M_N(m)$ is $\binom{d+N}{d}$, which grows polynomially in N *or* d but still very fast, as pointed out in [38, 53, 54]. On the other hand, however, many polynomial optimization problems encountered in practice have a sparse structure that can be exploited to decrease computational complexity; i.e., the polynomial p_i only contains a small fraction of the overall variables; e.g., see [40]. If the set of indices of variables in each $p_i(x)$ satisfies the so-called *running intersection property*, the size of the LMIs can be significantly reduced, i.e., at most $\binom{N+\xi}{N}$ where ξ is the largest number of variables appearing in each polynomial; e.g., see [39]. We now state the definition of this property.

Definition 2.9 (Running Intersection Property) *Let I_k, $k = 1, \ldots, \ell$, be the subsets of variables $X \doteq \{x_1, \ldots, x_d\}$ satisfying $\bigcup_{k=1}^{\ell} I_k = X$. If*

i) each constraint polynomial $p_i(x)$ uses only variables in I_k for some k;

ii) the objective polynomial can be written as $p_0 = p_{0,1} + \cdots + p_{0,n}$ where each $p_{0,i}$ uses only variables in I_k for some k,

then the running intersection property *is satisfied in Problem* (2.1) *if the collection $\{I_1, \ldots, I_\ell\}$ obeys the following condition:*

$$I_{k+1} \cap \left(\cup_{j=1}^{k} I_j \right) \subseteq I_s \text{ for some } s \leq k, \tag{2.14}$$

for every $k = 1, \ldots, \ell - 1$.

Similar to the result stated in Theorem 2.5 and Theorem 2.6, for a sparse polynomial optimization problem that satisfies the *running intersection property*, one can construct convex LMI relaxations as well. For simplicity, we denote $M_N(m, I_k)$ the moment matrix

for the reduced variables in the set I_k; and denote $M_{N_i}(p_j m, I_k)$ the localizing matrix with the reduced variables in I_k with respect to the polynomial $p_j(x)$. In the spirit of results in [38, 39], we have the following two theorems.

Theorem 2.10 (Convex Relaxations for Sparse Polynomial Optimization) *Assume that* (2.1) *satisfies the* running intersection property *in Definition 2.9 and let*

$$\begin{aligned} p_N^* = \quad & \min_{m} \quad \sum_{\alpha} p_{0,\alpha} m_\alpha \\ & s.t. \quad M_N(m, I_k) \succeq 0, \quad k = 1, \ldots, \ell \\ & \qquad M_N(p_i m, I_{k(i)}) \succeq 0, \quad i = 1, \ldots, d, \end{aligned} \tag{2.15}$$

where $p_i(x)$ contains only variables in $I_{k(i)}$. Then,

$$p_N^* \uparrow p_{\mathcal{K}}^* \text{ as } N \to \infty. \tag{2.16}$$

Theorem 2.11 (Sparse Polynomial Optimization with SOS) *Assume that* (2.1) *satisfies the* running intersection property *in Definition 2.9 and $p_0(x)$ has a* sparse *SOS representation on $\mathcal{K}$, i.e.,*

$$p_0(x) - p_{\mathcal{K}}^* = \sum_{k=1}^{L} \left(t_{k,0}^2(x) + \sum_{i=1}^{d} p_i(x) t_{k,i}^2(x) \right) \tag{2.17}$$

Then,

$$p_N^* = p_{\mathcal{K}}^*. \tag{2.18}$$

2.4 Conclusion

In this chapter, notation and some important preliminary results in polynomial optimization are introduced. Starting from the next chapter, we introduce some important problems in robust and probabilistic optimization, and show how they can be formulated and solved with the help of polynomial optimization theory.

Chapter 3

Robust and Probabilistic Optimization

In this chapter, we study a class of robust design problems with polynomial dependence on the uncertainty. One of the main motivations for considering these problems comes from robust controller design, where one often encounters systems which depend polynomially on the uncertain parameters. In classical robust control theory, it is required the constraints are satisfied for all possible values of the uncertainty. In contrast, in the area of probabilistic robustness, a probabilistic relaxation of the original robust problem is adopted, thus requiring the satisfaction of the constraints not for all possible values of the uncertainty, but for most of them.

The results presented in this chapter can be seen as integrated under the framework of probabilistic robustness. Being different from the randomized approach for tackling probabilistic relaxations, which is only guaranteed to provide soft bounds on the probability of satisfaction, we present a deterministic approach based on the novel concept of *kinship function* introduced in this chapter. This allows the development of an original framework, which leads to easily computable deterministic convex relaxations of the probabilistic problem. To make the convex relaxation computationally trackable, *optimal polynomial* kinship functions are introduced, which can be computed *a priori and once for all* and provide the "best convex bound" on the probability of constraint violation. More importantly, it

is proven that the solution of the relaxed problem converges to that of the original robust optimization problem as the degree of the optimal polynomial kinship function increases. Therefore, the proposed approach establishes an inherent connection between robust control and probabilistic control. Furthermore, by relying on quadrature formulae for computation of integrals of polynomials, it is shown that the computational complexity of the proposed approach is polynomial in the number of uncertain parameters. Finally, unlike other deterministic approaches to robust polynomial optimization, the number of variables in the ensuing optimization problem is not increased by the proposed approximation. An important feature of this approach is that a significant amount of the computational burden is "shifted" to a one-time off-line computation whose results can be stored and provided to any end user.

This chapter is organized as follows. First we introduce the concept of kinship functions followed by a motivation example. It is shown that there exists a whole family of possible kinship functions, and that any member of this family can be used to construct a deterministic *convex* relaxation of probabilistic robustness problem. Then, we show how to select, in an optimal way, *polynomial* kinship functions. This particular choice allows to obtain relaxations which are both i) computable and ii) asymptotically tight. Also, we stress that these optimal kinship functions can be computed offline and *once for all*. Furthermore, we introduce particular combinations of quadrature formulae, the so-called Smolyak formulae, which allow one to rewrite the optimization problem so as to ensure *polynomial-time* complexity of the algorithm. We also briefly discuss possible extensions of the proposed approach to the case of *vector* kinship functions, which allow one to directly tackle aggregated chance-constrained problems. Numerical examples are presented for illustration.

3.1 Introduction

It is a well-known fact that many problems arising in the context of analysis and control of systems subject to uncertainty can be recast in the form of *robust convex optimization problems*; see, for instance, [11, 2]. This deterministic approach has proven very successful in tackling problems where one has "simple" uncertainty structures. For instance, when the

robust problem to be solved depends linearly on the uncertainty, extreme point results can be easily applied; e.g., see [12].

For more general linear fractional dependence on norm-bounded *unstructured* uncertainty, efficient methods based on robust linear matrix inequalities (LMI) have been successfully developed; e.g., see [13]. However, the problem is significantly harder to solve when it involves structured uncertainty and computable solutions can be in general obtained only at the expense of introducing conservatism. Polynomial dependence on the uncertainty is a typical example. In this case, one is faced with substantial computational complexity barriers resulting from the semi-algebraic constraints.

In this chapter, we study a class of such problems and propose a framework for their solution. More precisely, we consider the general class of convex optimization problems subject to polynomial parametric uncertainty. That is, we are interested in robust optimization problems of the form

$$(RO): \quad \min_{x\in\mathcal{X}} \quad c^\top x \qquad \text{subject to} \quad f_i(x,q)\le 0 \text{ for all } q\in\mathcal{Q}, \quad i=1,2,\ldots,m \tag{3.1}$$

where q denotes uncertainty, $\mathcal{Q}$ is the uncertainty support set, and $f_i : \mathcal{X}\times\mathcal{Q}\to\mathcal{R}$ are convex functions in the design parameter x for fixed q and are polynomials in q for fixed x. This formulation is quite general. Not only it captures many robust control problems, it also encompasses many robust convex optimization problems that arise in diverse fields, ranging from finance [55, 56], truss structure optimization [57], beamforming [58], operations management [59] and analog IC design [60].

Recently, the interest in polynomially-constrained optimization problems has been revitalized by the introduction of approaches based on sum-of-squares (SOS) relaxations [61, 54], and other results of real algebraic geometry; e.g., see [38, 61, 53]. In particular, in [62], an algebraic approach is proposed to tackle robust optimization problems of the form of (3.1), for the case where the cost function f is linear in the design variable x, with the additional constraint that the admissible set for x is defined by LMI constraints. In [63], a matrix-version of SOS representation is given to construct relaxations for computing upper bounds

of robust optimization problems when uncertain parameters are constrained by polynomial matrix inequalities. These approaches systematically build an hierarchy of convex relaxations of (3.1). Such "lifted" problems provide a guaranteed – and hence feasible but not optimal – solution to the original problem. The nice feature of these methods, that makes them preferable to other conservative approaches, is that under mild assumptions one can prove asymptotic convergence of the solutions of the convex relaxations to the solution of (3.1) as one increases the order of the relaxation. In this sense, approaches like the ones mentioned above can be said to be *sufficient and asymptotically necessary*, since they provide a sequence of solutions whose cost converges *from above* to the optimal value of problem (3.1).

It is obvious that, to gradually remove conservatism, one has to increase the degree of such relaxations. This is done at the price of progressively introducing additional variables and constraints, hence increasing the problem size. In this way, the added computation burden can be quite large. For instance, for a polynomial optimization problem with d variables, if one uses a relaxation of order N, the size of the LMIs that one is required to solve is $\begin{pmatrix} d+N \\ d \end{pmatrix}$ and the number of optimization variables involved is $\begin{pmatrix} d+2N \\ d \end{pmatrix}$; e.g, see [38, 54, 53]. The added complexity as one uses better approximations is not surprising, since problem (3.1) is known to be, in general, NP-hard; see for instance [3].

In this chapter, a different approach is taken, and a probabilistic counterpart of (3.1) is considered. The idea is to assume a probabilistic nature for the uncertainty q and require probabilistic guarantees. More precisely, assuming that q is random with known probability density function (pdf) $\mu(q)$, one can fix an admissible level of violation probability ϵ and formulate the following probabilistic relaxation of (3.1)

$$(PO): \quad \min_{x \in \mathcal{X}} \quad c^\top x \qquad (3.2)$$
$$\text{subject to} \quad \text{Prob}\left\{q \in \mathcal{Q} \ : \ f_i(x,q) > 0\right\} \leq \epsilon, \quad i = 1, 2, \ldots, m.$$

Problems of this kind are well known in the statistics literature, where the constraints on the probability of violation of their deterministic counterpart are usually referred to as *chance constraints*; e.g., see [64] and references therein. From a robustness point of view,

motivations for considering a probabilistic approach are numerous, ranging from philosophical to computational; e.g., see [15, 16]. Here, we point out that this approach can be seen as "complementary" to the algebraic approach to robust polynomial optimization presented for instance in [62]. Indeed, the probabilistic relaxation (3.2) provides solutions which may be in general infeasible for the original problem (3.1). In contrary, the approach in [62] provides feasible solutions that are conservative and sub-optimal, in the sense that the relaxation give rise to an uncertainty set larger than the true one (hence leading to stranger constraints and leading to a subset of the true feasible set). Thus a solution of (3.2) leads to a corresponding functional value which is usually *lower* than the optimal one, while [62] provides an upper bound.

However, we remark that the probabilistic formulation (3.2) is by no means easier than the original problem (3.1). Indeed, problem (3.2) is immediately seen to be a hard nonconvex optimization problem, and the mere evaluation of the probabilities involved in the chance constraints requires, in general, the computation of complex multivariate integrals. For this reason, in recent years approaches based on randomized algorithms have been widely used for finding *approximate solutions* to (3.2). In particular, it has been shown in [17, 18] that one can use stochastic approximation algorithms to address this problem. Also, by randomly sampling the uncertainty, one can get a "large number" of constraints (rather than infinitely many) to obtain an approximate solution with given confidence [65]. Extensions of this approach to the case of nonconvex optimization problems have been recently proposed in [21].

However, the results obtained by these techniques are only shown to hold with a given *degree of confidence*. In other words, given the stochastic nature of the algorithms, the solutions are always *soft* solutions, in that they always entail a strictly positive (though arbitrary small) probability of being "far away" from the optimal one. This unavoidable risk of failure of the algorithm may be undesirable in many practical situations.

Motivated by these considerations, in this chapter we take a different viewpoint. The main objective of the proposed approach is to take advantage of the best features of the two approaches discussed above. More precisely, although still considering a probabilistic

formulation, we depart from the sampling-based techniques and approach problem (3.2) in a deterministic way. This is accomplished by the introduction of a novel concept: the *kinship function* (KF). This function can be viewed as a *convex approximation of the indicator function*. Using this concept enables one to preserve convexity of the optimization problem, while still providing an upper bound on the probability of constraint violation. That is, a kinship function $\kappa(\cdot)$ has the following key property

$$\int_{\mathcal{Q}} \kappa[f(x,q)]\,\mu(q)\,\mathrm{d}q \geq \mathrm{Prob}\,\{q \in \mathcal{Q} \;:\; f(x,q) > 0\} \tag{3.3}$$

where μ is the probability density function of q; see Section 3.4 for a precise definition of this class of functions. Using this concept, a series of convex *deterministic* relaxations of problem (3.2) is constructed using optimal polynomial kinship functions of increasing degree. Similarly to the behavior of SOS and algebraic methods, this hierarchy of relaxations is guaranteed to provide at each step deterministic (*hard*) solutions to the problem (3.2), and it is shown to *asymptotically converge* to the robust solution of (3.2) as the degree of the optimal kinship function increases. Moreover, particular combinations of quadrature formulae (the so-called *Smolyak formulae*) can be applied to ensure *polynomial-time* complexity with respect to the dimension of the uncertainty.

In this way, a significant amount of the computational burden is "shifted" from the user, as the most demanding part of the calculations is performed offline. More precisely, nodes and weights used in quadrature formulae are computed once and for all, stored in a repository and made available to any end user. Moreover, the number of optimization variables remains unchanged and the increased computational burden imposed on the end user, as better approximations are used, is only reflected on the number of elementary computations performed; i.e., the number of multiplications and sums performed by each step of the algorithm.

We also point out that the concept of kinship function introduced in this chapter constitutes an extension of the results in [66]. However, those results only address the problem of estimating the probability of performance satisfaction of a given system (by checking positivity of a polynomial), while we consider the more difficult problem of determining

the system parameters so to optimize the probability of performance satisfaction. In other words, using a system theoretical analogy, we may say that [66] addresses an *analysis* problem, while the approach introduced in this chapter provides algorithms that explicitly address probabilistic *design* problems.

3.2 A Motivating Example

To further motivate the proposed framework and to show its strong connections to classical robust design problems, a fixed-order SISO robust design example is now provided which is inspired by the examples in [67, 16].

We consider an uncertain plant described by the following transfer function

$$G(s,q) = 2(1 + 20q_1q_2 + 10q_1^2)\frac{s^2 + 1.5(1+q_2)s + 1}{(s-(2+q_3))(s+(1+10q_3q_4))(s+0.236)}$$

where $q = [q_1 \ q_2 \ q_3 \ q_4]^\top$ collects the uncertain terms acting respectively on the DC-gain, the numerator damping, and the pole locations of the plant. In this example, we assume that the support set of q is the hyper-rectangle

$$\mathcal{Q} = \left\{ q \in \mathbb{R}^4 \ : \ |q_1| \leq 0.05, |q_2| \leq 0.05, |q_3| \leq 0.1, |q_4| \leq 0.05 \right\}.$$

We first note that the above uncertain plant can be rewritten in the form

$$G(s,q) = \frac{b_2(q)s^2 + b_1(q)s + b_0(q)}{s^3 + a_2(q)s^2 + a_1(q)s + a_0(q)},$$

where $a_i(q)$ and $b_j(q)$ $i,j = 0,1,2$ are polynomials in q. The goal is to design the parameters $x = [x_1 \ x_2 \ x_3]^\top$ of a first order controller (if they exist)

$$C(s,q) = \frac{x_2 s + x_1}{s + x_3}$$

so that the closed-loop characteristic polynomial belongs to the following target stable

interval polynomial family for all admissible q:

$$\mathcal{P} \doteq \{p(s) = s^4 + c_3 s^3 + c_2 s^2 + c_1 s + c_0 \,|\, c_i \in [\underline{c}_k,\, \overline{c}_k],\, k = 0, 1, 2, 3\}$$

where $\underline{c} = [6 \quad 31.25 \quad 57 \quad 38.25]^\top$ $\overline{c} = [14 \quad 45.25 \quad 77 \quad 54.25]^\top$. If we define

$$A(q) = \begin{bmatrix} b_0(q) & 0 & a_0(q) \\ b_1(q) & b_0(q) & a_1(q) \\ b_2(q) & b_1(q) & a_2(q) \\ 0 & b_2(q) & 1 \end{bmatrix}, \quad d(q) = \begin{bmatrix} 0 \\ a_0(q) \\ a_1(q) \\ a_2(q) \end{bmatrix}$$

then simple algebraic manipulations show that the robust synthesis conditions are satisfied if and only if

$$\underline{c} \leq A(q)x + d(q) \leq \overline{c} \quad \text{for all } q \in \mathcal{Q}.$$

Applying the approach in [16] to the above robust linear constraints, we also associate a linear objective vector $c^\top \doteq [0\ 1\ 0]$, which amounts to seeking the robustly stabilizing controller having the smallest high-frequency gain. We thus obtain the robust linear program

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & c^\top x \\ \text{subject to} \quad & \mathcal{A}x + \mathcal{B} \leq 0 \end{aligned}$$

where

$$\mathcal{A} \doteq \begin{bmatrix} A(q) \\ -A(q) \end{bmatrix}, \quad \mathcal{B} \doteq \begin{bmatrix} d(q) - \overline{c} \\ -d(q) + \underline{c} \end{bmatrix}; \tag{3.4}$$

which is clearly a problem of the form (RO).

One should note that the numerical solution of this simply formulated problem is not "easy," since the coefficients $a_i(q), b_i(q)$ do not lie in independent intervals, and depend in a nonlinear (polynomial) way on q. Therefore, approaches such as the one in [67] cannot be directly applied in this case. However, as shown in Section 3.7, the results in this chapter provide a set of tools that enables one to efficiently tackle this problem.

3.3 Problem Formulation

For notation ease, and without loss of generality[1], in the sequel we consider robust optimization problems of the form (3.1) with a *single* functional constraint ($m = 1$), that is problems of the form

$$(RO): \quad \min_{x \in \mathcal{X}} \quad c^\top x \qquad \text{subject to} \quad f(x,q) \leq 0 \text{ for all } q \in \mathcal{Q}. \tag{3.5}$$

In the remaining of the chapter, the following standing assumptions on (3.5) are adopted.

Assumption 3.1 (On problem (3.5)**)**

(i) the set $\mathcal{X} \subset \mathcal{R}^{n_x}$ is a bounded convex set,

(ii) for fixed $\tilde{q} \in \mathcal{Q} \subseteq \mathcal{R}^d$, the function $f(x, \tilde{q})$ is convex in x,

(iii) for fixed $\tilde{x} \in \mathcal{X}$, the function $f(\tilde{x}, q)$ is polynomial in q with total degree less than or equal to σ,

(iv) the function $f(\cdot, \cdot)$ is bounded from below in $\mathcal{X} \times \mathcal{Q}$ by -1, i.e., $f : \mathcal{X} \times \mathcal{Q} \to [-1, \infty)$,

(v) for fixed $q = \tilde{q}$, a (sub)gradient $\partial_f(x, \tilde{q})$ of $f(x, \tilde{q})$ with respect to x is available.

Assumption 3.1.(iv) is equivalent to requiring that a finite lower bound on f is available. For instance, such a lower bound can be computed using a relaxation of the problem of minimizing $f(x, q)$ in $(\mathcal{X}, \mathcal{Q})$. Once this lower bound is obtained, one can normalize the function f in such a way that its minimum value is not smaller than -1.

Remark 3.2 (On the convexity assumption) *The assumption on the function f being convex in the design variable x is made so as to obtain a computationally meaningful framework. For this reason, the kinship functions are specially tailored to maintain convexity of the final formulation. However, one should point out that the main ideas proposed in this*

[1] One may consider problems with multiple constraints ($m > 1$) as well; see the discussion in Section 3.8.

chapter can be extended to non-convex formulations. In this case, all the derived results would still hold, with the understanding that one might only obtain local optimality. Moreover, in the non-convex case, one can relax the convexity condition on the kinship functions in order to obtain better approximations of the probability of constraint violation. See [68] for a possible way of designing polynomial approximations for estimating the probability of constraint violation.

As for the probabilistic counterpart of (3.5),

$$(PO): \quad \min_{x\in\mathcal{X}} \quad c^\top x \qquad \text{subject to} \quad \text{Prob}\{q\in\mathcal{Q} \ : \ f(x,q)>0\}\leq\epsilon, \tag{3.6}$$

the only additional assumption made is that q has independently distributed entries with finite moments.

Assumption 3.3 (On problem (3.6)**)**

(i) The probability density function $\mu(q)$ and its support set $\mathcal{Q}$ can be written as

$$\mu(q) = \mu_1(q_1)\mu_2(q_2)\cdots\mu_d(q_d) \tag{3.7}$$

$$\mathcal{Q} = \mathcal{Q}_1\times\mathcal{Q}_2\times\cdots\times\mathcal{Q}_d. \tag{3.8}$$

(ii) All the moments of $\mu(\cdot)$ are finite.

3.4 Kinship Functions and Convex Relaxations

In this section, we define the central concept of this chapter: the kinship function (KF). Furthermore, it is shown how to use this tool to construct a convex relaxation of problem (3.6).

Definition 3.4 (Kinship function) *A function $\kappa:[-1,\infty)\to\mathbb{R}$ is said to be a* kinship function *if it satisfies the following properties:*

(a) $\kappa(0) = 1$,

(b) $\kappa(y)$ *is convex for* $y \in [-1, \infty)$,

(c) $\kappa(y)$ *is nonnegative and nondecreasing function for* $y \in [-1, \infty)$.

It should be noted that the definition above introduces an entire *family* of kinship functions (all possible scalar functions satisfying properties *(a)-(c)*). The next theorem shows that any function that complies with Definition 3.4 provides a way of computing an upper bound on the probability of constraint violation.

Theorem 3.5 (Main property of KF) *Let* $\kappa(\cdot)$ *be a kinship function, and define the integral quantity*

$$V_\kappa(x) \doteq \int_{\mathcal{Q}} \kappa[f(x,q)]\, \mu(q)\, \mathrm{d}q. \tag{3.9}$$

Then, it holds that

$$Prob\{q \in \mathcal{Q} \ : \ f(x,q) > 0\} \leq V_\kappa(x).$$

Proof. By definition, the kinship function $\kappa[f(x,q)]$ is non-negative in $[-1, \infty)$ and greater than one if $f(x,q) \geq 0$. Thus, for any probability measure μ on $\mathcal{Q}$,

$$\begin{aligned} V_\kappa(x) &\geq \int_{\{q \in \mathcal{Q}: f(x,q) > 0\}} \kappa[f(x,q)] \mu(q) \mathrm{d}q \\ &\geq \int_{\{q \in \mathcal{Q}: f(x,q) > 0\}} \mu(q) \mathrm{d}q \\ &= \mathrm{Prob}\{q \in \mathcal{Q} \ : \ f(x,q) > 0\}. \end{aligned}$$

□

Theorem 3.5 is of fundamental importance for the framework developed in this chapter, since it provides a simple and direct relaxation of the probabilistic optimization problem (3.6). This is obtained by simply substituting the chance constraint in problem (3.6)

with its upper bound $V_\kappa(x)$. We refer to this problem as the *kinship-relaxed* problem

$$(PO\kappa) : \min_{x\in\mathcal{X}} \quad c^\top x \qquad \text{subject to} \quad V_\kappa(x) \leq \epsilon. \tag{3.10}$$

Moreover, properties of the kinship functions imply that the kinship-relaxed problem (3.10) is a convex problem, as stated next.

Theorem 3.6 (Convexity of the kinship-relaxation) *The relaxed optimization problem* (3.10) *is convex in* x*.*

Proof. As assumed, $f(x, q)$ is convex in x for any fixed q. Moreover, the kinship function $\kappa(\cdot)$ is non-decreasing and convex. Thus, the composite function $\kappa[f(x, q)]$ is convex in x for fixed q. Hence, the integral $V_\kappa(x)$ is convex since μ is a probability (non-negative) measure on $\mathcal{Q}$. Note that non-negative weighted integration is a standard operation that preserves convexity and composition of a convex function by a convex and non-decreasing function also preserves convexity; see e.g. [69]. □

As already remarked, Theorems 3.5 and 3.6 introduce a whole family of possible convex relaxations of problem (3.6). A natural question that arises is the following: *is it possible to select in some "optimal" way a particular KF, in order to obtain relaxations which are i) computable and ii) tight?* The answer to this question is given in the rest of this chapter. In particular, in Section 3.5, we define an optimal polynomial kinship function of given degree ϱ, and show how to efficiently construct it. With these at hand, we prove that, as the degree increases, the solution of the kinship-relaxed problem will converge to the solution of problem (3.1). Then, in Section 3.7, we show how to efficiently numerically solve the ensuing minimization problem.

3.5 Optimal Polynomial Kinship Functions

With the aim of constructing the "best possible" kinship function, we first formally introduce the following optimality criterion.

Definition 3.7 (Optimal kinship function) *An* optimal *kinship function is defined as the solution of the optimization problem*

$$\kappa_\star(\cdot) \doteq \arg\min_{\kappa\in\mathcal{C}} \quad \int_{-1}^{0} \kappa(y)dy \qquad (3.11)$$
$$\text{subject to} \quad \text{(a)-(c) in Definition 3.4},$$

where $\mathcal{C}$ is some subset of continuous functions defined on the interval $[-1,\infty)$.

Remark 3.8 (Interpretation of optimality) The above optimization problem is indeed a problem of seeking a continuous convex function in $\mathcal{C}$ over $[-1,\infty)$ that minimizes its L_1 norm on $[-1,0]$. The function $\kappa_*(\cdot)$ here is acting as a weighting function for $f(x,q)$. In this way, one can expect that the subset $\{q\in\mathcal{Q} : f(x,q)<0\}$ has the smallest possible contribution to the integral $V_\kappa(x)$. In more general cases, an additional weighting function $w(\cdot)$ may be introduced, to minimize the weighted norm of $\kappa(\cdot)$ on $[-1,0]$, i.e., to minimize $\int_{-1}^{0}\kappa(y)w(y)\mathrm{d}y$.

In our framework, we are mostly interested in seeking the optimal kinship function over *polynomials*. The following lemma shows that in this case, the conditions in Definition 3.4 can be substantially simplified, thus leading to a simpler form when considering optimal polynomial kinship functions (OPKFs).

Proposition 3.9 (Optimal polynomial KFs) *The optimal polynomial kinship function*

of degree ϱ, denoted as $\kappa_\varrho(y)$, is the solution of the following optimization problem

$$\min_{a_0,\ldots,a_\varrho} \quad \int_{-1}^{0} p(y)dy \tag{3.12}$$

$$\text{subject to} \quad p(y) = a_0 + a_1 y + \cdots + a_\varrho y^\varrho.$$

$$p(0) = 1, \tag{3.13}$$

$$p(-1) = 0, \tag{3.14}$$

$$p'(-1) = 0, \tag{3.15}$$

$$p''(y) \geq 0 \text{ for } y \in [-1, \infty). \tag{3.16}$$

Proof. See Appendix A.2. □

It is easy to see that the objective function is a linear function of the coefficient $a_0, \ldots, a_\varrho$, and the first three conditions (3.13), (3.14) and (3.15) are also linear. Moreover, according to Theorem 2.7 and Theorem 2.8, a one-dimensional non-negative polynomial on an interval, which might be finite, infinite or semi-infinite, can be expressed as sum-of-squares. As a direct consequence of these two theorems, condition (3.16) can be shown to be equivalent to a set of LMI constraints given below. These facts are summarized in the following corollary.

Corollary 3.10 (Computation of optimal polynomial KFs) *Define two series of Hankel matrices $H_{1,k} \in \mathcal{R}^{(n_1+1)\times(n_1+1)}$ and $H_{2,k} \in \mathcal{R}^{(n_2+1)\times(n_2+1)}$, with $n_1 = \lfloor \frac{\varrho-2}{2} \rfloor$ and $n_2 = \lfloor \frac{\varrho-3}{2} \rfloor$, as*

$$H_{i,0} = \begin{pmatrix} 1 & 0 & \ldots \\ 0 & 0 & \ldots \\ \vdots & \vdots & \vdots \end{pmatrix}, H_{i,1} = \begin{pmatrix} 0 & 1 & 0 & \ldots \\ 1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \ldots, H_{i,2n_i} = \begin{pmatrix} 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix}$$

for $i = 1, 2$; and $H_{i,k} = 0$ for $k < 0$ or $k > 2n_i$. Then, the optimization problem in Lemma

3.9 can be reformulated as follows.

$$
\begin{aligned}
\min_{a_0,\ldots,a_\varrho,Y_1,Y_2} \quad & \sum_{i=0}^{\varrho} \frac{(-1)^i}{i+1} a_i \\
\textit{subject to} \quad & a_0 = 1, \\
& \sum_{i=0}^{\varrho} (-1)^i a_i = 0, \\
& \sum_{i=1}^{\varrho} i(-1)^{i-1} a_i = 0, \\
& Tr(Y_1 H_{1,k}) + Tr(Y_2 H_{2,k}) - \sum_{i=k+2}^{\varrho} \frac{i!(-1)^{i-k-2}}{k!(i-k-2)!} a_i = 0, \ \textit{for } k = 0, \ldots, \varrho - 2, \\
& Y_k \succeq 0, \ \textit{for } k = 1, 2.
\end{aligned}
$$

Proof. One needs only to prove the equivalence of (3.16) to the conditions above. The remaining parts are immediate. As a direct consequence of Theorem 2.7, $p''(y)$ is non-negative on $[-1, \infty)$, or equivalently, $p''(y-1)$ is non-negative on $[0, \infty)$, if and only if it has the following representation,

$$p''(y-1) = \sum_i r_i^2(y) + y \sum_i s_i^2(y),$$

where $r_i(y)$ and $s_i(y)$ are polynomials of degree n_1 and n_2, respectively. Then, as shown in Theorem 2.8, the representation above is equivalent to the following conditions,

$$p''_k = \mathrm{Tr}(Y_1 H_{1,k}) + \mathrm{Tr}(Y_2 H_{2,k}), k = 0, \ldots, \varrho - 2,$$

where $p''_k = \sum_{i=k+2}^{\varrho} \frac{i!(-1)^{i-k-2}}{k!(i-k-2)!} a_i$ is the coefficient of the term y^k in $p''(y-1)$, and Y_1, Y_2 are positive semi-definite matrices. □

This corollary shows that the computation of the optimal polynomial KF of given degree ϱ amounts to solving a *finite dimensional* SDP problem, for which efficient algorithms exist, see e.g. [70, 71]. More importantly, there are two points that need to be stressed. First, the kinship polynomial functions are scalar polynomials, hence, the non-negative

polynomial in equation (3.16) can always be represented by sum-of-squares. Therefore, there is no need to build SDP relaxations and the computational cost is relatively small comparing to that needed to solve problem (PO_κ) in the proposed algorithm. Second, these functions can be computed off-line, *a priori* and *once for all*. For illustration purposes, in Fig. 3.1 we depict OPKFs of degrees $\varrho = 3, 5, 10$. These curves allow also to derive an intuitive interpretation of the concept of kinship function: indeed a KF can be seen as a convex upper bound of the indicator function, hence giving rise to a (convex) upper bound on the probability of constraint violation.

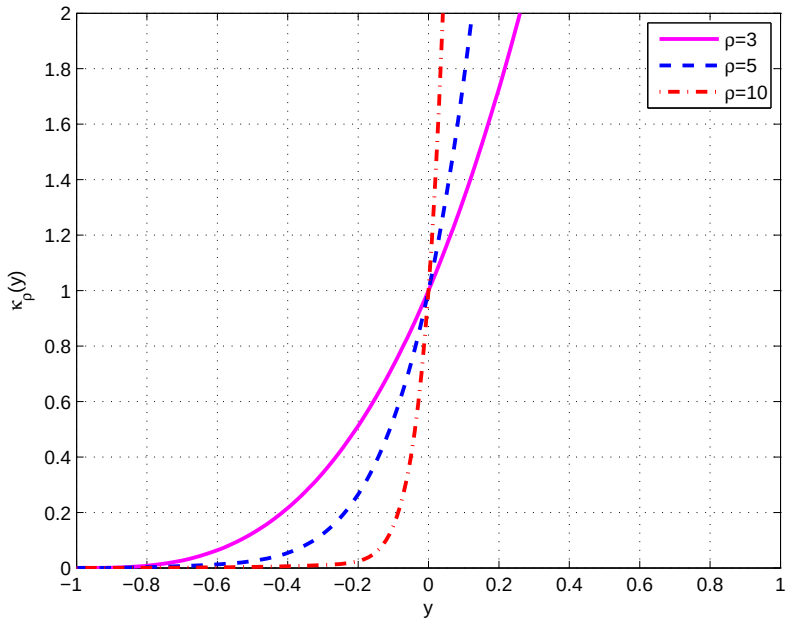

Figure 3.1: Optimal polynomial kinship functions of degrees $\varrho = 3, 5, 10$.

Remark 3.11 (Numerical issues) *We remark that it is a common problem when dealing with polynomial optimization to run into numerical difficulties, due to the well-known fact that the canonical polynomial basis is numerically ill-conditioned. For this reason, in an efficient implementation, the use of other polynomial bases, for example Chebychev polynomials, can be adopted to improve numerical stability; e.g. see [72].*

Once a repository of optimal polynomial KFs of increasing degree is computed and made available, they can be employed for constructing a sequence of approximations to the original problem (3.1), as shown in the next section.

3.6 Asymptotic Tightness of the Proposed Relaxation

With optimal polynomial kinship functions at hand, the intuition is that, for any given $x \in \mathcal{X}$, the set $\{q \in \mathcal{Q} : f(x,q) < 0\}$ contributes less in calculating the integral $V_\kappa(x)$ as the degree ϱ increases. Moreover, the contribution of the set $\{q \in \mathcal{Q} : f(x,q) \geq 0\}$ becomes dominant for large values of ϱ. This intuition is indeed true, as formally stated in the next theorem, which provides the asymptotic properties of the relaxed problem (3.10) using OPKFs, and constitutes one of the main results of this chapter.

Theorem 3.12 (Asymptotic convergence) *Assume that the uncertainty set $\mathcal{Q}$ is compact, and that the pdf $\mu(q)$ is strictly positive and bounded on $\mathcal{Q}$. Moreover, let the original robust problem* (3.1) *admit a unique solution x^*. Let $\epsilon > 0$, then, for ϱ sufficiently large, the set*

$$\{x \in \mathbb{R}^n : V_\kappa(x, \varrho) \leq \epsilon\}$$

where

$$V_\kappa(x, \varrho) \doteq \int_{\mathcal{Q}} \kappa_\varrho[f(x,q)]\mu(q)\mathrm{d}q \tag{3.17}$$

is nonempty. Furthermore, let x^*_ϱ be the solution of

$$(PO\kappa_\varrho): \quad x^*_\varrho \doteq \arg\min_{x \in \mathcal{X}} c^\top x \quad s.t. \quad V_\kappa(x, \varrho) \leq \epsilon. \tag{3.18}$$

Then,

$$\lim_{\varrho \to \infty} x^*_\varrho = x^*.$$

Proof. See Appendix A.3. □

In other words, Theorem 3.12 formally guarantees that the solution of the relaxed problem (3.18) can be made arbitrarily close to the unique solution of the robust problem (3.5), as long as the degree ϱ is chosen large enough.

Remark 3.13 (On the assumptions in Theorem 3.12) *It should be noted that, the assumption on robust feasibility in Theorem 3.12 is necessary to guarantee convergence of the proposed approach. The convergence is a direct consequence of the convexity constraint on the optimal kinship function, needed for convexity of the overall problem. This may lead to a conservative estimate of the probability of violation since the value of optimal kinship function $\kappa_\varrho(y)$ for $y > 0$ tends to infinity as $\varrho \to \infty$. In other words, when the probability of violation is strickly positive for all admissible values of x, the value of $V_\kappa(x, \varrho)$ might tend to infinity as $\varrho \to \infty$ and, hence, convergence cannot be assured.*

Similarly, we remark that Assumption 3.1.(iv) on the availability of a lower bound is not a restrictive one in terms of the convergence of the proposed algorithm. Indeed, although "loose" lower bounds might lead to possibly more conservative estimates of the probability of violation, this does not affect the convergence results in Theorem 3.12. This is a consequence of the fact that the value of the optimal kinship function $\kappa_\varrho(y)$ tends to zero as $\varrho \to \infty$ for any $-1 \leq y < 0$.

Finally, we note that the assumption on the compactness of $\mathcal{Q}$ is essentially made for guaranteeing the existence of the robust deterministic solution x^*.

Remark 3.14 (Relations with the dilation integral approach) *As already mentioned in Section 3.1, the idea of kinship function is mainly motivated by the recent literature on dilation integrals, e.g., see [66]. In particular, in [66] it is shown that, for fixed $x \in \mathcal{X}$, the dilation integral*

$$\int_{-1}^{\infty} (1 + \alpha f(x, q))^{2k} \, \mathrm{d}q, \quad k = 1, 2, \ldots, \quad \alpha > 0 \tag{3.19}$$

provides an upper bound on the violation volume, i.e., the volume of the set $\{q \in \mathcal{Q} : f(x, q) > 0\}$. Thus, the result in [66] is an "analysis" result. Our formulation can be seen as an extension of the dilation approach to the "design" case. To see this, notice that the

function

$$\kappa(y) = (1 + \alpha\, y)^{2k}, \quad k = 1, 2, \ldots$$

can be easily shown to comply with Definition 3.4, hence it is a member of the class of kinship functions. However, this function is not an optimal KF. In particular, the choice of KFs which are optimal and polynomial is what allows us to extend the asymptotic convergence properties of the dilation results (which are limited to analysis) to the solution of design problems.

We also note that straightforward symbolic calculations (e.g., using the MATLAB Symbolic Math Toolbox) can be used to integrate the polynomial functions involved in $V_\kappa(x, \varrho)$. However, it should be remarked that this approach may not be tractable for high dimensions, since the number of monomials resulting from symbolic integration increases exponentially. Indeed, when ϱ and/or d are large, the required intermediate symbolic expressions involved in the computation can easily lead to an amount of data exceeding MATLAB memory limitations. For this reason, in the next section we extend the computational approach proposed in [66] and present a computationally efficient approach for evaluating $V_\kappa(x, \varrho)$ when using polynomial kinship functions.

3.7 An Efficient Numerical Solution to the Kinship-relaxed Problem

3.7.1 Preliminaries on Quadrature Formulae

The main idea underlying the technique proposed in this chapter for numerically solving the optimization problem (3.18) is to compute the integral in (3.17) using the *quadrature formula* (QF). It is a well-known fact that, given a scalar function $g : \mathcal{Q}_j \to \mathbb{R}$ and a weighting function $\mu_j : \mathbb{R} \to \mathbb{R}$ with bounded moments, the one-dimensional definite *weighted* integral

$$I[g] = \int_{\mathcal{Q}_j} g(q_j)\mu_j(q)\mathrm{d}q_j$$

can be approximated using an N-point quadrature formula of the form

$$\mathsf{Q}_{j,N}[g] = \sum_{k=1}^{N} \omega_{j,k}\, g(\theta_{j,k}) \tag{3.20}$$

where $\theta_{j,k}$, $\omega_{j,k}$, $k = 1, \ldots, N$ are, respectively, the nodes and weights of the quadrature. In this one dimensional case, it can be proven that, for a so-called *Gauss formulae* QF constructed over N nodes, the integral $I[g]$ of any polynomial g of degree no greater than $2N - 1$ can be evaluated exactly; e.g., see [73]. It is evident that this approach can be extended in a straightforward way to the general multidimensional problem of computing a multivariate integral, as long as the density function μ satisfies Assumption 3.3. That is, if one chooses N_j nodes for x_j, the equation

$$I[g] \equiv \int_{Q_1} \cdots \int_{Q_d} g(q)\mu(q)dq = \sum_{k_1=1}^{N_1} \cdots \sum_{k_d=1}^{N_d} (\omega_{1,k_1} \cdots \omega_{d,k_d}) g\left(\theta_{1,k_1}, \ldots, \theta_{d,k_d}\right) \tag{3.21}$$

holds for all multivariate polynomials $g(x_1, \ldots, x_d)$ of degree no greater than $2N_j - 1$ in x_j, $j = 1, \ldots, d$; e.g., see [73]. The multidimensional quadrature formula appearing on the right-hand side of equation (3.21) is sometimes referred to as *tensor* QF.

Unfortunately tensor QFs can become intractable due to the exponential growth of the number of functional evaluations as dimension d increases. To avoid this computational difficulty, we propose an approach based on a combination of sequences of low-order quadrature formulae, the so-called *Smolyak formulae* [74]. These formulae are based on the evaluation of the integrand on so-called *sparse grids* (see [75] and references therein), which are highly non-uniform grids composed of a subset of the nodes of a tensor QF. In particular, a d-dimensional Smolyak formula with precision level ℓ is an $\mathsf{N}_d^{(\ell)}$-point multidimensional QF with nodes $\Theta_k \in \mathbb{R}^d$ and weights $\mathsf{w}_k \in \mathbb{R}$, $k = 1, \ldots, \mathsf{N}_d^{(\ell)}$:

$$\mathsf{S}_d^{(\ell)}[g] = \sum_{k=1}^{\mathsf{N}_d^{(\ell)}} \mathsf{w}_k g(\Theta_k). \tag{3.22}$$

The number of nodes $\mathsf{N}_d^{(\ell)}$ of the Smolyak formula depends on the dimension d and on

a precision index ℓ. Fig. 3.2 provides an example for $d = 3$ and different values of ℓ. These formulae enjoy two interesting properties which are fundamental for our successive developments: i) they have a number of nodes which is polynomial in d, ii) they have degree of exactness (DoE) $2\ell+1$, i.e. they can be used to evaluate integrals of polynomials of degree no greater than $2\ell + 1$ exactly. For ease of reading and completeness of the chapter, a brief

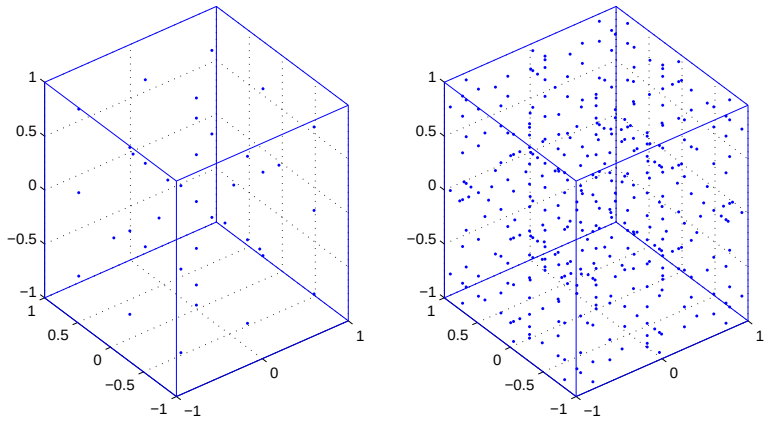

Figure 3.2: Three dimensional sparse grids for $\ell = 3$ and $\ell = 7$. The number of grid points are respectively 39 and 399.

introduction to Smolyak formulae is reported in Appendix A.1, where these fundamental properties are discussed in details.

An important point which should be remarked is that Smolyak formulae do not depend on any specific integrands. More precisely, for given dimension d, precision index ℓ, and weighting function μ, the nodes (and respective weights) can be computed *a priori and once for all*, and then stored for later computations. Specific algorithms for fast construction of the coefficients of Smolyak formulae are provided in [76]. Since the procedure for computing the nodes and weights is a time consuming job, a repository of nodes and weights for different values of the polynomial dimension and total degree, and different pdfs, has been created and is available on request. Some of them can be downloaded at [77].

3.7.2 A Finite Representation of Problem (3.18)

With the Smolyak formulae given in the previous section, we are ready to present the following theorem, which provides a finite representation of the integral $V_\kappa(x, \varrho)$ appearing in the optimization problem (3.18).

Theorem 3.15 (A reformulation of (3.18)) *Consider problem (3.18) and let Θ_k, w_k, $k = 1, \ldots, \mathsf{N}_d^{(\ell)}$ be the nodes and weights of a Smolyak formula constructed using a sequence of QF satisfying Definition A.1 in Appendix A.1, with $\ell = \left\lceil \frac{\sigma\varrho - 1}{2} \right\rceil$. Then, using the notation in (3.22), the optimization problem ($PO\kappa_\varrho$) is equivalent to the following one*

$$\min_{x \in \mathcal{X}} \quad c^\top x \tag{3.23}$$

$$\text{subject to} \quad \sum_{k=1}^{\mathsf{N}_d^{(\ell)}} \mathsf{w}_k \kappa_\varrho[f(x, \Theta_k)] \leq \epsilon. \tag{3.24}$$

Moreover, for fixed ϱ, the number of points $\mathsf{N}_d^{(\ell)}$ is polynomial in d. ⋆

Proof. The theorem is a direct consequence of Lemma A.2 and A.3 reported in Appendix A.1. □

Problem (3.23-3.24) is now formulated as a classical convex optimization program, and standard tools such as (sub)gradient descent or ellipsoidal/cutting plane localization methods (e.g., see [69, 78, 79]) can be applied to solve it. In particular, the (sub)gradient of V_κ can be immediately obtained. This is shown in the next proposition, which can be verified by direct computation.

Proposition 3.16 (Subgradient of V_κ) *For given $q \in \mathcal{Q}$, let $\partial_f(x, q)$ be a (sub)gradient of the function f with respect to x. Consider a polynomial kinship function of degree ϱ, $\kappa_\varrho(y) = \sum_{j=0}^{\varrho} a_j y^j$. Then*

$$\partial_{V_\kappa}(x) \doteq \sum_{k=1}^{\mathsf{N}_d^{(\ell)}} \mathsf{w}_k \partial_f(x, \Theta_k) \left(\sum_{j=0}^{\varrho} j\, a_j [f(x, \Theta_k)]^{j-1} \right) \tag{3.25}$$

is a (sub)gradient of $V_\kappa(x, \varrho)$.

Remark 3.17 (Connections with the scenario approach) *In [65], the authors derived explicit bounds on the number of random samples necessary to guarantee that (using our notation)*

$$Prob\{Prob\{q \in \mathcal{Q} :\ f(\tilde{x}, q) > 0\} \leq \epsilon\} \geq 1 - \beta, \tag{3.26}$$

where $\tilde{x}$ *is the solution of the so-called scenario problem, obtained by replacing the semi-infinite constraints in problem* (3.1) *with a finite number* N *of sampled scenarios, constructed by evaluating the constraints on randomly selected values* $q^{(i)} \in \mathcal{Q}$. *In other words, the authors derived bounds on the sample complexity* N *needed to obtain a solution having accuracy* ϵ *which has a confidence level of* $1 - \beta$. *The best currently available sample complexity bounds for scenario problems are given in [20], while in [21] the setup is extended to the case when a violation level can be fixed a-priori. Also, in [21] explicit sample complexity results are derived for non-convex scenario problems.*

One can argue that the approach proposed in this chapter is, in some sense, related to the scenario approach mentioned above, since it also uses points drawn from the uncertainty set to approximately solve the robust optimization problem. However, there are fundamental differences between the two methodologies. First, as already mentioned in the Section 3.1, the approach considered here does not involve any second level of probability. One may argue that this could be minor, since one may choose the confidence level $1 - \beta$ *very high (say* $\beta = 10^{-12}$*), but it should be stressed that they are conceptually and philosophically very different, since a result that holds with probability close to one is by no means a deterministic guarantee. Second, the nodes that we choose from the uncertainty set are carefully selected, in a thoroughly deterministic way, so that they provide* exact *computations.*

From a computation point of view, it is hard to compare these two approaches, since the number of nodes/samples needed depends on different quantities. In the case of the approach presented in this chapter, the number of nodes depends on the order of the polynomials and on the dimension of the uncertainty set, while in the scenario approach the number of samples depends on the the so-called accuracy and confidence levels.

On the other hand, it should be noted that, in the scenario approach, the complexity of

the optimization algorithm to be solved increases as the number of samples increase; i.e., one has an optimization problem that has an increasing number of constraints. In our case, the number of constraints remains invariant and, in this sense, the complexity of the "final optimization problem" to be solved does not increase with the number of nodes.

3.8 Extensions: Aggregate Chance-constrained Problems

In this section, we show how our approach may be extended to problems involving *aggregate* chance constraints, i.e. problems of the form

$$(PO): \quad \min_{x\in\mathcal{X}} \quad c^\top x \tag{3.27}$$
$$\text{subject to} \quad \text{Prob}\left\{q\in\mathcal{Q} \ : \exists i\in\{1,\dots,m\} \text{ such that } \ f_i(x,q)>0\right\}\leq\epsilon,$$

for $m>1$.

In other words, differently from the problem (3.2) formulated in Section 3.1, in this problem we aim at finding x that guarantees, with high probability, the *simultaneous* satisfaction of *all* inequalities $f_i(x,q)\leq 0$, $i=1,\dots,m$.

It is not difficult to see that (scalar) kinship functions can still be used to construct relaxations for the above problem. In particular, one may assign an OPKF for each constraint, to construct a convex relaxation of the following form,

$$\begin{aligned}\min_{x\in\mathcal{X}} \quad & c^\top x \\ \text{subject to} \quad & V_\kappa^{(i)}(x)\leq\epsilon_i, \quad i=1,2,\dots,m,\end{aligned}$$

where

$$V_\kappa^{(i)}(x)\doteq\int_{\mathcal{Q}}\kappa[f_i(x,q)]\mu(q)\mathrm{d}q. \tag{3.28}$$

Then, the probability of simultaneous satisfaction of all the constraints is no greater than $\sum_{i=1}^n\epsilon_i$. Alternatively, one may directly consider a single constraint formed by the sum of

the integrals, i.e.,

$$\sum_{i=1}^{m} V_{\kappa}^{(i)}(x) \leq \epsilon. \tag{3.29}$$

Though these approaches provide an upper bound of the probability of violation, they can be, in general, very conservative. Indeed, a set of scalar *optimal* kinship functions is by no means "optimal" for the aggregate problem. This fact motivates us to introduce the concept of the *vector*-kinship function.

Definition 3.18 (Vector-kinship function) *A function $\kappa : [-1, 1]^m \to \mathbb{R}$ is said to be a* vector-kinship function *if*

(a) $\kappa(y)$ is convex for $y \in [-1, 1]^m$,

(b) $\kappa(y)$ is non-decreasing in each component y_i when all other components y_j, $j \neq i$, are fixed,

(c) $\kappa(y) \geq 0$ for $y \in [-1, 0]^m$ and $\kappa(y) \geq 1$ otherwise.

It should be emphasized that unlike the scalar kinship function, the vector-KF requires its domain to be compact. Hence, the constraints f_is in the above problem are assumed to be bounded both from below and from above. In other words, without loss of generality, we assume $f_i : \mathcal{X} \times \mathcal{Q} \to [-1, 1]$. Consequently, defining the integral quantity

$$V_{\kappa}(x) \doteq \int_{\mathcal{Q}} \kappa[f_1(x, q), \ldots, f_m(x, q)]\, \mu(q)\, \mathrm{d}q, \tag{3.30}$$

leads to results similar to the ones stated in Theorem 3.5 and Theorem 3.6; i.e., $V_{\kappa}(x)$ is a convex upper bound on the probability of violation of the aggregate constraints.

Theorem 3.19 *Let $\kappa(\cdot)$ be a vector-kinship function, then,*

$$\mathit{Prob}\,\{q \in \mathcal{Q} \ : \exists i \text{ such that } f_i(x, q) > 0\} \leq V_{\kappa}(x).$$

Moreover, the relaxed optimization problem (3.2) *with $V_k(x)$ defined in (3.30) is convex in x.*

Proof. The proof of the first part is similar to that of Theorem 3.5, and thus is omitted. Moreover, since $f_i(x, q)$ is convex in x for any i as assumed, and the vector kinship function $\kappa(\cdot)$ is convex and non-decreasing in each component, the composition $\kappa[f_1, \ldots, f_m]$ is convex; see e.g. [69]. Hence, $V_k(x)$, the non-negative weighted integral of $\kappa[f_1, \ldots, f_m]$, is convex. □

The above theorem guarantees that the optimization problem constructed using vector kinship functions is a convex relaxation of the original problem. Next, analogous to the developments of Section 3.5, we proceed by defining the optimal polynomial vector-kinship function (OPVKF) as follows.

Definition 3.20 (Optimal polynomial vector-KF) *The optimal polynomial vector-kinship function of total degree ϱ is defined as the solution of the following optimization problem:*

$$\min_{p \in \mathbb{P}_\varrho} \int_{[-1,0]^m} \kappa(y) dy \tag{3.31}$$

$$\textit{subject to} \qquad p \textit{ is a vector-KF.}$$

Similar to Lemma 3.9, the optimal polynomial vector-KF of total degree ϱ can be written as the solution of an optimization problem, as shown next.

Lemma 3.21 *A polynomial vector-KF of total degree ϱ is optimal if it is the solution of the following problem.*

$$\min_{\kappa \in \mathbb{P}_\varrho{}^m} \int_{[-1,0]^m} \kappa(y) dy \tag{3.32}$$

$$\begin{aligned}
\textit{subject to} \qquad & \nabla^2 \kappa(y) \succeq 0 \textit{ for } y \in [-1,1]^m, && (3.33)\\
& \frac{\partial}{\partial y_i} \kappa(y) \geq 0 \textit{ for } y \in [-1,1]^m, i = 1, \ldots, m, && (3.34)\\
& \kappa(-1, \ldots, -1) = 0, && (3.35)\\
& \kappa(-e_i) = 1, i = 1, \ldots, m, && (3.36)
\end{aligned}$$

where e_i is the i-th column of the identity matrix. □

We briefly comment on the constraints in the above problem: condition (3.33) requires the Hessian matrix being positive semi-definite over the hyper-cube, while condition (3.34) amounts to m polynomials being positive over the hyper-cube. Both conditions are semi-infinite, and need to be converted to tractable ones to efficiently solve the optimization problem.

To this end, we recall that positive multi-dimensional polynomials can be *relaxed* as sum-of-squares; see e.g. [54]. In particular, condition (3.34) can be relaxed by using LMI constraints. Moreover, recent research shows that similar techniques can be used to represent positive definite polynomial matrices; see [63] for details. For reader's convenience, in the next lemma we restate a result specific to our case.

Lemma 3.22 *If the Hessian matrix $\nabla^2\kappa(y)$ is positive definite on $[-1,1]^m$, there exist $\epsilon > 0$ and SOS polynomial matrices $S_0(y), S_1(y), \ldots, S_m(y)$ of dimension $m \times m$ such that*

$$\nabla^2\kappa(y) = S_0(y) + \sum_{i=1}^{m} S_i(y)(1 - y_i^2) + \epsilon I.$$

□

Once the degree of each element of the polynomial matrices $S_0(y), \ldots, S_m(y)$ is fixed, one can construct LMIs by comparing the coefficients of both sides according to the above lemma; see [63] for details. Therefore, one can construct a hierarchy of relaxations, which are standard SDP problems, for approximately solving problem (3.32). The exact formulation is omitted for brevity.

We choose not to dwell into cumbersome implementation details and only give a brief description of main features of the vector-KF approach. To this end, first note that, even though the computed polynomials are in general sub-optimal, they can be improved by increasing the degree of the polynomial matrices $S_0(y), \ldots, S_m(y)$. Once again, we stress that these polynomial vector-KFs are computed offline and *once and for all*. Finally, once

a polynomial vector-KF of a given total degree is computed, the numerical approach introduced in Section 3.7 can be applied to solve problem $(PO\kappa_\varrho)$ without any substantial modification.

3.9 Numerical Examples

3.9.1 A Robust LP Problem

We first revisit the motivating example introduced in Section 3.2. To formulate the problem in the form of (3.18), let us define

$$f_i(x,q) \doteq \mathcal{A}_i x + \mathcal{B}_i, \quad i = 1,\dots,m,$$

with $\mathcal{A}_i$, $\mathcal{B}_i$ being the i-th row of the matrices $\mathcal{A}$, $\mathcal{B}$ defined in (3.4). Then, to tackle this problem, an optimal kinship function (OPKF) of given degree ϱ is used to construct a relaxation in the form of $(PO\kappa_\varrho)$, i.e. the integral

$$V^{(i)}_{\kappa_\varrho}(x) \doteq \int_{\mathcal{Q}} \kappa_\varrho[f_i(x,q)]\mathrm{d}q$$

is adopted to formulate the relaxation.

In the numerical simulation, the degree of the OPKF is chosen as $\varrho = 8$. Note that $A(q)$ and $B(q)$ are polynomial in q with total degree no more than 3. Thus, by Lemma A.3, the integral $V_{\kappa_\varrho}(x)$ can be evaluated exactly by using Smolyak formula with $\ell = 12$ and $d = 4$. The number of the nodes used in the formula is 8,705. Moreover, the feasible set $\mathcal{X}$ is assumed to be an ellipsoid with shape matrix $W = 10^3 I_{3\times 3}$, i.e., $\mathcal{X} \doteq \{x : x^T W^{-1} x \le 1\}$. Since f_i are linear in x, the lower bounds on f_i can be computed using the polynomial optimization toolbox provided in [80]. Then, they were appropriately scaled so as to be lower bounded by -1. The bound ($\epsilon = 0.01$) is set on the sum of all $V^{(i)}_{\kappa_\varrho}(x)$ as in (3.29). This guarantees that, as stated in Section 3.8, the probability of the constraints being satisfied is greater than or equal to $1-\epsilon$.

Then, a standard ellipsoid algorithm was chosen to solve the convex relaxation; the

interested reader is referred to [69] for a detailed description of this algorithm. The algorithm was run with initial condition $x_0 = [0 \quad 0 \quad 0]^\top$ and initial shape matrix W, and returned the following numerical solution

$$x^* = \begin{bmatrix} 6.2922 & 20.4695 & 9.7037 \end{bmatrix}^\top.$$

Finally, 1,000,000 uniformly distributed random samples were generated in the uncertainty set $\mathcal{Q}$. We found that none of them violates the linear constraints.

3.9.2 A One-period Portfolio Problem

In this section, we consider an application of the results of this chapter in the area of finance. More precisely, we consider a one-period portfolio problem and assume that there are n investment options. Denote by x_i the amount of budget initially allocated to the i-th option. Also, let r_i be the corresponding return. The goal is, roughly speaking, to minimize the risk while maximizing the total return $x^T r$.

In the example, we consider the Value at Risk (VaR) approach to this problem, and choose the factor model for the returns. Then, for a given *risk level* $\eta \in (0, 1)$, the portfolio selection problem can be stated as follows; see e.g. [81, 56]:

$$\begin{aligned} \min_{\gamma, x \in \mathcal{X}} \quad & \gamma \\ \text{subject to} \quad & k(\eta)\sqrt{x^T(A\Sigma A^T + D)x} - x^T A\hat{\phi} \leq \gamma. \end{aligned}$$

The variable γ to be optimized is usually referred to as the *loss level*, while $k(\eta)$ is a given function of η. The parameters $\hat{\phi}$, Σ and D are computed from real market data. In this example, the sensitivity matrix A is assumed to be uncertain of the form

$$A(q) = A_0 + q_m \left(\sum_i q_{i,1} A_{i,1} + \sum_{i,j} q_{i,j} q_{i,1} A_{i,j} \right)$$

where $q \in \mathcal{Q} \doteq \{q : \|q\|_\infty \leq 1\}$, q_m is the magnitude of uncertainty and A_{ij} has the i,j-th

element being $[A_0]_{i,j}$ and all others being zero. The allowable investment policy set $\mathcal{X}$ is defined as

$$\mathcal{X} := \left\{ x \Big| \sum_{i=1}^{n} x_i = 1, x_i \geq 0, i = 1, 2, ..., n \right\}.$$

The Convex Feasibility Problem

Note that $h(x, q)$ is not polynomial in q. However, if one fixes the loss level γ, and aims at finding an investment policy x such that

$$k(\eta)\sqrt{x^T(A(q)\Sigma A(q)^T + D)x} - x^T A(q)\hat{\phi} \leq \gamma$$

for all uncertainties $q \in \mathcal{Q}$, then our problem is equivalent to finding x such that

$$f(x, q) \leq 0 \quad \text{for all } q \in \mathcal{Q}$$

where

$$\begin{aligned} f(x, q) = x^T \left(A(q) \left(k^2(\eta)\Sigma - \hat{\phi}\hat{\phi}^T \right) A(q)^T + k^2(\eta) D \right) x \\ -2\gamma x^T A(q)\hat{\phi} - \gamma^2, \end{aligned}$$

One can see that $f(x, q)$ is indeed polynomial in q. Moreover, by assuming that

$$k^2(\eta)\Sigma - \hat{\phi}\hat{\phi}^T \succ 0 \tag{3.37}$$

we obtain a function $f(x, q)$ that is also convex in x. Hence, the problem addressed in this example fits the framework described in this chapter. It is worth pointing out that assumption (3.37) is indeed satisfied by real market data for low risk levels; e.g., see Table 2 in [82] for $\eta = 5\%$. Finally, we note that a lower bound of the function $f(x, q)$ can be easily calculated.

$$\begin{aligned} f(x, q) &\geq -2\gamma x^T A(q)\hat{\phi} - \gamma^2 \\ &\geq -2\gamma \max_q \{\|A(q)\hat{\phi}\|_\infty\} - \gamma^2 \end{aligned}$$

Hence, one can always pick a constant $\alpha > 0$ such that $\alpha f(x, q) \geq -1$ for all x and q.

Numerical results

As a numerical example, we considered a portfolio problem with the number of assets $n = 3$ and number of factors $m = 2$. The risk level was chosen to be $\eta = 5\%$. The nominal sensitivity matrix A_0, the covariance matrices Σ and D, and the expect value of the factors $\hat{\phi}$ were randomly generated.

$$A_0 = \begin{pmatrix} 0.4666 & -0.6952 \\ 0.2447 & -0.5934 \\ 0.9796 & 0.6386 \end{pmatrix},$$

$$D = \text{diag}\{\begin{pmatrix} 0.1902 & 0.5995 & 0.2923 \end{pmatrix}^T\},$$

$$\Sigma = \begin{pmatrix} 0.2009 & 0.1791 \\ 0.1791 & 0.4489 \end{pmatrix}, \quad \hat{\phi} = \begin{pmatrix} 0.0584 \\ 0.5385 \end{pmatrix}.$$

In this example, it is assumed that q is uniformly distributed over $\mathcal{Q}$. First, we fixed the uncertainty magnitude $q_{max} = 0.05$, and looked for an investment x (if any) such that the VaR is less than a given loss level γ. This is relaxed, in the proposed framework, to find a policy x such that $\text{Prob}\{f(x, q) > 0\} \leq \epsilon$. The optimal kinship function was chosen to be of order $\varrho = 3$, which led to a total degree $\nu = 12$. Then the integral $\int_{\mathcal{Q}} f(x, q)dq$ was evaluated using Smolyak rules with $l = 6$ and $d = nm = 6$. The number of the nodes in the formula is 4,161. The simulation was run for values of γ increasing from 1.8 to 2.45. Figure 3.3 shows the upper bound on the probability of constraint violation as a function of γ.

At the point $\gamma = 2.45$, the best policy is found to be $x = (0.4545, 0.1866, 0.3589)^T$, which leads to the upper bound $V^*_{\varrho 3} = 0.05$. By sampling 1,000,000 points over the set $\mathcal{Q}$, we found actually that none of them violates the inequality $f(x, q) \leq 0$.

Next, we fixed the probability threshold at $\epsilon = 1\%$, and increased the uncertainty magnitude q_m, to find the minimal loss level γ that ensures $\text{Prob}\{\text{VaR} > \gamma\} \leq \epsilon$; see Figure 3.4.

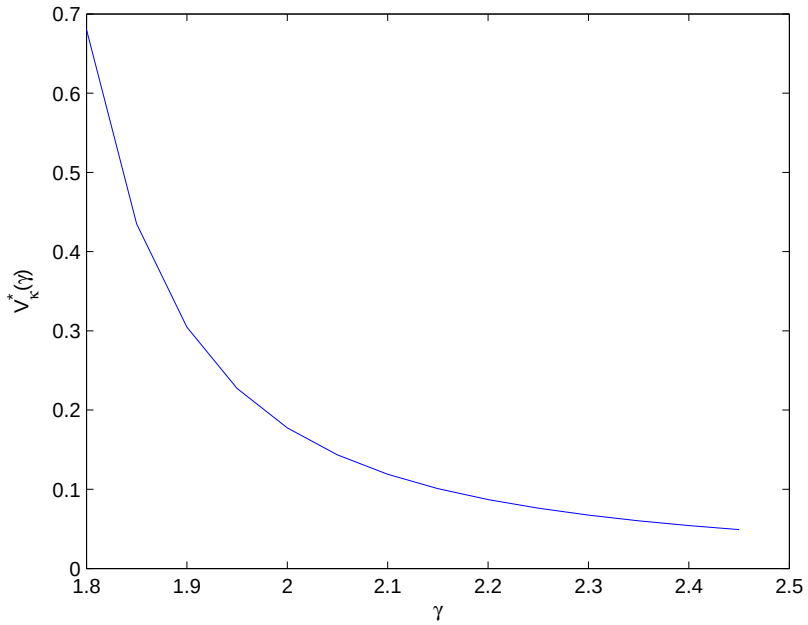

Figure 3.3: Numerical results for the one-period portfolio problem: $V^*_{\kappa_3}(\gamma)$ v.s. γ for $q_{max} = 0.05$.

3.10 Concluding Remarks

In this chapter, a novel approach to robust optimization problems with polynomial dependence on uncertainty has been presented. To solve this problem, we take a probabilistic viewpoint similar to the one considered in the area of probabilistic robustness; i.e, we do not require a solution that "works" for all possible values of uncertainty, but we allow a pre-specified risk of failure. However, in contrast to most of the results in this area, the algorithm presented here is completely deterministic.

To accomplish this, we introduce the concept of kinship function, and use it to obtain deterministic convex upper bounds on the probability of constraint violation. This concept is further refined and the so-called *optimal* polynomial kinship functions are introduced. These optimal functions, which can be computed off line and are available for the end users, provide the "best convex upper bound" on the probability of constraint violation by integrating on the uncertainty set among all polynomial kinship functions of a given degree.

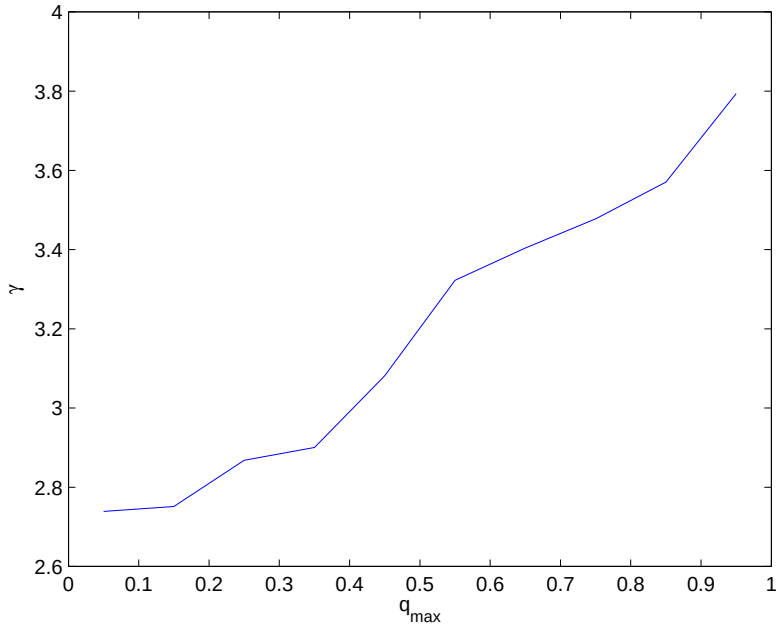

Figure 3.4: Numerical results for the one-period portfolio problem: The loss level γ v.s. the uncertainty magnitude q_{max}.

Moreover, the integrals involved in solving the resulting convex optimization problem can be efficiently computed using quadrature formulae. It is shown that computational complexity increases polynomially with the dimension of the uncertainty. Another important property of this approach is that, the solutions obtained are shown to converge to the solution of the original robust optimization problem as the degree of the optimal kinship function increases.

In further research, it is of interest to develop *matrix* kinship functions specifically tailored to the case where the robust constraints are robust Linear Matrix Inequalities with polynomial dependence on the uncertainty. Also, the polynomial base used in this chapter can lead to numerical difficulty if the degree of the desired kinship function is large. Thus, it is also of interest to further develop similar theory and to improve the computational efficiency by using different polynomial bases.

In the next chapter, we will discuss the so-called "distributional robust" problem. Unlike the probabilistic framework discussed in this chapter, where a PDF is presumed on the uncertainty, in distributional robust framework one finds the *worst-case* distribution that

leads to the worst-case expected value of a cost function.

Chapter 4

Distributional Robustness

In this chapter, we mainly address the following question: what is the the *worst-case* expected value of a continuous function (worst-case performance) over a class of admissible distributions. While it is difficult to answer in general for all continuous functions, a specific set of semi-algebraic functions including polynomials is considered. The first part of the chapter provides a sufficient condition for the *uniformity principle* to hold, i.e., the worst-case expected value is achieved with the uniform distribution. In cases where this principle does not hold, it is shown that to answer this question, for the class of distributions considered, it suffice to solve a convex optimization problem for which efficient linear matrix inequality (LMI) relaxations are available. Finally, the proposed approach is applied to estimate *hard bounds* of the worst-case probability of a semi-algebraic function being negative. Numerical examples are then presented which illustrate the applications of the results presented.

4.1 Introduction

During last two decades, the study of probabilistic approaches for system analysis and design under uncertainty has attracted considerable attention. From a robustness perspective, there are various motivations for the study of this topic, ranging from philosophical to computational; e.g., see [15, 16]. Numerous results are available in the literature, which

can be roughly classified into two categories. In the first category, the so-called probabilistic approaches are based on sampling the uncertainty with known distributions. They are also known as scenario approaches; e.g., see [21, 20, 19]. The basic idea is using the random generated uncertainty samples to "approximate" the robust conditions. In robustness analysis, a system with the nominal model and a uncertainty sample is evaluated and the overall performance on robustness is estimated by evaluating such systems using all the samples. In robust design problems, each uncertainty sample is substituted into the robust condition to generate a deterministic condition. Then the robust design is approximated by a normal design problem with a number of deterministic constraints. On the other hand, there is a different category of approaches to probabilistic robustness analysis. This is usually referred to as *distributional-free* approaches; e.g., see [23, 22, 24, 25, 26]. More specifically, in this approach, the uncertainty parameters are assumed to be random without detailed statistical description. The only assumption made is on the "shape" of the probability density functions (PDFs). For example, in [22] the PDFs are assumed to be symmetric and non-increasing with respect to the absolute value of the parameters. In this set-up, the questions motivated us to answer are: i) what is the *worst-case* distribution of the uncertainty parameters in the pre-defined class that leads to the worst-case expected performance of the system? and ii) how to determine the worst-case expected value?

In this chapter, we follow the second line of the study on probabilistic robustness analysis and consider the problem of determining the worst-case expected value of a nonlinear performance function in a distribution-free manner. The framework adopted in this chapter was first introduced in [22]. Having this framework as a starting point, here we aim at solving the following worst-case expected performance problem

$$\sup_{f \in \mathcal{F}} \quad \mathcal{E}_f[g(q)], \tag{4.1}$$

where the vector of uncertainty parameters $q \doteq (q_1, \ldots, q_d)$ is random with probability density function $f(q)$ belonging to a given distribution class $\mathcal{F}$. In addition, the distribution class $\mathcal{F}$ is assumed to be independent, symmetric and non-decreasing with respect to the absolute value of q_i. However, the results can be applied to other classes of distributions; see

the remark at the end of Chapter 4.3. The function $g(q)$ is assumed to be a semi-algebraic function as defined in [83], which is a generalization of polynomial functions, see Chapter 4.2 for detailed information.

The contribution of this chapter lies in three parts. First, we provide a sufficient condition to verify if the uniform distribution is indeed one of the optimal solutions for problem (4.1), i.e.

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)] = \mathcal{E}_{\mathcal{U}}[g(q)],$$

where $\mathcal{U}$ denotes the uniform distribution. The above equation is also called *uniformity principle* as introduced in [22]. Second, we show that problem (4.1) can be recast as a infinite dimensional convex optimization problem. In addition, a hierarchy of converging linear matrix inequality (LMI) relaxations is provided. Finally, by introducing the concept of *bounding approximations* of indicator functions, we show that it is possible to compute hard bounds on the probability of a semi-algebraic function being negative.

To this end, we first precisely define the admissible distribution class $\mathcal{F}$, which was first introduced in [22].

4.1.1 Admissible Distributions

Consistent with the distribution class discussed in [27, 22, 28, 25], we assume that q_i are independent random variables with bounded support. Without loss of generality, it is assumed that each q_i is supported on $[-1, 1]$. Hence, the parameter vector q is supported on the hypercube $\mathcal{Q} \doteq [-1, 1]^d$. Moreover, for each underlying probability density function f_i, only minimal a priori information is assumed on its "shape". That is, f_i is unknown but assumed to be symmetric about zero and non-increasing with respect to $|q_i|$. More precisely,

Definition 4.1 *A probability density function f is said to belong to the class $\mathcal{F}$ if*

1. $f(q) = f_1(q_1)f_2(q_2)\dots f_d(q_d)$,
2. $f_i(q_i) = 0$ *for* $q_i \notin [-1, 1]$ *and non-increasing with respect to* $|q_i|$.

$\star$

From a practical point of view, being equipped with an admissible distribution implies that the physical parameters are independent, less likely to have large deviations than small deviations from their nominal values, and equally likely to have positive or negative deviations. Examples of admissible distributions for q_1 are plotted in Figure 4.1.

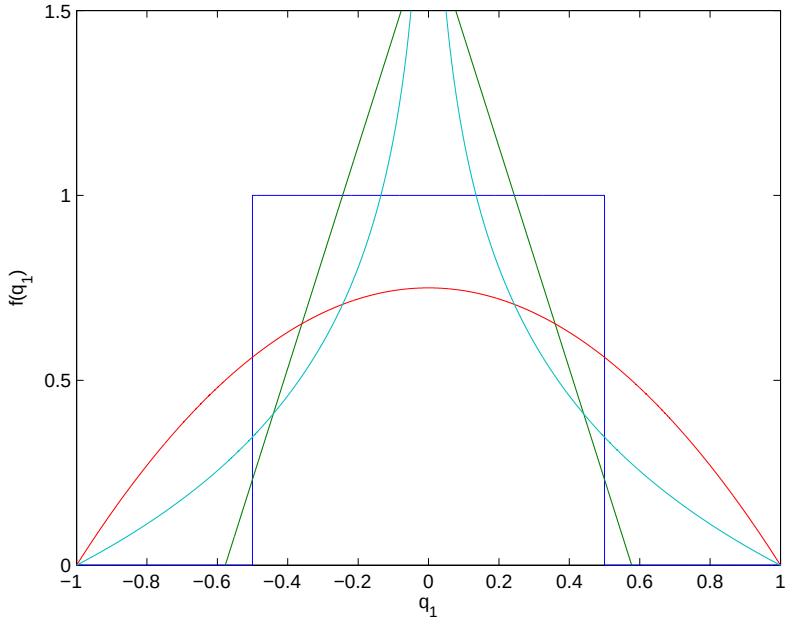

Figure 4.1: Admissible Probability Density Functions

4.1.2 Motivating Example

Consider a RLC amplifier shown in Figure 4.2. The resistors R_1 and R_2 are known with values $R_1 = 30\Omega$, $R_2 = 350\Omega$. The capacitor C and the inductor L are uncertain but known to be varying in the intervals $[0.75 \times 10^{-6}F, 2.05 \times 10^{-6}F]$ and $[0.01H, 0.03H]$, respectively.

Thus, the parameters C and L can be written as

$$\begin{aligned} C &= 1.4 \times 10^{-6} + 6.5 \times 10^{-7} q_1[F]; \\ L &= 0.02 + 0.01 q_2[H], \end{aligned}$$

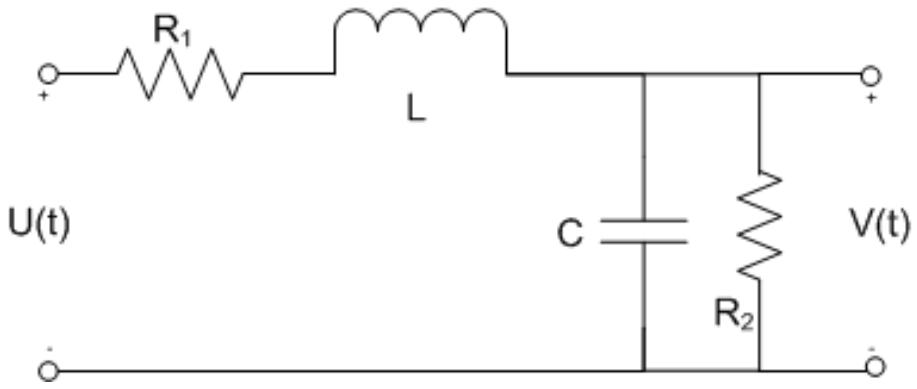

Figure 4.2: Motivating Example: A Simple RLC Amplifier

where the normalized uncertain parameters satisfy $(q_1, q_2) \in [-1, 1] \times [-1, 1]$. With $U(t)$ as input and $V(t)$ as output, the transfer function can be easily calculated and the corresponding magnitude of the frequency response is

$$A_\omega(q_1, q_2) = \frac{R_2}{[(R_1 + R_2 - \omega^2 R_2 LC)^2 + \omega^2 (L + R_1 R_2 C)^2]^{1/2}}. \tag{4.2}$$

Given a frequency ω, we consider the amplification $A_\omega(q_1, q_2)$ as a performance function of parameters (q_1, q_2), and aim at computing its supremal expected value over all admissible distributions,

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[A_\omega(q_1, q_2)].$$

Then there are several questions that need to be addressed. First, does the uniformity principle hold in this problem? In other words, is the uniform distribution the *worst-case* distribution that leads to the supremum of the expected values? Second, if this is not the case, how can we estimate the supremum of the expected values?

4.1.3 Connection to Previous Work

The framework adopted in this chapter was first introduced in [22]. The uniformity principle, which can be regarded as a theoretical justification for use of the uniform distribution in Monte Carlo simulations, is proposed and applied to probabilistic stability analysis of a linear system with parametric uncertainty. The worst-case properties of the uniform dis-

tribution was further developed in [23], showing that the uniform distribution requires the minimum number of samples to attain a specific performance level for a class of Lipschitz continuous functions. In [27], it has been also proven that the worst-case expected value of multi-linearly parameterized H_∞ norm is attained by the uniform distribution. However, it turns out that this principle is not universally valid. For example, in the problem of determining Hurwitz stability of a polynomial, if its coefficients involve nonlinear uncertainty structures, only the so-called *truncated uniformity principle* holds in general. That is, the worst-case distribution is a so-called *truncated uniform distribution*. This truncation phenomenon is considered in [28, 29].

In terms of applications, this framework has been adopted to many robustness/performance analysis problems, especially in circuit networks. In [24], a new Monte Carlo paradigm was proposed to analysis of resistive networks based on *distributional robustness*. More recently, an iterative algorithm was given to compute hard bounds on the probability of performance to circuit analysis, under the assumption that a property oracle exists [26]; and the worst-case probability of stability for polynomials with multi-linear uncertainty is studied in [25].

In this chapter, we aim at finding sufficient conditions that guarantee the uniform principle and, if it is not the case, estimating the worst-case expected performance by using adequate linear matrix inequality (LMI) relaxations. One important property is that these estimates are *hard bounds* of the "true" value. That is, unlike the ones estimated by using Monte Carlo sampling as in [25, 29], which always entail a certain *risk of failure*, a computed hard bound gives a *deterministic* estimate on the probability.

4.2 A Uniformity Principle for Semi-algebraic Functions

In this section, a certificate on the uniformity principle is provided to verify if the uniform distribution on the uncertainty leads to the worst-case expected value of a *semi-algebraic function* that belongs to a specific algebra $\mathcal{A}$.

To begin with, we start with the definition of the algebra$\mathcal{A}$, which is generated by basic monodic operations ($|\cdot|$, $(\cdot)^{1/n}$, $n = 1, 2, \ldots$) and dyadic operations ($+$, $\times$, $/$, $\max(\cdot,\cdot)$,

$\min(\cdot,\cdot)$) on polynomials. More precisely, this class is defined as follows.

Definition 4.2 (A subclass of semi-algebraic functions) *The algebra $\mathcal{A}$ is a class of functions containing all polynomials and whose membership satisfies the following properties,*

a) $g \in \mathcal{A} \rightarrow |g| \in \mathcal{A}$;

b) $g \in \mathcal{A} \rightarrow g^{1/p} \in \mathcal{A}, \quad p = 1, 2, \ldots$;

c) $g_1, g_2 \in \mathcal{A} \rightarrow g_1 + g_2,\ g_1 \cdot g_2,\ g_1/g_2,\ \max\{g_1, g_2\},\ \min\{g_1, g_2\} \in \mathcal{A}$.

$\star$

A key property of functions in $\mathcal{A}$ named *basic semi-algebraic lifting* is defined as the following.

Definition 4.3 (Definition 3.3 in [83]) *A function $g(q) \in \mathcal{A}$ is said to have a basic semi-algebraic lifting (b.s.a.l.) if there exist additional variables $z \in \mathcal{R}^m$ and some polynomials $t_i(q, z)$, $s_j(q, z)$, and a basic semi-algebraic set*

$$\mathcal{K} \doteq \{(q, z) : q \in [-1, 1]^d; t_i(q, z) \geq 0, i = 1, \ldots, n_g; s_j(q, z) = 0, j = 1, \ldots, n_e\}$$

such that the graph of g (denoted by Ψ_g) satisfies

$$\Psi_g \doteq \{(q, g(q)) : q \in [-1, 1]^d\} = \{(q, z_m) : (q, z) \in \mathcal{K}\}. \tag{4.3}$$

$\star$

As illustrated in [83], every well-defined function $g \in \mathcal{A}$ has a basic semi-algebraic lifting.

Lemma 4.4 (Lemma 3 in [83]) *Every semi-algebraic function $g(q) \in \mathcal{A}$ where $q \in [-1, 1]^d$ has a basic semi-algebraic lifting.* $\square$

Though the definition of $\mathcal{A}$ is complicate, many problems in system and control can be formulated by using some function belongs to $\mathcal{A}$. One immediate example is the frequency response function $A_\omega(q)$ defined in (4.2). Also, we use this example to show how to construct b.s.a.l. for this magnitude function. Let

$$\begin{aligned} z_1 &= [(R_1 + R_2 - \omega^2 R_2 LC)^2 + \omega^2 (L + R_1 R_2 C)^2]^{1/2}, \\ z_2 &= z_1^{-1}, \end{aligned}$$

and define polynomial functions (note R and C are linear in q_1 and q_2, respectively)

$$\begin{aligned} t_1(q,z) &\doteq z_1, \\ s_1(q,z) &\doteq z_1^2 - (R_1 + R_2 - \omega^2 R_2 LC)^2 - \omega^2 (L + R_1 R_2 C)^2, \\ s_2(q,z) &\doteq z_1 z_2 - R_2. \end{aligned}$$

Then, $A_\omega(q)$ has a b.s.a.l.

$$\Psi_A = \{(q, z_2) : (q,z) \in \mathcal{K}\}$$

where

$$\mathcal{K} \doteq \{(q,z) : q \in [-1,1]^2; t_1(q,z) \geq 0; s_1(q,z) = 0; s_2(q,z) = 0\}.$$

Now let us consider the following question: given a set of parametrized semi-algebraic functions, does the Uniformity Principle holds for all of them; i.e., is the uniform distribution the worst-case distribution for all elements in the set.

This question is, in our view, important since if one establishes that the Uniformity Principle indeed holds for all members of the set, then for any member in this (possibly large) set one can estimate the worst-case expected performance just by using a Monte Carlo method where the samples are generated using a uniform distribution.

We now provide a precise definition of the Uniformity Principle.

Definition 4.5 (Uniformity Principle) *Let g be a given function belongs to $\mathcal{A}$, where $\mathcal{A}$*

is the algebra defined in Definition 4.2, the uniformity principle holds for Problem (4.1) *if*

$$\sup_{f\in\mathcal{F}} \mathcal{E}_f[g(q)] = \mathcal{E}_{\mathcal{U}}[g(q)]. \tag{4.4}$$

$\star$

Consider a class of semi-algebraic functions

$$\mathcal{S} \doteq \{g(q) : g(q) = \tilde{g}(q,\delta) \text{ for some } \|\delta_i\| \le \epsilon_i, i = 1,\ldots,l\} \tag{4.5}$$

where $\tilde{g}(q,\delta)$ is a semi-algebraic function and belongs to $\mathcal{A}$ with respect to (q,δ). In the exposition to follow, sufficient conditions are provided for the Uniformity Principle to hold for class $\mathcal{S}$. As a firs step, define two sets of semi-algebraic functions

$$\tau_k(x,q,\delta) \doteq \tilde{g}(q_1,\ldots,q_{k-1},x,q_{k+1},\ldots,q_d,\delta) + \tilde{g}(q_1,\ldots,q_{k-1},-x,q_{k+1},\delta), \tag{4.6}$$

$$\omega_k(x_1,x_2,q,\delta) \doteq \tau_k(x_1,q,\delta) - \tau_k(x_2,q,\delta), \quad k = 1,\ldots,d. \tag{4.7}$$

Consequently, $\omega_k(x_1,x_2,q,\delta)$ belongs to the algebra $\mathcal{A}$ for all $k = 1,\ldots,d$. Hence, followed by Lemma 4.4, $\omega_k(x_1,x_2,q,\delta)$ has a b.a.s.l., i.e.

Lemma 4.6 For any $1 \le k \le d$, $\omega_k(x_1,x_2,q,\delta)$ has a basic semi-algebraic lifting

$$\Psi_{\omega_k} = \{(q,\delta,x_1,x_2,z) : (q,x_1,x_2) \in \mathcal{K}_{\omega_k}\} \tag{4.8}$$

where

$$
\begin{aligned}
z &= \omega_k(q, \delta, x_1, x_2), \\
\mathcal{K}_{\omega_k} &\doteq \left\{ (q, \delta, x_1, x_2, z): \begin{array}{l} q \in [-1, 1]^d; \\ 0 \le x_1 \le x_2 \le 1; \\ \delta_i \in [-\epsilon_i, \epsilon_i], i = 1, \ldots, l; \\ t_j(q, \delta, x_1, x_2, z) \ge 0, j = 1, \ldots, n_g; \\ s_k(q, \delta, x_1, x_2, z) = 0, k = 1, \ldots, n_e. \end{array} \right\},
\end{aligned}
$$

and t_j, s_k are polynomial functions that define the polynomial dependency of z and (q, δ, x_1, x_2) as illustrated in Definition 4.3. □

It is shown that the positivity of $\omega_k(q, \delta, x_1, x_2)$, $k = 1, \cdots, d$, provides a certificate for the uniformity principle on any member g in the class $\mathcal{S}$.

Theorem 4.7 Given a class of semi-algebraic function $\tilde{g}(q, \delta) \in \mathcal{A}$, the corresponding function $\omega_k(q, \delta, x_1, x_2)$ defined in (4.7) and its b.s.a.l. Ψ_{ω_k} defined in (4.8), consider the following polynomial optimization problem

$$
\begin{aligned}
\Omega_k^* \doteq \inf_{q, \delta, x_1, x_2, z} \quad & z \\
\text{s.t.} \quad & t_i(q, \delta, x_1, x_2, z) \ge 0, i = 1, \ldots, n_g, \\
& s_i(q, \delta, x_1, x_2, , z) = 0, i = 1, \ldots, n_e, \\
& v_i \doteq 1 - q_i^2 \ge 0, i = 1, \ldots, d, \\
& v_{d+1} \doteq x_1 \ge 0, \\
& v_{d+2} \doteq x_2 - x_1 \ge 0, \\
& v_{d+3} \doteq 1 - x_2 \ge 0, \\
& v_{d+3+i} \doteq \epsilon_i^2 - \delta_i^2, i = 1, \ldots, l.
\end{aligned} \tag{4.9}
$$

Then, if

$$\Omega_k^* \geq 0, \quad \text{for all } 1 \leq k \leq d, \tag{4.10}$$

the uniformity principle holds for all $g(q) \in \tilde{g}(q)$.

Proof: See Appendix B.1. □

Solving polynomial optimization problem (4.9) exactly is difficult in general. However, checking positivity of Ω_k^* can be recast as checking if a set is semidefinite programs has a positive solution, as stated in the next section.

4.2.1 LMI Relaxation

As introduced in Chapter 2, one of the most efficient methods to solve the polynomial optimization problems above is the so-called moment-based approach. In the approach, the polynomial optimization problem is lifted and converted to a convex program by using the characterization of the moment set of all possible probability measures supported on a compact semi-algebraic set. Therefore, the N-truncated moment matrices introduced in Chapter 2.2.1 are used to formulate LMI relaxations for the polynomial optimization problem (4.9). It also has been proven that the solutions of the relaxations monotonically converge to the desired value as N increases to infinity.

To sum up, for a fixed value of N, the corresponding LMI relaxation for numerically solving problem (4.9) is given as,

$$\Omega_{k,N}^* \doteq \max \quad m_{0\cdots01} \tag{4.11}$$

$$\begin{aligned} \text{s.t.} \quad & M_N(\mathbf{m}) \succeq 0; \\ & M_{N_{g,k}}(t_{k,i}, \mathbf{m}) \succeq 0, i = 1, \ldots, n_g, \\ & M_{N_{e,k}}(s_{k,i}, \mathbf{m}) = 0, i = 1, \ldots, n_e, \\ & M_{N-1}(v_{k,i}, \mathbf{m}) \succeq 0, i = 1, \ldots, d + l + 3, \end{aligned} \tag{4.12}$$

where $t_{k,i}$, $s_{k,i}$ and $v_{k,i}$ are the constraint polynomials defined in problem (4.9) for ω_k. Hence, the following result is a direct consequence of Theorem 4.2 in [7].

Theorem 4.8 *There exists at least one positive solution to* (4.9) *if and only if there exists a relaxation order* N_k *such that*

$$\Omega^*_{k,N} \geq 0,$$

for all $1 \leq k \leq d$. □

Remark 4.9 *Once the uniformity principle is proven to be held on a given semi-algebraic function class, Monte Claro methods can be used to estimate the worst-case expected value of any function that belongs to the class. In other words, one can generate uniformly distributed samples over* $q \in [-1, 1]^d$*, and evaluate the function by using these samples directly. Then, the average of all evaluated values can be regarded as an estimate of the worst-case expected value of the function.*

4.3 Estimation on Worst-case Expected Performance

In the previous section, a sufficient condition is provided that guarantees the uniformity principle on evaluating the worst-case expected value of semi-algebraic function. However, there are many cases in which the worst-case distribution is not the uniform distribution; e.g., see [29, 84]. Therefore, in this section we consider problem (4.1) when the uniformity principle is not valid and show that it can be reformulated as a convex optimization problem. Moreover, LMI relaxations which are computationally tractable are provided. To this end, two cases are addressed in this section. First, the function $g(q)$ is assumed to be a semi-algebraic function such that $g(q) \in \mathcal{A}$. Second, $g(q)$ is considered as a general continuous function and polynomial approximations are used to estimate its worse-case expected value.

4.3.1 Case I: Semi-algebraic Functions

Assume that a function $g(q) \in \mathcal{A}$ is well-defined, and, hence, it always has a basic semi-algebraic lifting as defined in Definition 4.2, i.e.,

$$\Psi_g = \{(q, z_m) : (q, z) \in \mathcal{K}\}, \tag{4.13}$$

where

$$\mathcal{K} = \{(q, z) : q \in [-1, 1]^d; z \in \mathbb{R}^m; t_i(q, z) \geq 0, i = 1, \ldots, n_g; s_j(q, z) = 0, j = 1, \ldots, n_e\}.$$

Moreover, we define the moment set

$$\mathcal{M}_q^{\mathcal{U}_t} \doteq \left\{\{m_i\} : m_i = \mathcal{E}_f[q^i] \text{ for some } f(q) \in \mathcal{U}_t\right\},$$

and the projection operator Γ_q that projects the moment sequence of the variables (q, z) to the space of the moment of q only. Then, it can be proven that,

Lemma 4.10 *Given any semi-algebraic function $g(q) \in \mathcal{A}$, it follows that*

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{\Gamma_q m \in conv(\mathcal{M}_q^{\mathcal{U}_t}), m \in \mathcal{M}_\mathcal{K}} m_{0\cdots 1}.$$

Proof: See Appendix B.2. □

Remark 4.11 *It is worth pointing out that the above lemma holds also for other classes of distributions. That is, to compute the worst-case expected value for a semi-algebraic function over any distribution class, it always suffices to consider the convex hull of its moment set.*

Therefore, to solve problem (4.1), it suffices to characterize the set $\mathcal{M}^{\mathcal{U}_t}$ and the moment set $\mathcal{M}_\mathcal{K}$. For the second one, as elaborated in Chapter 2.2.1, $\mathcal{M}_\mathcal{K}$ can be approximated using LMIs consisting of moment matrices. The only problem left now is to characterize the moment set of all truncated uniform distributions, $\mathcal{M}^{\mathcal{U}_t}$. In fact, it can be proven

that $\text{conv}(\mathcal{M}^{\mathcal{U}_T})$ is a linear transformation of the known Hausdorff moment set $\mathcal{M}_{[-1,1]^d}$, as shown in Lemma 4.10 in Appendix B.3.

To sum up, we state the main theorem of this section as the following.

Theorem 4.12 Assume that the semi-algebraic performance function $g(q)$ has a basic semi-algebraic lifting Ψ_g as defined in (4.13). Then,

$$\sup_{f\in\mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{m\in\mathcal{M}_{\mathcal{K}},\Gamma m\in\Pi\mathcal{M}_{[-1,1]^d}} m_{0\cdots01}, \tag{4.14}$$

where Π is a diagonal matrix with elements

$$\pi_{\mathbf{i},\mathbf{i}} = \begin{cases} \frac{1}{(i_1+1)\cdots(i_d+1)} & \text{, if } i_1,\ldots,i_d \text{ are even;} \\ 0 & \text{, otherwise.} \end{cases} \tag{4.15}$$

Proof: See Appendix B.3. □

To solve this recast optimization problem, it suffice to characterize the moment sets supported on $\mathcal{K}$ and on $[-1,1]^d$. As elaborated in Chapter 4.2.1, the characterization of a moment set supported on a compact semi-algebric set is equivalent to positivity of infinite dimensional moment matrices and localizing matrices. Therefore, given a fixed relaxation order N, the relaxed conditions for $m \in \mathcal{M}_{\mathcal{K}}$ are

$$M_N(m) \succeq 0, \tag{4.16}$$

$$M_{N-1}(p_i, m) \succeq 0, i = 1,\ldots,d, \tag{4.17}$$

$$M_{N-d_{t_i}/2}(t_i, m) \succeq 0, i = 1,\ldots,n_g, \tag{4.18}$$

$$M_{N-d_{s_i}/2}(s_i, m) = 0, i = 1,\ldots,n_e, \tag{4.19}$$

where $p_i(q,z) = 1 - q_i^2$ and $t_i(q,z)$, $s_i(q,z)$ are the polynomials defined in (4.13).

On the other hand, $\Gamma_q m$ is the projection of the moment of (q,z) on the space of the moment of q Therefore, for the same relaxation order N, the relaxed conditions for

$\Gamma_q m \in \Pi\mathcal{M}_\mathcal{K}$ are

$$M_N(\Pi^{-1}\Gamma_q m) \succeq 0, \tag{4.20}$$

$$M_{N-1}(p_i, \Pi^{-1}\Gamma_q m) \succeq 0, i = 1, \ldots, d. \tag{4.21}$$

Hence, the condition that a sequence belongs to the moment sets is equivalent to an infinite linear matrix inequality. Therefore, one can construct a LMI relaxation by considering a finite value of N. In [7], it is proven that the solution of the relaxation monotonically converges to the desired value as N increases to infinity.

To sum up, for a fixed value of N, the corresponding LMI relaxation for numerically solving problem (4.14) is given as,

$$\max \quad m_{0\cdots01} \tag{4.22}$$

$$\text{s. t.} \quad (4.16) - (4.21) \tag{4.23}$$

Hence, the following result is a direct consequence of Theorem 4.2 in [7].

Theorem 4.13 *Let g_N^* be the solution of problem* (4.22)*. Then, as N increases to infinity,*

$$p_N^* \downarrow \sup_{f \in \mathcal{F}} \mathcal{E}_f[p(q)]. \tag{4.24}$$

□

4.3.2 Case II: Continuous Functions

In this section, we consider the case where $g(q)$ is continuous but not necessary being a semi-algebraic function. In this case, we first approximate $g(q)$ by semi-algebraic functions if possible, or, in particular, approximate it by polynomials. Since currently there does not exist any systematic way to approximate a continuous function by semi-algebraic functions in $\mathcal{A}$, we focus on the polynomial approximations in this section. However, once such approximations are available (which may have better accuracy than using polynomial

approximations), the results stated in this section can be directly applied.

Note that any continuous function defined on a compact set can be uniformly approximated by polynomials. There are many algorithms available, most of which amount to the selection of interpolation points [85, 86].

For example, $A_\omega(q_1, q_2)$ defined in (4.2) for $\omega = 7,500Hz$ can be approximated by a polynomial with total degree $\varrho = 5$,

$$\begin{aligned}\tilde{A}_\varrho(q) = 1.18 - 0.67q_1 - 0.80q_2 - 0.063q_1^2 - 0.17q_1q_2 + \\ 0.041q_2^2 + 0.17q_1^3 + 0.66q_1^2q_2 + 0.72q_1q_2^2 + \\ 0.24q_2^3 - 0.014q_1^4 - 0.0051q_1^3q_2 + 0.013q_1^2q_2^2 - \\ 0.084q_1q_2^3 - 0.074q_2^4 + 0.053q_1^5 - 0.10q_1^4q_2 - \\ 0.31q_1^3q_2^2 - 0.356q_1^2q_2^3 - 0.13q_1q_2^4 + 0.052q_2^5,\end{aligned}$$

with the maximal error in the domain

$$\max_{q\in\mathcal{Q}} |A(q) - \tilde{A}_\varrho(q)| \approx 0.0538.$$

The following theorem guarantees that a good estimate of the worst-case expected value of a continuous function can be immediately obtained once the counterpart of its polynomial (semi-algebraic) approximations is evaluated.

Theorem 4.14 *Given a continuous function $g(q)$, where q is a random variable supported on a compact set $\mathcal{Q}$ with some density function $f \in \mathcal{F}$. Then for any $\epsilon > 0$, there always exists a polynomial (semi-algebraic) function $p(q)$ that $\max_{q\in\mathcal{Q}} |p(q) - g(q)| < \epsilon$. Moreover,*

$$\left|\mathcal{E}_f[p(q)] - \mathcal{E}_f[g(q)]\right| < \epsilon,$$

and, hence,

$$\left|\sup_{f\in\mathcal{F}} \mathcal{E}_f[p(q)] - \sup_{f\in\mathcal{F}} \mathcal{E}_f[g(q)]\right| < 3\epsilon. \tag{4.25}$$

Proof: See Appendix B.4 □

Remark 4.15 *According to Theorem 4.14, if the uniformity principle holds for $p(q)$, then its worst-case expected value can be evaluated by a simple Monte Carlo simulation using sufficient many uniform samples of q. If this not true, i.e., the uniformity principle is not valid, one may use Theorem 4.13 and solve the related LMI relaxations to estimate the worst-case expected value of $p(q)$, since polynomials are the simplest functions that belong to $\mathcal{A}$. Hence, the supremal expected value of $g(q)$ can be estimated using* (4.25).

4.4 Hard Bounds for Probabilistic Objective Functions

In this section, we consider a probabilistic version of problem (4.1). That is, we want to estimate the worst-case probability

$$\inf_{f\in\mathcal{F}} \text{Prob}\{h(q) \geq 0\} \tag{4.26}$$

for any given semi-algebraic function $h(q) \in \mathcal{A}$.

It is obvious that, given a density function f,

$$\text{Prob}\{h(q) \geq 0\} = \int_{\mathcal{Q}} I[h(q)]f(q)dq = \mathcal{E}_f[I(h(q))],$$

where $I(y)$ is the indicator function of the set $\{y : y \geq 0\}$. However, the results developed so far in the previous sections cannot be applied to estimate

$$\inf_{f\in\mathcal{F}} \mathcal{E}_f[I(h(q))],$$

since $I(h(q))$ is not a continuous function.

To overcome the difficulty, in line with the results developed in [87], we adopt an algorithm to compute the so-called upper bound and lower bound semi-algebraic approximations for $I(y)$. There are many choices to construct such bounding semi-algebraic functions. In this section, two cases are considered: polynomials and piecewise linear functions. To bound the probability, these approximating functions have to satisfy specific conditions so that the corresponding estimates are greater/less than the "true" probability for any

probability measure f.

Definition 4.16 (Bounding Approximations) Assume α and β are two known positive numbers. A function $I^+(y)$ is called an upper bound approximation of the indicator function $I(y)$ on $[-\alpha, \beta]$ if $I^+(y) \geq 0$ for $y \in [-\alpha, 0]$ and $I^+(y) \geq 1$ for $y \in [0, \beta]$. Similarly, a function $I^-(y)$ is called a lower bound approximation of $I(y)$ if $I^-(y) \leq 0$ for $y \in [-\alpha, 0]$ and $I^-(y) \leq 1$ for $y \in [0, \beta]$. $\star$

4.4.1 Bounding Approximations

Piecewise Linear Bounding Approximations

The objective is to first construct an upper bound approximation $I^+(y)$ on the interval $[-\alpha, \beta]$ where $I^+(y)$ is a piecewise linear function. Indeed, it is not difficult to verify that the function

$$I_\delta^+(y) = \begin{cases} 0, & y \in [-\alpha, -\delta]; \\ 1 + \frac{y}{\delta}, & y \in [-\delta, 0]; \\ 1, & y \in [0, \alpha]. \end{cases},$$

for any fixed $\delta \in [-\alpha, 0)$ is an upper bounding approximation. Moreover, $I_\delta^+(y)$ is a semi-algebraic function that belongs to the algebra $\mathcal{A}$ defined in Definition 4.2. In fact, $I_\delta^+(y)$ can be rewritten as

$$I_\delta^+(y) = \min\{1, \max\{0, 1 + y/\delta\}\}, \tag{4.27}$$

and, hence, $I_\delta^+(y) \in \mathcal{A}$. Similarly, a lower bound approximation $I_\delta^-(y)$ can be constructed as the following,

$$I_\delta^+(y) = \begin{cases} 0, & y \in [-\alpha, 0]; \\ \frac{y}{\delta}, & y \in [0, \delta]; \\ 1, & y \in [\delta, \alpha]. \end{cases},$$

It can also be rewritten as

$$I_\delta^-(y) = \min\{1, \max\{0, y/\delta\}\}, \tag{4.28}$$

and, hence, $I_\delta^-(y) \in \mathcal{A}$. Some examples of possible piecewise linear bounding approximations of indicator function are illustrated in Figure 4.3.

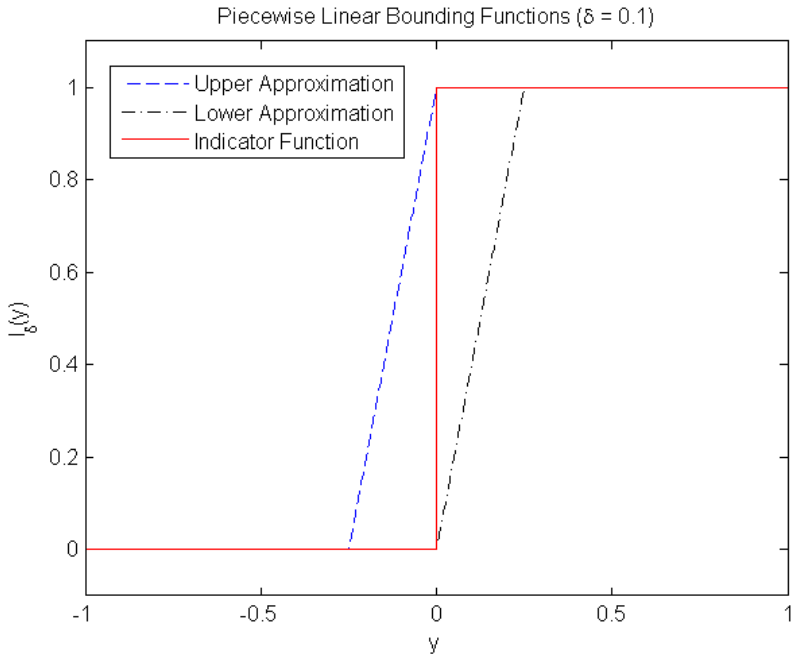

Figure 4.3: Examples of Piecewise Linear Bounding Approximations

Polynomial Bounding Approximations

The "optimal" upper bound polynomial approximation $I_p^+(y)$ on the interval $[-\alpha, \beta]$ is defined as the solution of the following optimization problem,

$$\begin{aligned} \min_{J \in \mathbb{P}_\varrho} \quad & \int_{-\alpha}^{\beta} W(y) J(y) dy \\ s.t. \quad & J(y) \geq 1, y \in [-\alpha, 0]; \\ & J(y) \geq 0, y \in [0, \beta], \end{aligned} \tag{4.29}$$

where $W(y)$ is any selected weighting function satisfying: i) $W(y)$ is positive and ii) $W(y)y^k$ is integrable on $[-\alpha, \beta]$ for $0 \leq k \leq \varrho$. The approximation is "optimal" in the sense that the weighted $\mathcal{L}_1$ norm of its difference to the indicator function, i.e. $\int_{-\alpha}^{\beta} W(y)|J(y) - I(y)|dy$, is minimized. Similarly, the "optimal" lower bound polynomial approximation $I^-(y)$ is

defined as the solution of the following problem,

$$\begin{aligned} \max_{J \in \mathbb{P}_\varrho} \quad & \int_{-\alpha}^{\beta} W(y) J(y) dy \\ s.t. \quad & J(y) \leq 1, y \in [-\alpha, 0] \\ & J(y) \leq 0, y \in [0, \beta] \end{aligned} \tag{4.30}$$

According to [88, 87], the above two problems can be converted to standard semi-definite programs (SDPs). Hence, given interval $[-\alpha, \beta]$, weighting function $W(y)$ and degree ϱ, the "optimal" polynomial approximations can be easily computed by standard SDP solvers. It is worth pointing out that, since $J(y)$ is a univariate polynomial, the above problems can be solved exactly. That is, unlike the LMI relaxations introduced in Chapter 4.2.1 for solving problem (4.14), here the "optimal" polynomial approximations can be always computed exactly. Some examples of possible polynomial bounding approximations of indicator function are illustrated in Figure 4.4.

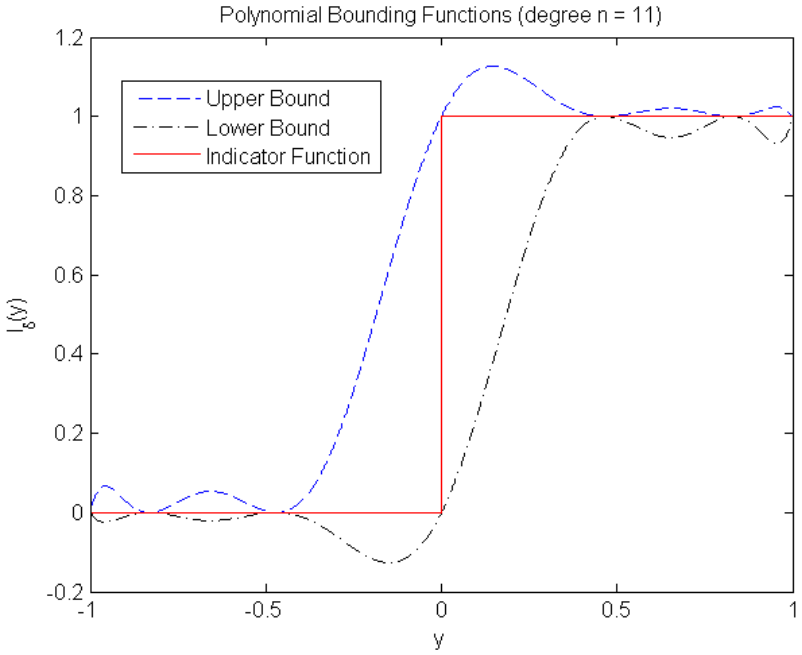

Figure 4.4: Examples of Polynomial Bounding Approximations

4.4.2 Estimation on Hard Bounds

Given bounding approximations $I^+(\cdot)$ and $I^-(\cdot)$ that satisfy Definition 4.16, one can see immediately that

$$\mathcal{E}_f[I^-(h(q))] \leq \text{Prob}\{h(q) \geq 0\} \leq \mathcal{E}_f[I^+(h(q))] \tag{4.31}$$

for any density function f. Let us define the semi-algebraic performance functions

$$\begin{aligned} I_h^+(q) &\doteq I^+[h(q)], \\ I_h^-(q) &\doteq I^-[h(q)]. \end{aligned}$$

Consequently, the following result is obtained.

Corollary 4.17 Given any polynomial $h(q)$,

$$p^- \leq \inf_{f \in \mathcal{F}} \text{Prob}\{h(q) \geq 0\} \leq p^+ \tag{4.32}$$

where

$$p^- \doteq \max\left\{0, \inf_{f \in \mathcal{F}} \mathcal{E}_f I^-[h(q)]\right\},$$

and

$$p^+ \doteq \min\left\{1, \inf_{f \in \mathcal{F}} \mathcal{E}_f I^+[h(q)]\right\}.$$

Moreover, p^- and p^+ can be estimated by solving a hierarchy of converging LMI relaxations.

Proof: Suppose that $\inf_{f \in \mathcal{F}} \text{Prob}\{h(q) \geq 0\}$ takes its infimum at f^*, then according to (4.31),

$$\begin{aligned} \inf_{f \in \mathcal{F}} \text{Prob}\{h(q) \geq 0\} &= \text{Prob}\{h(q) \geq 0\}_{f^*} \\ &\geq \mathcal{E}_{f^*}[I^-(h(q))] \geq \inf_{f \in \mathcal{F}} \mathcal{E}_f[I^-(h(q))]. \end{aligned}$$

On the other hand, suppose that $\inf_{f\in\mathcal{F}} \mathcal{E}_f I^+[h(q)]$ takes its infimum at f^*, then

$$\begin{aligned}\inf_{f\in\mathcal{F}} \mathcal{E}_f I^+[h(q)] &= \mathcal{E}_{f^*} I^+[h(q)] \\ &\geq \text{Prob}\left\{h(q)\geq 0\right\}_{f^*} \geq \inf_{f\in\mathcal{F}} \text{Prob}\left\{h(q)\geq 0\right\}.\end{aligned}$$

We conclude the proof due to the fact that $\text{Prob}\left\{h(q)\geq 0\right\}$ is a non-negative number less than 1. □

4.5 Numerical Examples

In this section, three examples are given that illustrate the approach proposed in this chapter.

4.5.1 Example 1

We first consider the circuit introduced in Chapter 4.1.2. For frequency $\omega = 7,500$ Hz, the corresponding amplification is

$$A(q) = \frac{350}{[a_1(q)^2 + a_2(q)^2]^{1/2}}.$$

where

$$\begin{aligned}a_1(q) &= 171.25 + 255.94q_1 + 275.63q_2 + 127.97q_1q_2, \\ a_2(q) &= 260.25 + 51.19q_1 + 75q_2.\end{aligned}$$

The bivariate function $A(q)$ can be well approximated by the hyper-interpolation at Morrow-Patterson-Xu cubature points for the product Chebyshev measure; see e.g. [86] and references therein. In the numerical computation, $A(q)$ is approximated by a polynomial $\tilde{A}_\varrho(q)$ with total degree $\varrho = 10$ and the corresponding maximal error is

$$\max_{q\in\mathcal{Q}} |A(q) - \tilde{A}_\varrho(q)| \approx 0.004.$$

Using the approach described in Chapter 4.3, the worst-case expected value was computed

$$\sup_{f\in\mathcal{F}} \mathcal{E}_f[\tilde{A}_\varrho(q)] = 1.168,$$

which is attained at the truncation point $t = [0.095, 0.913]$. Hence, according to Theorem 4.14,

$$1.156 \leq \sup_{f\in\mathcal{F}} \mathcal{E}_f[A(q)] \leq 1.18.$$

For comparison, we generated 1,000,000 uniform samples for q on $[-1, 1]^2$ and then carry out a Monte Carlo based optimization to determine the worst-case truncated uniform distribution. The estimate of the maximal expected amplification is

$$\max_{f\in\mathcal{F}} \mathcal{E}_f[A(q)] = \max_{f\in\mathcal{U}_T} \mathcal{E}_f[A(q)] \approx 1.17,$$

which is attained at the truncation point $t = [0.09, 0.92]$. In contrast, the expected amplification for the uniform distribution is

$$\mathcal{E}_{u_{[1,1]}}[A(q)] \approx 1.15.$$

4.5.2 Example 2

To show how to apply the approach proposed in Chapter 4.4 to estimate hard bounds of the worst-case probability, we consider the example in [29]. A two stage amplifier is shown in Figure 4.5. It has fixed parameters $R_1 = 1000\Omega$, $R_2 = 100\Omega$, $L = 0.01H$ and uncertain parameters

$$C_1 \in [1.0796 \times 10^{-6}F, 2.0196 \times 10^{-6}F];$$
$$C_2 \in [1.1135 \times 10^{-6}F, 4.9135 \times 10^{-6}F].$$

To illustrate our approach, we consider the frequency response at $\omega = 10^4$ Hz. The frequency is selected where the response is sensitive to the uncertainty. More precisely,

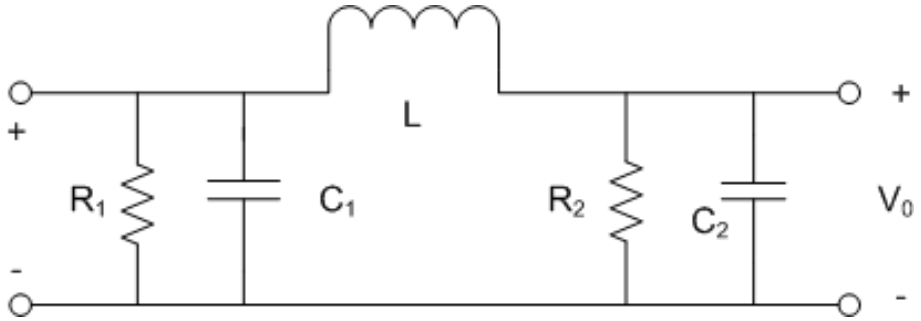

Figure 4.5: A Two Stage Amplifier

given the amplification,

$$\begin{aligned} A_\omega(q) &= \left|\frac{a_\omega(q)}{b_\omega(q)}\right| \\ &= \left|\frac{b_2(q)(j\omega)^2+b_1(q)(j\omega)+a_0(q)}{a_3(q)(j\omega)^3+a_2(q)(j\omega)^2+a_1(q)(j\omega)+a_0(q)}\right|, \end{aligned}$$

where

$$\begin{aligned} a_3(q) &= 10^{-9}(4.6698 + 1.4163q_1 + \\ &\qquad 2.9443q_2 + 0.893q_1q_2); \\ a_2(q) &= 10^{-5}(1.851 + 0.47q_1 + 0.19q_2); \\ a_1(q) &= 0.46631 + 0.047q_1 + 0.19q_2; \\ a_0(q) &= 1100; \\ b_2(q) &= 10^{-3}(3.0135 + 1.9q_2); \\ b_1(q) &= 10; \\ b_0(q) &= 100000, \end{aligned}$$

we want to estimate

$$\min_{f\in\mathcal{F}} \text{Prob}\left\{A_\omega(q) \le 316\right\}$$

for $\omega = 10^4$.

First we apply the approach introduced in Chapter 4.4 to compute hard bounds for the worst-case probability. Note that $A_\omega(q) \le 316$ is equivalent to that

$$h(q) \doteq -b_\omega(q)b_{-\omega}(q) + (316)^2 a_\omega(q)a_{-\omega}(q) \ge 0,$$

where $\omega = 10,000$. Then, the interval $[-\alpha, \beta]$ can be fixed by setting $\alpha = -\min_{q\in\mathcal{Q}} h(q)$ and $\beta = \max_{q\in\mathcal{Q}} h(q)$. Hence, the "optimal" lower bound and upper bound polynomial approximations defined in Chapter 4.4 can be computed once given degree ϱ and weighting function $W(y)$. Then, Corollary 4.17 can be applied to compute hard bounds for the worst-case probability.

The numerical result is listed in Table 4.1 for various values of ϱ. The weighting function is selected as $W(y) \equiv 1$.

ϱ	8	9	10	11	12	13	14
p^+	0.7850	0.7352	0.7302	0.7210	0.7113	0.6889	0.6524
p^-	0.1375	0.1443	0.1648	0.1825	0.2283	0.2614	0.2926

Table 4.1: Upper bounds and lower bounds of the worst-case probability

For comparison, we take a uniform sampling over the whole uncertainty set $\mathcal{Q}$ and use the generated 1,000,000 samples to carry out a Monte Carlo based optimization to determine the worst-case truncated uniform distribution. It is estimated that

$$\inf_{f\in\mathcal{F}} \text{Prob}\{A(q) \leq 316\} \approx 0.5962,$$

where the minimum is achieved at the truncation point $t = [0.5, 0]$. In contrast, the uniform distribution on $[-1, 1]^2$, i.e. $t = [1, 1]$, leads to the estimate

$$\text{Prob}\{A(q) \leq 316\} \approx 0.7705,$$

which is, obviously, over optimistic.

4.5.3 Example 3

Now we continue to consider the numerical problem discussed in Example 2. Instead of polynomial bounding approximations, the piecewise linear bounding functions, which are defined in Chapter 4.4.1, are used to estimate the hard bounds. Then, by applying Corollary 4.17 and Theorem 4.13, the hard bounds can be estimated by solving SDPs.

The numerical result is listed in Table 4.2 for various values of δ. The weighting function

is selected as $W(y) \equiv 1$.

δ	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
p^+	0.8067	0.8017	0.7959	0.7907	0.7848	0.7791	0.7712	0.7222
p^-	0.1170	0.1441	0.1790	0.2255	0.2906	0.3882	0.4023	0.4137

Table 4.2: Upper bounds and lower bounds of the worst-case probability

4.6 Concluding Remarks

In this chapter, the problem of determining the worst-case expected value of a polynomial performance function over a class of admissible distributions is addressed. A uniformity principle is proposed to verify that if the uniform distribution leads to the worst-case expected value of a semi-algebraic function. In the case this principle is not valid, an equivalent convex formulation is provided to compute the worst-case expected value. LMI relaxations are also proposed based on the results from the theory of moments and polynomial optimization. It is also pointed out that, if semi-algebraic (polynomial) approximations are available to a continuous function, the value can be well estimated by using these approximations. Moreover, by introducing the novel concept of *bounding approximations*, the results developed in this chapter are applied to the computation of hard bounds on probabilistic objective functions.

In further research, we aim at extending these result to other classes of distributions. For example, it is possible that the proposed approach can be applied to the extension of class $\mathcal{F}$ where symmetry requirement is eliminated. Hence, we aim at the developing similar algorithms for computing worst-case expected performance for other classes of distributions, which is also be of interest in practice.

So far, we have discussed solving important design and analysis problems in control and system theory under two different probabilistic frameworks: the chance-constraint problem and the distributional robustness problem. Starting from next chapter, we focus on solving some well-recognized and difficult problems in system identification and generalized fixed-order interpolation. By formulating them as very specific kinds of polynomial optimization, we propose computationally efficient algorithms to solve the problems.

Chapter 5

Hybrid System Identification

The problem of identifying discrete time affine hybrid systems with noisy measurements is addressed in this chapter. Given a finite number of measurements of input/output and a bound on the measurement noise, the objective is to identify a switching sequence and a set of affine models that are compatible with the a priori information, while minimizing the number of affine models. While this problem has been successfully addressed in the literature if the input/output data are noise-free or corrupted by process noise, results for the case of measurement noise are limited. In this chapter, we first recast the problem into a *sparse* polynomial optimization form by exploiting its inherent sparse structure. Yet even with sparsity, the new formulated problem can be used to solve very small size problems. Thus we develop and propose two different algorithms to overcome the computational difficulty: a randomized hit-and-run type algorithm and a deterministic algorithm. In the framework where a randomized algorithm is used, we first propose an iterative algorithm where computational cost is greatly reduced yet global convergence cannot be guaranteed. Then we propose a hit-and-run algorithm to circumvent the local convergence problem. In the deterministic algorithm, a finite dimensional semi-definite problem is given which is equivalent to the identification problem. Moreover, to address computational complexity issues, an equivalent rank minimization problem subject to deterministic LMI constraints is provided, as efficient convex relaxations for rank minimization are available in the literature. Numerical examples are provided, illustrating the effectiveness of both algorithms.

5.1 Introduction

In general, a hybrid system is a system whose behavior is determined by switching dynamics. These systems arise in many different contexts, for examples, circuit network, biological systems, systems with interaction with logic devices and continuous processes. In addition, they can be used to approximate nonlinear dynamics. Thus, due to the potential application to a vast set of practical problems, the problem of identifying input/output hybrid models has attracted considerable attention, and several approaches have been developed.

For the identification problem of switch affine systems, there are many results available in the literature. One may refer to a thorough review [30] for a summary of recent developments. In the case where measurements are noise-free, an algebraic procedure, known as Generalized Principal Component Analysis (GPCA), has been proposed in [31, 32] to efficiently solve the problem. The problem can be also formulated as a mixed linear integer optimization problem [33] or in terms of linear complementary inequalities [34], leading to generically NP-hard problems. More recently, a greedy algorithm has been proposed to identify the system while minimizing the number of switches [35]. For robust identification of switched affine systems subject to process noise, an efficient moment-based convex approach using convex relaxations on rank minimization has been proposed in [36] . A similar approach was also pursued in [37] to solve a different problem: segmenting a collection of noisy measurements into subspaces.

In this chapter, we consider the problem of identifying discrete time affine hybrid systems with noisy measurements. First, we provide an equivalent polynomial optimization problem, which inherently has a sparse structure and satisfies the so-called *running intersection property*. This sparse structure can be used to significantly reduce computational complexity, e.g. see [38, 39, 40]. The reasoning behind this polynomial optimization based formulation bears strong connection to hybrid decoupling used in GPCA. Yet it preserves all system parameters as a part of optimizing variables, while GPCA eliminates the structure of the parameters (this is elaborated in Chapter 5.3). Moreover, it is shown that the sparse polynomial optimization problem can be solved via a *fixed size* semi-definite program (SDP).

Furthermore, for larger size problems, an equivalent rank minimization problem is provided. In formulating this rank minimization problem, inspired by the results in [36], we isolate the system parameters from the unknown noise terms, and, hence, eliminate them from the decision variables in the formulated rank minimization problem. A major advantage of the approach proposed in this chapter is that the matrix in the objective function is symmetric and of much smaller size than the ones used in [36]. This feature notably reduces the computational burden in solving convex relaxations of rank minimization problems.

5.2 Set Membership Identification

In this section, we define the hybrid system identification problem. More specifically, we consider the problem of identifying single-input-single-output switched linear systems of the form

$$y_k = -\sum_{i=1}^{n} a_i(\sigma_k)(y_{k-i} - e_{k-i}) + \sum_{i=1}^{m} b_i(\sigma_k)u_{k-i} + e_k \tag{5.1}$$

where u, y and e denote input, output and noise, respectively. The magnitude of noise is bounded by $\bar{e} > 0$. Moreover, a_i and b_i are the parameters of the system, where $\sigma_k \in \{1, 2, \ldots, s\}$ denotes which sub-system is active at time k.

Without any additional restrictions, the problem admits infinitely many solutions. For example, one can assign one trivial model corresponding to each measurement. Thus, one needs to add additional constraints or objectives to make the problem meaningful. In this chapter, we aim at minimizing the number of sub-systems and/or minimizing the order of the linear models. For simplicity, let us assume that s, the number of switched sub-systems, is known and (n, m), the order of the linear models, is also known. This assumption does not imply loss of generality since one can always increase s and/or (n, m) until a meaningful solution is found. Then, the problem of interest can be formally stated as follows.

Problem 5.1 *Given input and corrupted output measurements u, y over the interval $[1, L]$, a bound $\bar{e}$ on the ℓ_∞ norm of measurement noise e (i.e. $|e_k| \leq \bar{e}$ for $k \in [1, L]$), the number of sub-models s and the order of sub-models (n and m), find a switched affine model of the*

form (5.1) that is consistent with all a priori information and the measurement data, or conclude that none exists.

Compared to the switched autoregressive exogenous (SARX) linear models considered in [34, 36], the model (5.1) assumes that one has measurement noise and does not contain unmodeled system dynamics.[1] In [89], the case with measurement noise is considered and a polynomial optimization problem is formulated for finding a compatible hybrid model, if any, using an algebraic procedure known as GPCA. In the next section, we use a related but different algebraic procedure, with two main advantages. First, the parameters of the linear models are explicitly included in the optimization variables, i.e., once the polynomial optimization problem is solved, the system parameters are known immediately. Hence, there is no need to use parameter recovering algorithms as in GPCA related approaches. Second, it is shown the polynomial optimization problem can be solved via a fixed size SDP program.

5.3 Algebraic Reformulation

In this section, based on the so-called *hybrid decoupling constraint* introduced in [32], one can see that equation (5.1) is equivalent to the polynomial equation

$$\begin{aligned} p_{0,k}(e,a,b) \;\doteq\; & \prod_{j=1}^{s}[(y_k + \sum_{i=1}^{n} a_i(j)(y_{k-i} - e_{k-i}) - \\ & \sum_{i=1}^{m} b_i(j)u_{k-i} - e_k)] = 0 \end{aligned} \tag{5.2}$$

which holds for all k. Then, the identification problem is equivalent to finding some admissible noise e and parameters $a_i(j)$ and $b_i(j)$, so that $p_k(e) = 0, k = 1, \ldots, L$.

To address this issue, we consider the following polynomial optimization problem.

Problem 5.2 *Given the number of the sub-systems s and the order of linear models (n, m),*

[1] This formulation can be easily modified to include unmodeled dynamics, see Remark 5.5.

find

$$\min_{e,a,b} \quad p_0(e,a,b) \qquad (5.3)$$
$$s.t. \qquad \|e\|_\infty \leq \bar{e},$$

where

$$p_0(e,a,b) = \sum_{k=1}^{L} p_{0,k}(e,a,b)^2.$$

The equivalence between the above problem and Problem 5.1 is established in the following theorem.

Theorem 5.1 *Given the number of the sub-systems s and the order of the sub-models (n,m), if there exists at least one compatible switched affine model for Problem 5.1, then there exist noise e and parameters (a,b) so that the minimum of Problem 5.2 is zero. The converse is also true.*

Proof: If there is a compatible hybrid model, then, given the true values of parameters and noise, one have $p_{0,k}(e,a,b) = 0$ for all k. Hence, $p_0 = 0$. Since p_0 is a SOS, $p_0 \geq 0$, hence, the minimum of (5.3) is zero. Conversely, if $p_0 = 0$ for some e, a, b, then $p_{0,k}(e,a,b) = 0$ for all k. Since $p_{0,k}$ is a product of s polynomials, then one of them is equal to zero. Denote the index of such a polynomial by $\sigma(k)$. Hence, a system with a, b being its parameters is a compatible model, and $\sigma(k)$ is the index of the active sub-model at time k. □

Alternatively, similar to the algebraic procedure, known as GPCA introduced in [32], one can re-write equation (5.2) as the following:

$$p_{0,k}(e,a,b) = \prod_{i=1}^{s} (\mathbf{b}_i^T \mathbf{r}_k) = \mathbf{c_s}^T v_s(\mathbf{r}_k) = 0, \qquad (5.4)$$

where $\mathbf{b}(i) = [1, a_1(i), \ldots, a_n(i), \ldots, b_1(i), \ldots, b_m(i)]$, $v_s(\cdot)$ is the Veronese map and $\mathbf{r}_k = [y_k - e_k, \ldots, y_{k-n} - e_{k-n}, u_{k-1}, \ldots, u_{k-m}]$. Then, a "noisy" Veronese matrix $V_s(\mathbf{r}, e)$ can be constructed and the identification problem is equivalent to finding admissible noise e so

that $V_s(\mathbf{r}, e)$ has a non-trivial null space, i.e. finding e and a non-zero vector x that

$$V_s(\mathbf{r}, e)x \doteq \begin{bmatrix} v_s(\mathbf{r}_1, e) \\ \vdots \\ v_s(\mathbf{r}_L, e) \end{bmatrix} x = 0. \tag{5.5}$$

Once such a null space is found, it can be used to recover the parameters of the sub-systems; e.g. see [32].

Now we present a simple example to illustrate how to construct Veronese matrix.

Example 5.1 *For $s = 2$ and the order $(n, m) = (1, 1)$, equation (5.2) can be written that*

$$\begin{aligned} &(y_k + a_1(y_{k-1} - e_{k-1}) - b_1 u_k - e_k) \cdot \\ &(y_k + a_2(y_{k-1} - e_{k-1}) - b_2 u_k - e_k) \quad = \quad 0. \end{aligned}$$

Hence, the k-th row of the noisy Veronese matrix V_s can be written as

$$v_k(\mathbf{r}, e) = \begin{pmatrix} (y_k - e_k)^2 \\ (y_{k-1} - e_{k-1})(y_k - e_k) \\ -u_k(y_k - e_k) \\ -u_k(y_{k-1} - e_{k-1}) \\ (y_{k-1} - e_{k-1})^2 \\ u_k^2 \end{pmatrix}^T$$

Remark 5.2 *As mentioned above, there is a connection between the formulation in (5.2) and GPCA type formulation as in (5.5). There are, however, substantial differences between these two. In GPCA-type approaches, an algebraic procedure is used to construct the so-called* Veronese *matrix. The related problem is to find the null space of the matrix and to* extract *system parameters from the null vector, e.g., see [32]. On the other hand, in the optimization problem (5.3), the system parameters are a part of the optimization variables. Hence, a necessary and sufficient condition is derived in Theorem 5.1; and once the optimization problem is solved, the parameters are determined immediately.*

Although Problem 5.1 and Problem 5.2 are equivalent, at first glance it seems that the reformulation is at least equally difficult to solve, since (5.3) contains variables a, b and e where the dimension of e is equal to the number of measurements. However, if one carefully checks the structure of the polynomial p_0, it can be found that it is *sparse*. In fact, p_0 is a sum of squares of $p_{0,k}$ and for each k, $p_{0,k}$ is a polynomial of variables a, b and $e_{k-n}, \ldots, e_k$. Hence, problem (5.3) inherently satisfies the *running intersection property* defined in Definition 2.9, with

$$I_k = \{a, b, e_k, \ldots, e_{k+n}\}, \quad k = 1, 2, \ldots, L. \tag{5.6}$$

Moreover, p_0 is a sum-of-square already, implying that, if the minimum of (5.3) is zero, $p_0 - p_0^*$ can be represented in the form of (2.17); i.e., one just needs to take $t_{k,0} = p_{0,k}$ and $t_{k,i} = 0$. Consequently, Problem 5.2 can be solved via a fixed and finite sized SDP. This is summarized as follows.

Theorem 5.3 *Given the number of the sub-systems s and the order of the sub-models (n, m), consider the following optimization problem*

$$\begin{aligned} p_s^* = \quad & \min_{m} \quad \sum_{\alpha} p_{0,\alpha} m_\alpha \\ & s.t. \quad M_N(m, I_k) \succeq 0, \quad k = 1, \ldots, L \\ & \qquad M_{N_i}(p_i m, I_{\beta(i)}) \succeq 0, \quad i = 1, \ldots, L+n, \end{aligned} \tag{5.7}$$

where $I_{\beta(i)} = \{a, b, e_{\beta(i)}, \ldots, e_{\beta(i)+n}\}$ is the partial variable set for p_i, $\beta(i) = i$ for $i = 1, \ldots, L$ and $\beta(i) = L$ for $i = L+1, L+n$; $N = 2s$ and $N_i = 2s-1$ for $i = 1, \ldots, L+n$; $p_i = \bar{e}^2 - e_i^2, i = 1, \ldots, L+n$. Then, if there exists at least one compatible switched affine model for Problem 5.1, $p_s^* = 0$ is the optimum of (5.7). Conversely, if $p_s^* = 0$, rank $M_N(m, I_k)$ = rank $M_{N_i}(m, I_k)$ for all k, and rank $M_N(m, I_k \cap I_j) = 1$ for all pairs (j, k) with $I_k \cap I_j \neq \emptyset$, then, there exists at least one compatible model.

Proof: This is a direct consequence of Theorem 2.11, given the fact that running intersection property holds for the collection of the variable sets I_k defined in (5.6) and the fact

that p_0 has a representation of (2.17) by taking $t_{k,0} = p_{0,k}$ and $t_{k,i} = 0$. $\square$

Remark 5.4 *The rank condition rank* $M_N(m, I_k) =$ *rank* $M_{N_i}(m, I_k)$ *and rank* $M_N(m, I_k \cap I_j) = 1$ *is a sufficient condition to guarantee that the optimum of the SDP relaxation is the same as that of the corresponding polynomial optimization problem, see e.g. [90, 39]. With this rank condition being satisfied, an algorithm is given in [90], which can always extract an optimal moment sequence corresponding to a probability measure with point support.*

Remark 5.5 *In the case where both input noise and output noise are considered, the results above still apply. One just need to add more variables for input noise terms in the polynomial optimization problem, which has a similar structure as shown above.*

The following simple example illustrates that exploiting the sparse structure of the problem substantially reduces computational complexity.

Example 5.2 *Consider a hybrid system with two sub-systems* ($s = 3$) *of order* $(n, m) = (1, 0)$, *and 20 measurements* ($L = 20$). *Then, in Problem 5.2, one would have to build a moment matrix of size* $\Theta_1 = \binom{21+3+6}{6} = 593,775$ *and 21 localizing matrices of size* $\Theta_2 = \binom{21+3+5}{5} = 118,755$ *in the SDP, for solving a general polynomial optimization problem. On the other hand, by taking into account the sparsity, one could build 20 moment matrices of size* $\Theta_3 = \binom{2+3+6}{6} = 462$ *and 21 localizing matrices of size* $\Theta_4 = \binom{2+3+5}{5} = 252$ *for constructing an equivalent SDP.*

As illustrated above, by taking into account the sparsity, the optimization problem can be solved if its size is relatively small. One important observation is that the complexity is proportional to the number of measurements, i.e. if one fixes the structure of the hybrid system, the maximum size of LMIs in the SDP remains the same and the number of LMIs increases proportionally. Hence, if the hybrid system is relatively simple, the identification problem can be solved via solving (5.7) directly. However, if the hybrid system becomes more complex, say, the number of sub-systems increases or the order of the sub-models increases, the problem becomes numerically difficult to solve. This comes from the fact

that the size of I_k is equal to $n_\nu = n + 1 + s(n + m)$, and the max size of the moment matrices is equal to $\binom{n_\nu+2s}{2s}$.

To overcome this numerical difficulty, we propose two algorithms based on different methodologies: randomized algorithm based on a hit-and-run method and a deterministic algorithm motivated by GPCA related algorithms.

5.4 A Hit-and-run Algorithm to an Equivalent Two-step Optimization Problem

In this section, we propose an iterative algorithm based on the condition (5.5). More precisely, we start from a random unit vector x_0, and solve (5.3) for $x = x_0$ to find e that minimizes the objective function. Then, we fix e and let x be the singular vector of $V_s(\mathbf{r}, e)$ corresponding to its minimal singular value. We repeat this iteration until x and e converge. This iteration problem is formally defined as follows.

Problem 5.3 *Given an initial unit vector x_0, find e^* and x^* such that*

$$e^*(x_0) = \arg \min_{e} \quad \|V_s(\mathbf{r}, e)x^*\|_2^2 \qquad s.t. \quad \|e\|_\infty \leq \bar{e}, \tag{5.8}$$

and

$$x^*(x_0) = \arg \min_{x} \quad \|V_s(\mathbf{r}, e^*)x\|_2^2 \qquad s.t. \quad \|x\|_2^2 = 1. \tag{5.9}$$

It is easy to see that the computational cost is significantly reduced by fixing a part of the unknown variables. On the negative side, however, one may encounter a convergence problem. That is, counterexamples can be constructed where the procedure above converges to some local minimum, depending on the initial vector x_0. Therefore, we propose a randomized algorithm that circumvents this problem to circumvent the local convergence problem.

First of all, let us state the following proposition that helps in the understanding the algorithm.

Proposition 5.6 *Let*

$$\gamma = \|V_s(\mathbf{r}, e^*(x_0))x^*(x_0)\|_2^2$$

where $e^*(x_0)$ and $x^*(x_0)$ are the solutions of Problem 5.3 for some initial unit vector x_0. Then, the minimum of Problem 5.2 is γ if and only if the minimum of the problem

$$\begin{aligned} \min_{e,\beta,d} \quad & \|V_s(\mathbf{r}, e)(x^* + \beta d)\|_2^2 - \gamma\|x^* + \beta d\|_2^2 \\ s.t. \quad & \|e\|_\infty \leq \bar{e}. \end{aligned} \tag{5.10}$$

is zero.

Proof: We prove it by contradiction. For the "if" part, assume that the minimum of Problem 5.2 is not γ. Then there exist some noise e and some unit vector $\tilde{x}$ that

$$\|V_s(\mathbf{r}, e)\tilde{x}\|_2^2 < \gamma.$$

Let $d = \tilde{x} - x^*$ and $\beta = 1$, then,

$$\|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 < \gamma.$$

Hence,

$$\begin{aligned} &\|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 - \gamma\|x + \beta d\|_2^2 \\ &\qquad = \|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 - \gamma < 0 \end{aligned}$$

which is a contradiction.

For the "only if" part, assume the minimum of problem (5.10) is less than zero for some direction d and some scalar β. Note this minimum is always non-positve, since the objective

is always zero if one take $\beta = 0$. Then, let the unit vector

$$\tilde{x} = \frac{x^* + \beta d}{\|x^* + \beta d\|_2},$$

one has

$$\begin{aligned} &\|V_s(\mathbf{r}, e)(x^* + \beta d)\|_2^2 - \gamma \|x^* + \beta d\|_2^2 \\ &\quad = \|x^* + \beta d\|_2^2 \cdot (\|V_s(\mathbf{r}, e)\tilde{x}\|_2^2 - \gamma) < 0 \end{aligned}$$

Hence, γ cannot be the minimum of Problem 5.2, which is a contradiction. □

Thus, solving problem (5.10) is equivalent to solving Problem 5.2, which is also a sparse polynomial optimization problem satisfying running intersection property. Now we are ready to state the randomized algorithm, which is summarized in Algorithm 5.1.

5.4.1 Computational Cost Analysis

In Step 1, since x is fixed, the objective polynomial in problem (5.3) becomes

$$p_0 = p_{0,1}(e_1, \ldots, e_{n+1}) + \ldots + p_{0,L}(e_L, \ldots, e_{L+n}).$$

Hence, for a fixed relaxation order N, the size of the moment matrices is $\binom{n+1+N}{N}$ and the size of localizing matrices is $\binom{n+N-1}{N-2}$, in the constructed SDP. The number of LMIs is $2L + 1$. Note also that since the total degree of p_0 is equal to $2s$, it usually does not require a high relaxation order to solve the optimization problem at or very close to the optimality.

Step 2 is a singular value decomposition for a constant matrix. The size of the matrix $V_s(\mathbf{r}, e)^T V_s(\mathbf{r}, e)$ is $\binom{n+m+s}{s}$, which is not related to the data length of measurement data.

In Step 4, to construct a SDP relaxation for a fixed relaxation order N, the size of the moment matrices is $\binom{n+2+N}{N}$ and the size of localizing matrices is $\binom{n+N}{N-2}$. The number of LMIs is $2L + 1$.

By the analysis above, we observe that this algorithm can be applied to complicated problems encountered in practice. However, it is unclear how many H&R iterations are

Algorithm 5.1 Hit-and-Run Algorithm

Step 0: Generate a random vector x_0 that $\|x_0\| = 1$
Set $x^* \leftarrow x_0$
repeat
repeat
Step 1: Set $x \leftarrow x^*$, solve

$$e^* \leftarrow \arg\min_{e} \quad \|V_s(\mathbf{r}, e)x\|_2^2 \qquad s.t. \qquad \|e\|_\infty \leq \bar{e}$$

Step 2: Set $e \leftarrow e^*$, solve

$$x^* \leftarrow \arg\min_{x} \quad \|V_s(\mathbf{r}, e)x\|_2^2 \qquad s.t. \qquad \|x\|_2^2 = 1$$

until

$$\min_{e} \ \|V_s(\mathbf{r}, e)x^*\|_2^2 \ \ s.t. \ \ \|e\|_\infty \leq \bar{e} \quad = \quad \min_{x} \ \|V_s(\mathbf{r}, e^*)x\|_2^2 \ \ s.t. \ \ \|x\|_2^2 = 1$$

Step 3: Set $\gamma \leftarrow \|V_s(\mathbf{r}, e^*)x^*\|_2^2$
repeat
Step 4: (H&R) Pick a random direction d, solve

$$(e, \beta) \leftarrow \arg\min \quad \|V_s(\mathbf{r}, e)(x + \beta d)\|_2^2 - \gamma\|x + \beta d\|_2^2 \qquad \text{s.t.} \qquad \|e\|_\infty \leq \bar{e}$$

until a convergence criterion is reached
until
return x^*

needed to find a singular value close enough to zero, given a solution exists. In our numerical examples, we found the algorithm usually terminates and gives a good estimate in less than 100 iterations.

5.5 A Deterministic Algorithm to an Equivalent Rank Minimization Problem

In this section, we formulate an equivalent rank minimization problem by using quadratic functions of rows of the Veronese matrix to eliminate the system parameters a, b in the

optimization problem. It can be seen that the main computational difficulty comes from the fact that all system parameters are a part of the decision variables in the polynomial optimization. Thus, if the structure of the hybrid system is complex, one may need high computational power to solve (5.7).

In this section, we show that, by manipulating the rows of the noisy Veronese matrix, a rank minimization problem with finite LMIs can be formulated and it is equivalent to the original identification problem. Though rank minimization problem is in general NP-hard, efficient relaxations are available especially for symmetric positive semi-definite matrices, which is the case encountered.

By revisiting the Veronese matrix condition in (5.5), it is shown the identification problem is equivalent to finding admissible noise e so that $V_s(\mathbf{r}, e)$ has a non-trivial null space, i.e. finding e and a non-zero vector x so that (5.5) holds. Once such a null space is found, it can be used to recover the parameters of the sub-systems; e.g. see [32]. This is summarized in the following proposition.

Proposition 5.7 (Theorem 1, 2 in [32]) *Given the number of the sub-systems s, the order of sub-models (n, m), if there exists at least one compatible switched affine model for Problem 5.1, then there exists some admissible noise e such that $V_s(\mathbf{r}, e)$ is rank deficient; and the system parameters a, b can be extracted from its null space.*

Remark 5.8 *The size of the Veronese matrix is $L \times \binom{n+m+s}{s}$. For each row of the matrix, the highest total degree of the polynomials (in e) is s, the number of sub-systems.*

Given the noisy Veronese matrix constructed above, let's define a matrix

$$Q(m) \doteq \sum_{i=1}^{L} \mathbf{E}_{\mu}\{v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)\} \tag{5.11}$$

where m is the truncated moment sequence corresponding to probability measure μ. Hence, $Q(m)$ is a symmetric matrix linear in m. Now, we are ready to propose the equivalent rank minimization problem.

Problem 5.4 Given the number of sub-systems s and the order of sub-models (n, m), find a rank deficient matrix $Q(m)$ defined in (5.11) subject to

$$M_N(m, I_k) \succeq 0, \quad k = 1, \ldots, L,$$
$$M_{N_i}(p_i m, I_{\beta(i)}) \succeq 0, \quad i = 1, \ldots, L+n,$$

where $I_{\beta(i)} = \{e_{\beta(i)}, \ldots, e_{\beta(i)+n}\}$ is the partial variable set for p_i, $\beta(i) = i$ for $i = 1, \ldots, L$ and $\beta(i) = L$ for $i = L+1, \ldots, L+n$; $N = s$ and $N_i = s-1$ for $i = 1, \ldots, L+n$; and $p_i = \bar{e}^2 - e_i^2, i = 1, \ldots, L+n$.

Then, we observe that this problem is indeed equivalent to the problem of finding an admissible noise sequence that results in a rank deficient Veronese matrix.

Theorem 5.9 *If there exists an admissible noise sequence e such that the Veronese matrix $V_s(\mathbf{r}, e)$ is rank deficient, then Problem 5.4 has a feasible solution. Conversely, if Problem 5.4 has a feasible solution m, and if rank $M_N(m, I_k) =$ rank $M_{N_i}(m, I_k)$ for all k and rank $M_N(m, I_k \cap I_j) = 1$ for all pairs (j, k) with $I_k \cap I_j \neq \emptyset$, then, $V_s(\mathbf{r}, e)$ is rank deficient for some admissible noise e.*

Proof: See Appendix C.1. □

Although rank minimization is NP-hard, efficient convex relaxations are available. In particular, good approximate solutions can be obtained by using a log-det heuristic that relaxes rank minimization to a sequence of convex problems, e.g., see [91, 36]. Furthermore, as stated in Theorem 5.9, it suffices to find a rank deficient solution. Thus, we use a modification of log-det heuristic that aims at dropping the rank by one, as illustrated in the following.

Remark 5.10 *Given a rank deficient matrix $Q(m^*)$ is found with a moment sequence m^*, a null vector can be easily determined by computing its eigenvectors and eigenvalues. Moreover, by Theorem 5.9, there exists some admissible noise e that the Veronese matrix $V_s(\mathbf{r}, e)$ is rank deficient. One can extract such a noise value e from the moment sequence m^* using*

Algorithm 5.2 Drop Rank

Set $X \leftarrow Q(m)$, $X_0 \leftarrow I$, $k \leftarrow 0$.
repeat
Solve

$$X_{k+1} \leftarrow \begin{array}{ll} \arg\min & \mathrm{Tr}(X_k + \delta I)^{-1} X \\ \text{s.t.} & m \in \mathcal{M} \end{array}$$

where $\mathcal{M}$ is the feasible set in Problem 5.4.
Decompose the symmetric matrix $X_k = T^{-1} D T$.
Set $\delta \leftarrow \min \mathrm{diag}(D)$.
Set $k \leftarrow k + 1$.
until a convergence criterion is reached.
return X_k

the algorithm introduced in [90]. Once the noise values are estimated, the problem can be converted to the noise free case by plugging the noise estimates in to the Veronese matrix and the system parameters can be computed using the procedure introduced in [31].

5.6 Numerical Examples: Randomized Algorithm

In this section, we present some numerical results for the proposed randomized algorithm. In the first part, the CPU time needed for solving the sub-problems in Algorithm 5.1 is given. In the second part, we use an academic example to illustrate the effectiveness of the proposed hit-and-run algorithm.

5.6.1 CPU Time of Each Iteration for Different Size Problems

In this example, we run the code in Matlab2007R on a computer with a 1.8GHz CPU and 1 GB RAM. Moreover, we used the package provided by Waki et al. [92] to solve the sparse polynomial optimization problem. The CPU time used for formulating the optimization problem and solving the SDP relaxations is listed in Table 5.1.

Remark 5.11 *As shown in the table, if one fixes the relaxation order N, the CPU time for solving the optimization problems is proportional to the number of measurements L. On the other hand, the conversion time (that is, the time used to formulate the sparse optimization*

Table 5.1: Estimated and true values of parameters

N	L	s	Conversion Time (sec)	SeDuMi Time (sec)
2	20	2	2.21	0.19
4	20	2	2.48	1.02
8	20	2	4.51	20.89
2	40	2	8.28	0.31
4	40	2	9.33	2.18
8	40	2	13.29	49.64
2	80	2	112.12	0.60

N: relaxation order; L: number of measurements;
s: number of sub-systems.

problems into standard SeDuMi forms) is long when L is large. However, one should note this formulation can be done symbolically once the structure of the hybrid system is fixed. Thus it can be expected that the algorithm can be applied to handle problems of much larger size, if the coding is optimized,

5.6.2 Illustrative Example

In this section, we consider an illustrative example to demonstrate the effectiveness of the proposed algorithm. It is assumed that there are three sub-systems with the order $(n, m) = (2, 1)$. That is,

$$y_k = -a_{1,i}(y_{k-1} - e_{k-1}) - a_{2,i}(y_{k-2} - e_{k-2}) + b_i u_{k-1} + e_k$$

where $i = 1, 2, 3$. We take $L = 90$ measurement with $s = 5$ switches. The simulation is run for two different noise levels, i.e. $\bar{e} = 0.05$ and $\bar{e} = 0.2$ (note that the GPCA approach usually cannot handle noisy data well, especially for large noise; see [36]). The input signal u is random ranging from -1 to 1. In Table 5.2, we present the values of the system parameters identified by the algorithm for the noise levels considered.

Moreover, the measurements are classified as coming from one of the the sub-models, indicating which one is active at each time k. This is shown in Figure 5.1.

It can be seen that, in the noise-free case, all measurements are correctly classified. in the

Table 5.2: Estimated and true values of parameters

		True	Noise $e < 0.05$	Noise $e < 0.2$
Sub-model 1	a_1	1.8000	1.7642	1.9320
	a_2	0.8100	0.8261	0.7034
	b_1	1.5000	1.5134	1.5820
Sub-model 2	a_1	0.5000	0.4742	0.3394
	a_2	0.0600	0.0436	0.0240
	b_1	1.0000	0.9821	1.3740
Sub-model 3	a_1	-1.4000	-1.3627	-1.1360
	a_2	0.7400	1.7642	0.6722
	b_1	1.0000	1.0461	1.3836

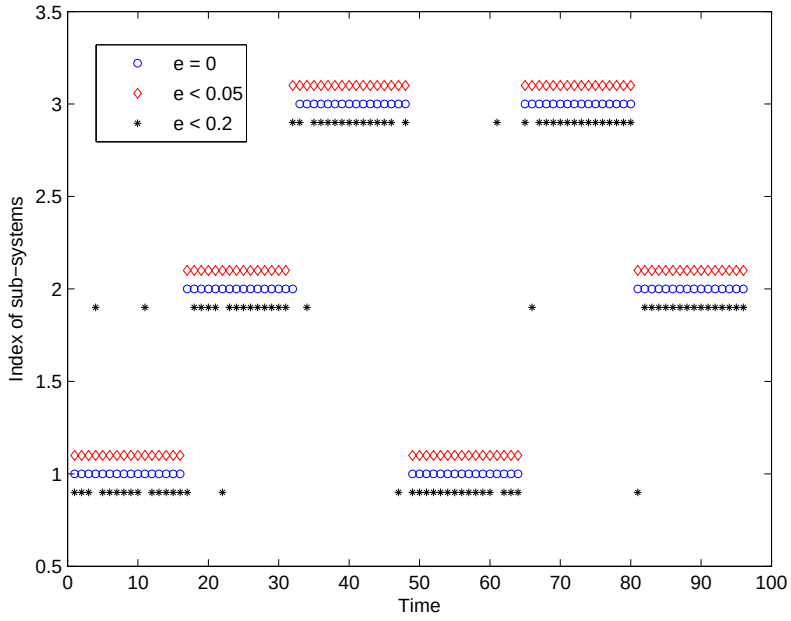

Figure 5.1: Identified and true active sub-systems v.s. time

case of small noise, the parameters of the sub-systems can be closely recovered and almost all (except one) measurements are correctly classified. For the case of large noise, one could conclude that the measurements are "misclassified" at several time instants. However, it should be noted that, given the model identified above, the classification obtained is indeed compatible with the measured data and a priori information on the magnitude of the noise. In fact, if the dynamic of sub-models are close and the noise is large, one may get a hybrid

system with less sub-models. This should not be regarded as a failure of the algorithm, as the system identified is compatible with all the given information.

5.7 Numerical Examples: Deterministic Algorithms

In this section, we present several numerical examples for illustrating the proposed algorithms. In the first example, a simple hybrid system with limited measurements is considered. The identification problem is then addressed by solving the SDP problem (5.7) directly, according to Theorem 5.3. In the second example, a more complex hybrid model is considered. It is then shown that the rank minimization relaxation algorithm in Algorithm 5.2 works efficiently on identifying the system parameters. A few comments are also given regarding how the noise bound may affect identification results.

5.7.1 Example I: via Sparse Polynomial Optimization

In this example, we consider a switched affine system with $s = 3$ and $(n, m) = (1, 0)$, where the sub-models are

$$\begin{aligned} y_k &= -0.9(y_{k-1} - e_{k-1}) + u_{k-1} + e_k \quad \text{(Submodel 1)} \\ y_k &= -0.5(y_{k-1} - e_{k-1}) + u_{k-1} + e_k \quad \text{(Submodel 2)} \\ y_k &= +0.7(y_{k-1} - e_{k-1}) + u_{k-1} + e_k \quad \text{(Submodel 3)} \end{aligned}$$

and the system is modeled as (5.1), i.e.,

$$y_k = -a_1(\sigma_k)(y_{k-1} - e_{k-1}) + u_{k-1} + e_k,$$

where $\sigma_k \in \{1, 2, 3\}$ depending on which sub-model is active at time k. In the simulation, we set $\sigma(k) = 1$ for $k = [1, 5]$, $\sigma(k) = 2$ for $k = [6, 10]$ and $k = [16, 20]$, $\sigma(k) = 3$ for $k = [11, 15]$. The experimental data were obtained with a unit step input and with uniformly randomly generated noise bounded by $\bar{e}$. First we set the noise bound to $\bar{e} = 0.1$ and then increase the bound by setting $\bar{e} = 0.3$. The identification problems are addressed based on Theorem 5.3

via solving a SDP relaxation built from the equivalent sparse polynomial optimization problem.

The parameter values used in the simulation and their estimated values with respect to different noise bounds are illustrated in Table 5.3.

Table 5.3: Estimated and True Values of Parameters

		True	$\bar{e} = 0.1$	$\bar{e} = 0.3$
Submodel 1	a_1	0.9000	0.9352	0.5825
Submodel 2	a_1	0.5000	0.4867	0.5825
Submodel 3	a_1	-0.7000	-0.6841	-0.6652

Moreover, the active sub-system at time k can be determined once the sub-models are identified. The results are shown in Figure 5.2.

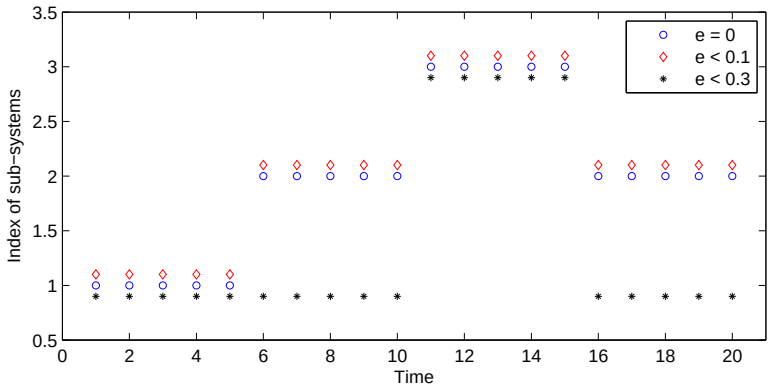

Figure 5.2: Identified and true active sub-systems v.s. time

Remark 5.12 *As one can see from Table 5.3 and Figure 5.2, if the noise bound is large, there are compatible hybrid systems with a smaller number of sub-systems. However, one should note that this is not a failure of our algorithm. In fact, the hybrid system obtained is still compatible with all the measurements and a priori information. This should not be surprising since when the measurements are limited, the difference between model dynamics could be covered by the noise, especially when the noise is large. Moreover, when the noise is large, one may get a compatible hybrid system with linear sub-models of lower orders.*

5.7.2 Example II: via Rank Minimization

In this example, we consider a more complex switched affine system to show the computational efficiency of using Theorem 5.9 and Algorithm 5.2. It is assumed that there are three sub-systems with the order $(n, m) = (2, 1)$. That is,

$$y_k = -a_{1,i}(y_{k-1} - e_{k-1}) - a_{2,i}(y_{k-2} - e_{k-2}) + b_{1,i}u_{k-1} + e_k$$

where $i \in \{1, 2, 3\}$ depend on which sub-system is active at time k. We take 120 measurements with 11 switches among these three sub-systems. The simulation is run for two different noise levels, i.e. $\bar{e} = 0.1$ and $\bar{e} = 0.3$. The input signal u is uniformly randomly generated between -1 and 1. In Table 5.4, we present the values of the system parameters identified by the algorithm for the noise levels considered.

Table 5.4: Estimated and true values of parameters

		True	Noise $e < 0.1$	Noise $e < 0.3$
Submodel 1	a_1	0.9000	0.9183	0.9586
	a_2	0.1800	0.1736	0.2379
	b_1	0.2000	0.2310	0.2863
Submodel 2	a_1	0.5000	0.5155	0.5494
	a_2	0.0600	0.0564	0.1014
	b_1	1.0000	1.1310	1.0802
Submodel 3	a_1	-1.0000	-0.9716	-0.9618
	a_2	0.3000	0.3144	0.3287
	b_1	0.6000	0.6369	0.6991

Moreover, the active sub-system at time k can be determined once the sub-models are identified. The results are shown in Figure 5.3.

5.8 Concluding Remarks

This chapter addresses the identification problem of a discrete-time affine hybrid system with input/output data corrupted by measurement noise. The proposed approach first formulates the identification problem as a polynomial optimization problem, for which there

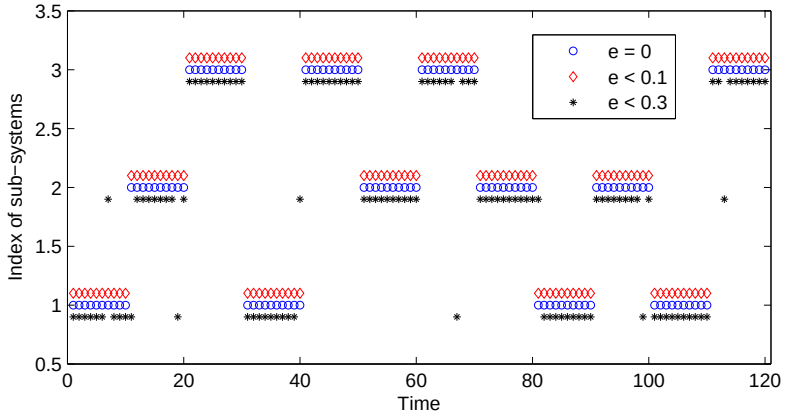

Figure 5.3: Identified and true active sub-systems v.s. time

exists convergent SDP relaxations. It is shown that the polynomial optimization problem obtained inherently has a sparse structure, which can be used to significantly reduce the size of the SDP, and, hence, to reduce computational cost. Moreover, since the objective function in the polynomial optimization has a SOS representation, its SDP relaxation with appropriate relaxation size is equivalent to the original identification problem.

If the hybrid system is complex, however, solving the sparse optimization requires tremendous memory and computation power. Thus, we develop two different algorithms to address the computational difficulty. In the first randomized algorithm, we propose a two-step iterative algorithm to significantly reduce computational cost and use hit-and-run type algorithm to circumvent the local convergence problem. For the second deterministic algorithm, a drop-rank approach is given to address the computational difficulties when the hybrid system is complex. Here the size of relaxation is determined exclusively by the number of sub-models of the hybrid system. Though rank minimization problem is in general NP-hard, efficient convex relaxations are available in the literature. Numerical examples are provided to illustrate the effectiveness of the proposed algorithms.

As shown in this chapter, by isolating system parameters from the decision variables in the optimization problem, the computational cost can be substantially reduced. However, the structure information of system parameters is then lost. Hence, in future research, we aim at developing approaches to separate the search of admissible noise and systems

parameters while preserving the structural information.

In the next chapter, we consider a generalized fixed-order interpolation problem. It is known that the interpolation problem has strong connection with system design and identification. We show in the next chapter that a polynomial optimization based formulation is equivalent to the generalized fixed-order interpolation problem. Computationally efficient algorithms are also proposed to solve the problem.

Chapter 6

Generalized Fixed Order Interpolation

In this chapter, we address the problem of finding a fixed order plant that interpolates given data in time and frequency domain, and satisfies additional constraints such as stability and passivity. The comprehensive framework developed in this chapter can be used to address a wide range of complex control problems such as system identification with a priori bound on its order, fixed-order controller design and spectral estimation. To solve the proposed interpolation problem, it is first shown that it can be recast as finding a point in a properly defined semi-algebraic set. Then, an efficient numerical algorithm based on convex relaxations of rank minimization is proposed to solve the problem. Numerical examples are provided to illustrate the efficiency of the algorithm.

6.1 Introduction

In this chapter, we provide efficient numerical algorithms aimed at finding fixed order Single-Input/Single-Output (SISO) systems that interpolate given time-domain and frequency-domain data. The framework provided also allows for the imposition of extra constraints on the interpolating system such as stability and (generalized) passivity.

More precisely, we consider the following generalized interpolation problem: Given a set

of $N_{\mathrm{NP}}+1$ distinct complex pairs of points

$$\mathcal{F} \doteq \{(z_0, F_0), (z_1, F_1), \ldots, (z_{N_{\mathrm{NP}}}, F_{N_{\mathrm{NP}}})\},$$

a set of $N_{\mathrm{CF}}+1$ real values

$$\mathcal{T} \doteq \{g_0, g_1, \ldots, g_{N_{\mathrm{CF}}}\}$$

and a quadratic supply function $s(u, y)$, where u and y denote system input and output, respectively, find a system $G(z)$ of order no greater that a given integer N, that satisfies the following conditions:

$$G(z_k) = F_k, \quad k = 0, \ldots, N_{\mathrm{NP}}; \tag{6.1}$$

$$G(z) = g_0 + g_1 z^{-1} + \ldots + g_{N_{\mathrm{CF}}} z^{-N_{\mathrm{CF}}} + \ldots; \tag{6.2}$$

$$G(z) \text{ is passive with supply function } s(u, y); \tag{6.3}$$

If passivity is not required and no bound on the order of the plant is provided, condition (6.1) amounts to the classical Nevanlinna-Pick interpolation problem [41, 42]. A classical Nevalinna-Pick interpolation problem (without degree constraint) is addressed in [41, 42], where it is shown a solution exists if and only if the so-called Pick matrix is positive semidefinite. Given $N+1$ interpolation points, it is known that an interpolant of degree at most N always exists. Moreover, a constructive algorithm is given based on convex optimization for finding all such interpolants in [43].

If instead condition (6.2), i.e., the interpolation conditions on time-domain data is required, it yields to a standard Caratheodory-Fejer interpolation problem [44]. A so-called *maximum entropy* solution is given in [45] to parameterize all rational covariance extensions of degree at most N, given $N+1$ interpolation points.

If both conditions (6.1) and (6.2) are required, it yields to mixed time/frequency-domain identification in the related literature of system identification [48]. These are well-known interpolation problems and have been extensively studied; see [45, 43, 46, 47] and references therein. Interpolatory identification algorithms based on linear matrix inequalities (LMIs)

have been proposed in [46, 47, 48] to solve mixed time/frequency robust identification problems, resulting interpolants of degree at most N given $N+1$ mixed domain interpolation points. Based on interpolation theory developed in [43], the problem of stabilizing controller synthesis is discussed in [49, 50].

Condition (6.3) allows one to address important problems such as fixed order controller design with proven H_∞ performance and spectral estimation. To see this, consider quadratic supply functions of the form

$$s(u,y) \doteq \begin{pmatrix} u \\ y \end{pmatrix}^T \Pi \begin{pmatrix} u \\ y \end{pmatrix},$$

Then, requiring that the system satisfies

$$\|G(z)\|_\infty \leq \gamma$$

is equivalent to requiring dissipativity of $G(z)$ with a supply function defined by the matrix

$$\Pi = \begin{pmatrix} -\gamma^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Moreover, requiring $G(z)$ to be positive real (an important property in many controller design and spectral estimation problems), i.e.,

$$G(e^{-j\omega}) + G(e^{j\omega}) \geq 0$$

is equivalent to having a dissipative $G(z)$ with a supply function defined by the matrix

$$\Pi = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

More examples on using passivity to address control problems can be found in [93].

The interpolation problem mentioned above is known to be a complex optimization

problem [43]. Therefore, the main contribution of this chapter is to provide computationally efficient relaxations of this problem. More precisely, we start by recasting the interpolation problem as finding a point in a properly defined semi-algebraic set. To find such a point or to conclude that none exists, we use tools from the field of polynomial optimization to develop convex semi-definite (SDP) relaxations that are proven to converge to a solution of the proposed semi-algebraic feasibility problem. Moreover, we exploit efficient convex relaxations of the rank minimization problem to improve the computational efficiency of the proposed approach.

It should be emphasized that polynomial based approaches (with or without consideration of sparsity) have been proposed and discussed in the literature of system identification; e.g. see [94] for hybrid system identification, see [51] for set-membership error-In-variables identification and see [52] for fixed-order identification with stability constraints. In this chapter, we go beyond results available in the literature and provide a comprehensive framework for fixed-order interpolation which enables one to address a wide range complex problems such as fixed order controller design, system identification and spectral estimation. Moreover, by relying on results on polynomial optimization and rank minimization, we provide efficient algorithms to address this complex problem.

6.2 Problem Formulation

In this chapter, we consider a generalized interpolation problem where a fixed order (possibly passive) SISO system is required to satisfy a set of interpolation conditions both in time and frequency domain. To begin with, we first define the set of allowable transfer functions.

Definition 6.1 *Given two integers n and m, a positive number $\rho \leq 1$ and a symmetric 2×2 real matrix*

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{12} & \pi_{22} \end{pmatrix},$$

a transfer function $G(z)$ of the form

$$G(z) = \frac{b(z)}{a(z)}, \tag{6.4}$$

is said to belong to $\mathcal{G}(\rho, \Pi, n, m)$ if it

a) is of fixed-order (n, m), i.e.,

$$\begin{aligned} a(z) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n}, \\ b(z) &= b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}; \end{aligned}$$

b) is exponentially stable with root radius ρ, i.e., all the zeros of $a(z)$ lie inside the disk with radius ρ;

c) is dissipative with respect to the quadratic supply function

$$s(u_i, y_i) \doteq \begin{pmatrix} u_i \\ y_i \end{pmatrix}^T \Pi \begin{pmatrix} u_i \\ y_i \end{pmatrix},$$

i.e., there exist a energy storage function V *such that*

$$V(\mathbf{x}_{k_2}) \leq V(\mathbf{x}_{k_1}) + \sum_{i=k_1}^{k_2} s(u_i, y_i), \tag{6.5}$$

for all $k_1 \leq k_2$ and for all possible trajectories $(u, y, \mathbf{x})$, where u_k, y_k and $\mathbf{x}_k$ denote input, output and the states of the system at time k, respectively. ⋆

Then, given the two sets of interpolation points

$$\mathcal{F} \doteq \{(z_0, F_0), (z_1, F_1), \ldots, (z_{N_{\mathrm{NP}}}, F_{N_{\mathrm{NP}}})\} \subset \mathcal{C}^2 \tag{6.6}$$

and

$$\mathcal{T} \doteq \{g_0, g_1, \ldots, g_{N_{\mathrm{CF}}}\} \subset \mathcal{R}, \tag{6.7}$$

we aim at finding a SISO system $G(z) \in \mathcal{G}(\rho, \Pi, n, m)$ that satisfies the Caratheodory-Fejer interpolation conditions

$$G(z) = g_0 + g_1 z^{-1} + \ldots + g_{N_{\mathrm{CF}}} z^{-N_{\mathrm{CF}}} + \ldots, \tag{6.8}$$

i.e., the values of the first $N_{\mathrm{CF}} + 1$ impulse response of $G(z)$ are equal to the g_ks, and the Nevanlinna-Pick interpolation conditions

$$G(z_k) = F_k, \quad k = 0, \ldots, N_{\mathrm{NP}}. \tag{6.9}$$

To sum up, we aim at solving the following problem:

Problem 6.1 Given (n, m), two sets of interpolation points $\mathcal{F}$ and $\mathcal{T}$, as well as a positive scalar $\rho \leq 1$ and a symmetric matrix $\Pi \in \mathcal{R}^{2\times 2}$, find a system $G(z)$ of the form of (6.4) that belongs to $\mathcal{G}(\rho, \Pi, n, m)$ and satisfies the interpolation conditions (6.8) and (6.9).

6.3 An Equivalent Polynomial Problem

In this section, we show that Problem 6.1 defined in the previous section can be recast as finding a point in a properly defined semi-algebraic set and provide algorithms aimed at solving this feasibility poblem. As the first step, we start by formulating the interpolation conditions and the conditions on (generalized) passivity as polynomial constraints. Then, an algorithm based on convergent SDP relaxations is proposed to solve the feasibility problem.

6.3.1 Conditions on Caratheodory-Fejer interpolation

First we consider the Caratheodory-Fejer interpolation condition (6.8). Note that the sequence g_i given in $\mathcal{T}$ are the first $N_{\mathrm{CF}} + 1$ points of the unit impulse response of $G(z)$. Therefore, with fixed order (n, m), we can rewrite the interpolation condition (6.8) as a set of linear equations in the system parameters $a_i, i = 1, \ldots, n$ and $b_k, k = 0, \ldots, m$. More precisely, we start by defining the *regression vector* ϕ_k and the *system parameter vector* Θ

as follows,

$$\phi_k = [g_{k-1}, \ldots, g_{k-n}, u_k, \ldots, u_{k-m}]^T, \tag{6.10}$$

$$\Theta = [-a_1, \ldots, -a_n, b_0, \ldots, b_m]^T, \tag{6.11}$$

where

$$u_k = \delta(k) = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the impulse response g_k can be written as

$$g_k = \Theta^T \cdot \phi_k, \quad k = 0, \ldots, N_{CF}. \tag{6.12}$$

Define polynomial equations

$$p_k^{CF}(\Theta) \doteq \Theta^T \cdot \phi_k - g_k = 0, \quad k = 0, \ldots, N_{CF}. \tag{6.13}$$

Then, the following result holds.

Lemma 6.2 The Caratheodory-Fejer interpolation condition (6.8) is satisfied if and only if (6.13) holds.

6.3.2 Conditions on Nevanlinna-Pick Interpolation

Now we consider the Nevanlinna-Pick interpolation condition (6.9). In this case, given the interpolation points $(z_k, F_k) \in \mathcal{F}$ and a fixed order transfer function $G(z)$ with parameters Θ defined in (6.11), we want

$$G(z_k) = \frac{b(z_k)}{a(z_k)} = F_k, \quad k = 0, \ldots, N_{NP},$$

or, equivalently,

$$b(z_k) - F_k a(z_k) = 0.$$

Note that the equation above includes complex numbers. Thus, we define polynomials

$$\begin{aligned} p_{2k}^{\mathrm{NP}}(\Theta) &= \mathrm{Re}\{b(z_k) - F_k a(z_k)\}, \\ p_{2k+1}^{\mathrm{NP}}(\Theta) &= \mathrm{Im}\{b(z_k) - F_k a(z_k)\} \end{aligned}$$

for $k = 0, \ldots, N_{\mathrm{NP}}$. Then, we have the following results.

Lemma 6.3 *The Nevanlinna-Pick interpolation condition* (6.8) *is satisfied if and only if*

$$p_k^{NP}(\Theta) = 0, \quad k = 0, \ldots, 2N_{NP} + 1. \tag{6.14}$$

6.3.3 Conditions on Stability

In this section, we consider the stability constraint used to define $\mathcal{G}$ in Definition 6.1. It is natural to consider the stability constraint in many interpolation problems. For example, system identification using interpolation approaches always requires the identified system to be strictly stable, as in practice identification approaches are only performed for stable systems. Therefore, it is desirable to assume $G(z)$ is exponentially stable with radius $\rho < 1$. To obtain polynomial constraints on the system parameters that are equivalent to stability, we use Jury's criterion, which is a necessary and sufficient condition for bounded-input-bounded-output (BIBO) stability; e.g., see [95]. Also see [51, 52] for incorporating stability as polynomial constraints to system identification. For completeness, we briefly restate this well-known result.

Table 6.1: Jury's array

a_n	a_{n-1}	a_{n-2}	$\cdots$	a_2	a_1	1
1	a_1	a_2	$\cdots$	a_{n-2}	a_{n-1}	a_n
$J_{1,n-1}$	$J_{1,n-2}$	$J_{1,n-3}$	$\cdots$	$J_{1,1}$	$J_{1,0}$	
$J_{1,0}$	$J_{1,1}$	$J_{1,2}$	$\cdots$	$J_{1,n-2}$	$J_{1,n-1}$	
$J_{2,n-2}$	$J_{2,n-3}$	$J_{2,n-4}$	$\cdots$	$J_{2,0}$		
$J_{2,0}$	$J_{2,1}$	$J_{2,2}$	$\cdots$	$J_{2,n-2}$		
$\vdots$	$\vdots$	$\vdots$				
$J_{n-2,2}$	$J_{n-2,1}$	$J_{n-2,0}$				

Lemma 6.4 (Jury's test [95]) *The system $G(z)$ in (6.4) is stable, i.e., all the roots of $a(z)$ locate inside the unit circle, if and only if the following inequalities hold,*

$$\begin{aligned} 1+\sum_{i=1}^{n} a_i &> 0 \\ 1+\sum_{i=1}^{n}(-1)^i a_i &> 0 \\ |a_n| &< 1 \\ |J_{k,n-k}| &< |J_{k,0}|, \quad 1 \le k \le n-2 \end{aligned}$$

where $J_{k,i}, k = 1, \ldots, n-2, i = 0, \ldots, n-k$ are the elements in the Jury's array shown in Table 6.1 such that

$$J_{k,i} \doteq \begin{vmatrix} J_{k-1,n-k} & J_{k-1,i} \\ 1 & J_{k-1,n-k-i} \end{vmatrix},$$

for $k = 1, \ldots, n-2$ and $i = 0, \ldots, n-k$. □

Note that every element in the Jury's array is polynomial in the system parameters $a_1, \ldots, a_n$. Hence, the set of system parameters corresponding stable $G(z)$ can be described by the following polynomial inequalities.

$$p_k^S(\Theta) > 0, k = 1, \ldots, n+2. \tag{6.15}$$

where

$$\begin{aligned} p_1^S &= 1+\sum_{i=1}^{n} a_i, \\ p_2^S &= 1+\sum_{i=1}^{n}(-1)^i a_i, \\ p_3^S &= 1+a_n, \\ p_4^S &= 1-a_n, \\ p_{k+4}^S &= J_{k,0}^2 - J_{k,n-k}^2, \quad k = 1, \ldots, n-2. \end{aligned}$$

Remark 6.5 *If exponential stability is presumed, i.e., we want to test if all roots have magnitude less than ρ, Lemma 6.4 can be modified by setting $\hat{a}_k = a_k/\rho^k, k = 1, \ldots, n$ and substitute them correspondingly in the polynomials $p_k^S(\Theta)$.*

6.3.4 Conditions on Dissipativity

In this section, we provide equivalent polynomial constraints on dissipativity. Dissipativity is an important system concept from both theoretical and practical point of view. In this chapter, we are particularly interested in this property because it provides a unified representation for many important properties in control theory such as $\mathcal{H}_\infty$ performance and positive realness. We first restate a well-known result on dissipativity using KYP Lemma, see [93] for details. The following theorem is an adopted version of Proposition 2.16 in [93] for discrete-time systems.

Theorem 6.6 *Assume that $G(z)$ is stable and has no poles on the unit circle, then it is dissipative with respect to the quadratic supply function*

$$s(u_i, y_i) \doteq \begin{pmatrix} u_i \\ y_i \end{pmatrix}^T \Pi \begin{pmatrix} u_i \\ y_i \end{pmatrix},$$

where $\Pi \doteq \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{12} & \pi_{22} \end{pmatrix} \in \mathcal{R}^{2\times 2}$ *is a symmetric matrix, if and only if*

$$\begin{pmatrix} 1 \\ G(z^{-1}) \end{pmatrix}^* \Pi \begin{pmatrix} 1 \\ G(z^{-1}) \end{pmatrix} \leq 0 \tag{6.16}$$

for all $z \in \mathcal{C}$, $|z| = 1$. □

The above theorem shows that if stability is assumed, dissipativity is equivalent to positivity of a scalar function on the unit circle. Indeed, it can be shown that, the inequality (6.16) can be rewritten as a positivity constraint imposed on a trigonometric polynomial

on the unit circle. More precisely, note that (6.16) is equivalent to

$$-(\pi_{11} + \pi_{12}G(z^{-1}) + \pi_{12}G(z) + \pi_{22}G(z)G(z^{-1})) \geq 0.$$

Or, equivalently,

$$\begin{aligned} R(z) \doteq -\big(&\pi_{11}a(z)a(z^{-1}) + \pi_{12}a(z)b(z^{-1}) + \\ &\pi_{12}b(z)a(z^{-1}) + \pi_{22}b(z)b(z^{-1})\big) \geq 0. \end{aligned} \tag{6.17}$$

for all $z \in \mathcal{C}$ with $|z| = 1$. Note that $R(z) = R(z^{-1})$. Hence,

$$R(z) = \sum_{k=-d_r}^{d_r} r_i z^i, r_i = r_{-i}. \tag{6.18}$$

where $d_r = \max\{m, n\}$ and r_k are (quadratic) polynomials in Θ. Note that $R(z)$ is a univariate polynomial. To obtain a certificate for its positivity on the unit circle, we first introduce the following theorem.

Lemma 6.7 (Theorem 2.5 in [96]) *The univariate polynomial $R(z)$ is non-negative on the unit circle if and only if there exist a positive semi-definite matrix*

$$Q \doteq \begin{pmatrix} q_{0,0} & \cdots & q_{0,d_r} \\ \vdots & \ddots & \vdots \\ q_{d_r,0} & \cdots & q_{d_r,d_r} \end{pmatrix} \succeq 0$$

such that

$$r_k = \sum_{i=k}^{d_r+k} q_{i,i-k}, k = 0, \ldots, d_r. \tag{6.19}$$

□

Note that the matrix Q in Lemma 6.7 is linear in the coefficients of $R(z)$ but polynomial (not linear) in the system parameters Θ. Thus, we introduce additional variables λ_i, $i =$

$1, \ldots, \frac{d_r(d_r+1)}{2}$, since the matrix Q has $\frac{(d_r+1)(d_r+2)}{2}$ free variables with additional d_r+1 linear constraints. To illustrate, consider the case where $n = 2$ and $m = 1$. Then,

$$Q = \begin{pmatrix} q_{00} & q_{01} & q_{02} \\ q_{10} & q_{11} & q_{12} \\ q_{20} & q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \lambda_3 & r_2 \\ * & \lambda_2 & r_1 - \lambda_1 \\ * & * & r_0 - \lambda_1 - \lambda_2 \end{pmatrix}.$$

Note that Q is positive semi-definite if and only if its principal minors are non-negative. In addition, every principal minor of Q is polynomial in its elements, and hence, polynomial in Θ and λ. Denote these $d_r + 1$ polynomials as $p_k^{DP}(\Theta, \lambda)$. Hence, we obtain equivalent polynomial constraints for 6.16. More precisely, we define the following polynomial inequalities

$$p_k^{DP}(\Theta, \lambda) \geq 0, k = 1, \ldots, d_r + 1. \tag{6.20}$$

Then (6.20) is equivalent to (6.16) and we have

Lemma 6.8 *Given a stable $G(z)$ with no poles on the unit circle, it is dissipative with respect to the quadratic supply matrix Π if and only if* (6.20) *holds for some λ.* □

6.3.5 Polynomial Optimization and its SDP Relaxations

From the discussion in the previous sections, the conditions on describing the allowable transfer function set $\mathcal{G}(\rho, \Pi, n, m)$ and the interpolation conditions can be recast as polynomial constraints (6.13), (6.14), (6.15) and (6.20). Hence, solving Problem 6.1 is equivalent to finding a feasible point in the semi-algebraic set defined by those polynomials.

Problem 6.2 (Polynomial Reformulation) *Find Θ and λ such that $(\Theta, \lambda) \in \mathcal{K}$ where*

$$\mathcal{K} \doteq \left\{ (\Theta, \lambda) : \begin{array}{ll} p_k^{CF}(\Theta) = 0, & k = 0, \ldots, N_{CF} \\ p_k^{NP}(\Theta) = 0, & k = 0, \ldots, 2N_{NP} + 1 \\ p_k^{S}(\Theta) \geq 0, & k = 1, \ldots, 2n \\ p_k^{DP}(\Theta, \lambda) = 0, & k = 1, \ldots, d_r + 1 \end{array} \right\}.$$

Then, the follow theorem establishes equivalence between Problem 6.1 and 6.2.

Theorem 6.9 *There exist at least one model $G(z) \in \mathcal{G}(\rho, \Pi, n, m)$ to Problem 6.1 if and only if there exists a feasible solution (Θ, λ) to Problem 6.2.*

Proof: This follows directly from Lemma 6.2, 6.3, 6.4 and 6.8. □

Remark 6.10 *Note that one can choose a polynomial objective function $p_0(\Theta)$ on the system parameters to formulate the above feasibility problem to a polynomial optimization problem. The polynomial p_0 can be selected arbitrarily depending on the context of the interpolation problem. For example, in system identification using the interpolation approach, it can be chosen $p_0 = a_1$ to find lower bounds of the system parameters.*

Based on the recent results on polynomial optimization as introduced in Chapter 2, an approach based on SDP relaxations is suggested to solve Problem 6.2. Since this SDP formulation has been repetitively used in the previous chapters, we simple state the SDP relaxation hereafter to focus our discussion on the main contribution of this chapter.

Problem 6.3 (SDP Relaxation) *Find m that satisfies*

$$\begin{cases} M_N(m) \succeq 0 \\ M_{N_i}(p_k^{CF} m) = 0, \quad k = 0, \ldots, N_{CF} \\ M_{N_i}(p_k^{NP} m) = 0, \quad k = 0, \ldots, 2N_{NP} + 1 \\ M_{N_i}(p_k^{S} m) \succeq 0, \quad k = 1, \ldots, n+2 \\ M_{N_i}(p_k^{DP} m) \succeq 0, \quad k = 1, \ldots, d_r + 1 \end{cases}, \tag{6.21}$$

where m is the truncated moment sequence for (Θ, λ), N is the relaxation order, $M_N(m)$ is the moment matrix, $M_{N_i}(p_k m)$ is the localizing matrix with respect to polynomial p_k. ⋆

Theorem 6.11 *If there exists $G(z) \in \mathcal{G}(\rho, \Pi, n, m)$ to Problem 6.1, then, Problem 6.3 is feasible for any relaxation order N. Conversely, if Problem 6.3 is feasible and there exists a solution m^* satisfies*

$$rank\ M_N(m^*) = rank\ M_{N_i}(m^*), \tag{6.22}$$

there exists at least one solution $G(z) \in \mathcal{G}(\rho, \Pi, n, m)$ to Problem 6.1 and the parameter Θ can be extracted from m^*.

Proof: See Section 2.1 in [90]. □

Remark 6.12 *The rank condition is a sufficient condition to guarantee the system parameter Θ can be extracted from the optimal solution m^*, see e.g. [90]. With this rank condition being satisfied, an algorithm is given in [90], which can always extract a moment sequence corresponding to a probability measure with point support.*

So far, we have discussed on how to formulate the generalized interpolation problem to a feasibility problem on a semi-algebraic set and on how to solve it via its SDP relaxations. In the next section, a refined algorithm based on rank minimization is discussed to enforce the rank condition in Theorem 6.11.

6.4 A Rank Minimization Based Algorithm

Motivated by the fact that the system parameters can be extracted from a feasible solution of Problem 6.3 if the rank condition is satisfied, in this section we impose a stronger condition on the rank of the moment matrix, and, hence, simplify (6.21) in order to improve computational efficiency.

Fix relaxation order $N = d_{\max}/2$ where $d_{\max}$ is the highest total degree of all polynomials in (6.13), (6.14), (6.15) and (6.20). Consider the following rank constrained problem,

Problem 6.4 *Find m that satisfies*

$$\begin{cases} M \succeq 0, \\ rank(M_N(m)) = 1, \\ \sum_\alpha p^{CF}_{k,\alpha} m^\alpha = 0, & k = 0, \ldots, N_{CF} \\ \sum_\alpha p^{NP}_{k,\alpha} m^\alpha = 0, & k = 0, \ldots, 2N_{NP}+1 \\ \sum_\alpha p^{S}_{k,\alpha} m^\alpha \geq 0, & k = 1, \ldots, n+2 \\ \sum_\alpha p^{DP}_{k,\alpha} m^\alpha \geq 0, & k = 1, \ldots, d_r+1, \end{cases} . \tag{6.23}$$

where $p^{CF}_{k,\alpha}$, $p^{NP}_{k,\alpha}$, $p^{S}_{k,\alpha}$ and $p^{DP}_{k,\alpha}$ are the coefficients of the corresponding polynomials and m^α are the corresponding elements in the moment matrix M.

Theorem 6.13 *There exists a solution $G(z) \in \mathcal{G}(\rho, \Pi, n, m)$ to Problem 6.1 if and only if Problem 6.4 is feasible.*

Proof: Assume that there exists such a solution to Problem 6.1. According to Theorem 6.9, there exists (Θ, λ) that satisfies all the conditions in Problem 6.2. Let $m^\alpha = (\Theta, \lambda)^\alpha$, i.e., the moment sequence m is assigned with a Dirac distribution at the point (Θ, λ). Then the moment matrix $M_N(m)$ is positive semi-definite and of rank one. In addition, according to Theorem 6.11, all the constraints in (6.23) are satisfied. Hence, m is a feasible solution to Problem 6.4.

On the other hand, let m be a feasible solution of Problem 6.4. Since $M_N(m)$ is positive semi-definite and of rank one, the moment sequence m is associated with a Dirac distribution. Assume the Dirac function is only non-zero at the point (Θ, λ), then, the transfer function $G(z)$ associated with parameter Θ is a solution to Problem 6.1. □

Since the rank condition $\text{rank}(M_N(m)) = 1$ is in general difficult to address directly, we rewrite Problem 6.4 as

$$
\begin{aligned}
m^* = \quad \arg\min_{m} \quad & \text{rank}(M_N(m)) && (6.24)\\
\text{s.t.} \quad & M_N(m) \succeq 0, && (6.25)\\
& \sum_\alpha p^{\text{CF}}_{k,\alpha} m^\alpha = 0, k = 0, \ldots, N_{\text{CF}}, && (6.26)\\
& \sum_\alpha p^{\text{NP}}_{k,\alpha} m^\alpha = 0, k = 0, \ldots, 2N_{\text{NP}} + 1, && (6.27)\\
& \sum_\alpha p^{S}_{k,\alpha} m^\alpha \geq 0, k = 1, \ldots, n + 2, && (6.28)\\
& \sum_\alpha p^{\text{DP}}_{k,\alpha} m^\alpha \geq 0, k = 1, \ldots, d_r + 1. && (6.29)
\end{aligned}
$$

It is easy to see (6.24) is equivalent to Problem 6.1 in the following sense.

Corollary 6.14 *There exists $G(z) \in \mathcal{G}(\rho, \Pi, n, m)$ to Problem 6.1 if and only if* (6.24) *has rank one solution.*

Remark 6.15 *Note that in* (6.23) *and* (6.24)*, the relaxation order N is fixed and is only determined by the highest total degree of all the polynomials p_0, p_k^{CF}, p_k^{NP}, p_k^S and p_k^{DP}. Hence, the value of N is known at a priori. There is no need to increase the value of N since we aim at finding rank one moment matrix.*

Although rank minimization is NP-hard, efficient convex relaxations are available. In particular, good approximate solutions can be obtained by using a log-det heuristic that relaxes rank minimization to a sequence of convex problems. Thus, we use the log-det heuristic algorithm proposed in [91], which has been proven to be efficient in practice, to solve problem (6.24), as summarized in Algorithm 6.3.

Algorithm 6.3 Rank Minimization

Set $X \leftarrow M(m)$, $X_0 \leftarrow I$, $\delta \leftarrow 0$, $k \leftarrow 0$.
repeat
Solve

$$X_{k+1} \leftarrow \begin{array}{ll} \arg\min & \mathrm{Tr}(X_k + \delta I)^{-1} X \\ \text{s.t.} & (6.25) \text{ - } (6.29). \end{array}$$

Decompose the symmetric matrix $X_k = T^{-1} D T$.
Set $\delta \leftarrow \min \mathrm{diag}(D) + \delta_0$.
Set $k \leftarrow k + 1$.
until a convergence criterion is reached.
return X_k

6.5 Application to Control System Problems

6.5.1 Fixed Order Stable System Identification

One of the most well-known applications of interpolation theory is system identification. In the literature, a model is said to be consistent with the experimental data if it satisfies interpolation conditions within a given noise level, as the measurements are always noisy in practice. For simplicity, we assume the time domain data corresponds to the impulse response. The case of general input is similar. In this way, the interpolation conditions

(6.8) and (6.9) become

$$\begin{aligned} g_k &= y_k - e_k^t, k = 1, \ldots, L_t, \quad (6.30) \\ G(e^{-j\omega_k}) &= F_k - e_k^f, k = 1, \ldots, L_f, \quad (6.31) \end{aligned}$$

where y_k is the impulse response of $G(z)$, e_k^f and e_k^t are measurement noise in frequency domain and time domain bounded by $\bar{e^f}$ and $\bar{e^t}$, respectively. Then, the corresponding polynomial constraints are a simple modification of (6.8) where we take noise into account as variables,

$$\begin{aligned} \Theta^T \phi_k + \Theta^T \Delta\eta_k &= y_k - e_k^t, \\ \bar{e^t} + e_k^t &\geq 0, \\ \bar{e^t} - e_k^t &\geq 0, \end{aligned}$$

for $k = 1, \ldots, L_t$ where

$$\Delta\eta_k = [v_{k-1}, \ldots, v_{k-n}, 0, \ldots, 0]^T,$$

and Θ and ϕ_k are defined in (6.11) and (6.10). The corresponding modifications to (6.9) are

$$|G^(e^{-j\omega_k}) - F_k| \leq \bar{e_f},$$

for $k = 1, \ldots, L_f$, which can be also converted to polynomial constraints as shown in [51, 52, 97]. Thus, in principle the algorithm proposed in Chapter 6.3 and 6.4 can be applied to identify fixed order stable systems with noisy data in time and frequency domain.

However, a direct application will lead to forbidden computational difficulty for large L_t and/or L_f. This is a consequence of the fact that the size of the moment matrix greatly increases with the number of added variables, i.e., noise terms; e.g. see [97]. To circumvent this difficulty, one may use results on sparse polynomial optimization to reduce computational burden by exploiting the inherent *running intersection property*. In addition, the idea on using rank minimization to increase computational efficiency introduced in

Chapter 6.4 can be applied as well, see [97] for details.

6.5.2 Fixed Order Controller Design

The interpolation approach in this chapter can also be applied to the design of fixed order stabilizing controllers. Consider a system as shown in Figure 6.1. Assume the controller

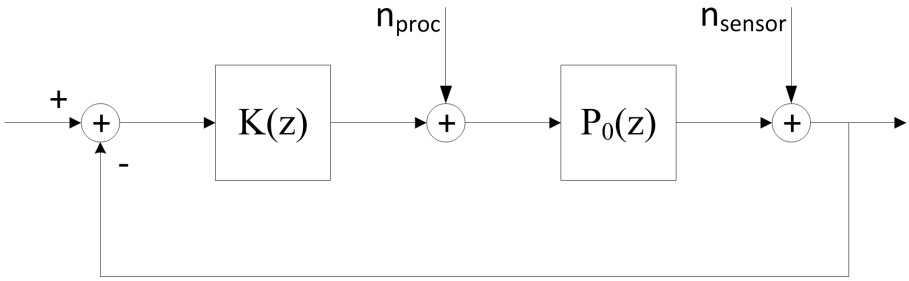

Figure 6.1: Fixed Order Controller Design

$K(z)$ has the form

$$K(z) = \frac{d(z)}{c(z)}$$

with fixed order (n, m). Let $p_1, \ldots, p_{\ell_1}$ be the unstable poles of $P_0(z)$ and let $z_1, \ldots, z_{\ell_2}$ be the non-minimum-phase zeros of $P_0(z)$. Then, the closed loop system

$$G_c(z) = P_0 K[1 + P_0 K]^{-1}$$

is internally stable if and only if

a) $G_c(z)$ is stable,

b) $G_c(p_k) = 1$ for $k = 1, \ldots, \ell_1$,

c) $G_c(z_k) = 0$ for $k = 1, \ldots, \ell_2$

Note that in this design problem, we require the closed-loop system $G_c(z)$ to be stable, whose coefficients are linear in the coefficients of $K(z)$. Hence, $G_c(z)$ is stable if and only if (6.15) holds by substituting a_is with polynomial functions of the coefficients of $c(z)$ and $d(z)$, which still leads to polynomial constraints. Moreover, condition (b) and (c) amount to $\ell_1 + \ell_2$

Nevanlinna-Pick interpolation conditions. Therefore, the controller design problem can be recast in the form of Problem 6.1 and solved by the algorithm developed on Chapter 6.3 and Chapter 6.4. Furthermore, additional constraints can be put on the $\mathcal{H}_\infty$ performance of the closed loop system using passivity.

6.5.3 Spectral Estimation

The problem of using interpolation theory to spectral estimation has been discussed in [43]. We restate a typical spectral estimation problem as the following. Consider a scalar zero-mean, stationary Gaussian stochastic process $\{y_i\}$ with $\Phi(e^{i\theta})$ being its power spectral density. Then,

$$\Phi(z) = f(z) + f(z^{-1})$$

where

$$f(z) = \frac{b(z)}{a(z)} = \frac{1}{2}f_0 + f_1 z^{-1} + \cdots .$$

The objective is to estimate Φ from a realization $\{y_i\}_{1\le i\le N}$ of the process. The interpolation approach introduced in [43] suggests using measurements $\{w_i\}_k$ after passing $\{y_i\}$ through a bank of stable filters

$$G_k(z) = \frac{1}{1 - p_k z^{-1}}, \quad k = 0, 1, \ldots, \ell,$$

and then constructs Nevanlinna-Pick interpolation constraints

$$f(p_k) = \frac{1}{2}(1 - p_k^2)c_k, \quad k = 0, 1, \ldots, \ell \tag{6.32}$$

where c_k is the estimated zeroth order covariance of $\{w_i\}_k$. Therefore, the spectral estimation problem with fixed degree constraint can be recast in the form of Problem 6.1, i.e., find $f(z) \in \mathcal{G}(1, \Pi)$ where

$$\Pi = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

that satisfies interpolation conditions (6.32). It should be emphasized that in [43], a complete parameterization has been given if $f(z)$ has degree at most ℓ. In our approach, the algorithm can be performed to find such a $f(z)$, if any, with degree $k \leq \ell$.

6.6 Numerical Results

In this section, we present two examples to illustrate the proposed algorithms.

6.6.1 Example 1: Fixed-order Stable System Identification

A third order system is considered with transfer function

$$G(z) = \frac{z^2 - 3z + 2}{z^3 - 0.6300z^2 - 0.7228z + 0.8043}.$$

This system is stable with H_∞ norm $\|G(z^{-1})\|_\infty = 58.32$ and root radius $\rho = 0.97$. To perform the identification process, the system was excited with $L_t = 100$ random input between -1 and 1 and process noise bounded by 0.2.

The identification is performed by using a priori information on stability and $\mathcal{H}_\infty$ norm. First we try to identify the system under the conditions that the the system is stable with root radius $\rho = 0.965$ and the $\mathcal{H}_\infty$ norm is less than 50. The algorithm proposed concludes there is no such a compatible system exists that is consistent to the measurement data. Then, the system is identified by assuming the system is stable with root radius $\rho = 0.97$ and the $\mathcal{H}_\infty$ norm is less than 60. The corresponding identified system is

$$\hat{G}(z) = \frac{0.9955z^2 - 2.9690z + 1.9690}{z^3 - 0.605z^2 - 0.7277z + 0.771}.$$

It can be verified that this identified system is stable with root radius 0.9685 and its $\mathcal{H}_\infty$ norm is 55.86.

6.6.2 Example 2: Stabilizing Controller Design

In this section, we consider a fixed order stable controller synthesis problem as indicated in Figure 6.2. More precisely, we consider an unstable system

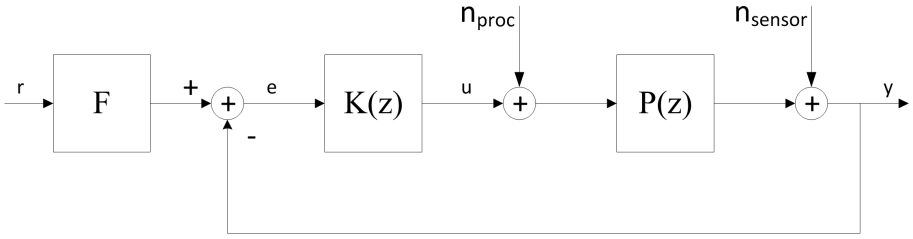

Figure 6.2: Stabilizing Controller Design with Fixed Order

$$P(z) = \frac{z(z-3)}{(z-2)(z+2)},$$

and aim at designing a forward gain F and a *first-order stable* controller,

$$K(z) = \frac{b_1 + b_0 z}{a_1 + z},$$

if any, such that

1) the closed loop system $G_c(z) = GK[1+GK]^{-1}$ is internally stable;

2) the DC gain from r to y is 1;

3) the controller $K(z)$ is stable;

4) the transfer function from r to controller output u is stable and has its $\mathcal{H}_\infty$ norm less than 2.

Base on the discussion in Chapter 6.5.2, this design problem can be formulated as a generalized interpolation problem of the form that is addressed in this chapter. Hence, the

problem is solved by applying Algorithm 6.3, which gives a feasible solution

$$\begin{aligned} F &= 0.0928, \\ K(z) &= -\frac{0.2361z + 1.417}{z}, \end{aligned}$$

that satisfies the required performance specifications.

6.7 Concluding Remarks

In this chapter, we provide an efficient numerical algorithm to address a generalized interpolation problem of finding a *fixed order* system in an allowable set that satisfies mixed domain interpolation conditions. Given interpolation points both in time-domain and frequency-domain, we aim at finding a fixed order plant that interpolates the given data and satisfies extra constraints such as stability and generalized passivity to the plant. By formulating conditions on interpolation and on the allowable model set as polynomial constraints, the problem is recast as a feasibility problem: finding a feasible point in a semi-algebraic set. Moreover, numerical algorithms based on results from polynomial optimization and rank minimization are proposed to efficiently solve this complex problem. Two numerical examples on identification and fixed-order controller design are provided to illustrate the efficiency of the proposed algorithms. It is of interest to improve the computational efficiency of the proposed algorithms and applying them to large size practical problems.

Chapter 7

Concluding Remarks

In this dissertation, several important yet difficult problems in system and control have been discussed. They are categorized into four different topics. The first one is about chance-constraint optimization and its applications in robust system design under the probabilistic framework. Under this topic, we study a class of robust design problems with polynomial dependence on the uncertainty. Under the probabilistic framework, a probabilistic relaxation of the original robust problem is adopted, thus requiring the satisfaction of the constraints not for all possible values of the uncertainty, but for most of them. As different from the randomized approach for tackling probabilistic relaxations, which is only guaranteed to provide soft bounds on the probability of satisfaction, we present a deterministic approach based on the novel concept of *kinship function*. This allows the development of an original framework, which leads to easily computable deterministic convex relaxations of the probabilistic problem. In particular, *optimal polynomial* kinship functions are introduced, which can be computed *a priori and once for all* and provide the "best convex bound" on the probability of constraint violation. More importantly, it is proven that the solution of the relaxed problem converges to that of the original robust optimization problem as the degree of the optimal polynomial kinship function increases. Furthermore, by relying on quadrature formulae for computation of integrals of polynomials, it is shown that the computational complexity of the proposed approach is polynomial in the number of uncertain parameters. Finally, unlike other deterministic approaches to robust polynomial

optimization, the number of variables in the ensuing optimization problem is not increased by the proposed approximation. An important feature of this approach is that a significant amount of the computational burden is "shifted" to a one-time offline computation whose results can be stored and provided to any end user.

It is noted that in chance-constraint optimization, a probability density function is always required for the uncertainty to impose probability constraints. Yet this piece of information is not necessarily available in practice. In such cases, the uniform distribution is commonly chosen as the probability distribution of the uncertainty. However, it is naturally to ask the question: can we justify the use of uniform distribution? As robustness is the most important concern, it is equivalent to ask if uniform distribution is the *worst-case distribution*, i.e., if it leads to the smallest probability guaranteeing robustness. Research in the literature has shown that uniform distribution is not always the worst-case distribution. Therefore, one of main objectives of this dissertation is to address the question: what is the *worst-case* performance on robustness for a given system over a class of admissible distributions on its uncertainty? While it is difficult to answer in general, we aim at finding the worst-case expected value of a specific set of semi-algebraic functions including polynomials as the worst-case performance. First, a sufficient condition is provided for the *uniformity principle* to hold, i.e., the worst-case expected value is achieved with the uniform distribution. In cases where this principle does not hold, it is shown that to answer this question, for the class of distributions considered, it suffice to solve a convex optimization problem for which efficient linear matrix inequality relaxations are available. Finally, the proposed approach is applied to estimate *hard bounds* of the worst-case probability of a semi-algebraic function being negative.

While the first topics are related to system design and analysis problems under probabilistic frameworks, hybrid system identification and generalized interpolation problems are discussed later in this dissertation. Though identification and interpolation are not directly related to the first two topics, the methods proposed in this dissertation share the same polynomial optimization technique on constructing convex approximations to solve the problems.

Under the topic of hybrid system identification, it is address to identify discrete time affine hybrid systems with noisy measurements. Given a finite number of measurements of input/output and a bound on the measurement noise, the objective is to identify a switching sequence and a set of affine models that are compatible with the a priori information, while minimizing the number of affine models. While this problem has been successfully addressed in the literature if the input/output data is noise-free or corrupted by process noise, results for the case of measurement noise are limited. In this dissertation, by recasting the identification problem as polynomial optimization, we develop randomized and deterministic algorithms to solve the problem, in which the inherent sparse structure is exploited to reduce the computational cost. For the randomized algorithm, we develop a hit-and-run type approach that leads to further computational complexity reduction, allowing for solving realistically sized problems. Alternatively, a finite dimensional semi-definite problem is given which is equivalent to the identification problem. Moreover, to address computational complexity issues, an equivalent rank minimization problem subject to deterministic LMI constraints is provided, as efficient convex relaxations for rank minimization are available in the literature.

In the last topic discussed in this dissertation, we address the problem of finding a fixed order plant that interpolates given data in time and frequency domain, and satisfies additional constraints such as stability and passivity. The comprehensive framework can be used to address a wide range of complex control problems such as system identification with a priori bound on its order, fixed-order controller design and spectral estimation. To solve the proposed interpolation problem, it is first shown that it can be recast as finding a point in a properly defined semi-algebraic set. Then, an efficient numerical algorithm based on convex relaxations of rank minimization is proposed to solve the problem.

While the proposed methods contributed by this dissertation have successfully demonstrated the efficiency on solving the problems addressed under various topics, they also bring opportunities for future research.

- In the topic of chance-constraint optimization, we propose the concept of scaler *kinship* function and provide algorithms to calculating the optimal polynomial kinship

functions. In further research, it is of interest to develop matrix kinship functions specically tailored to the case where the robust constraints are robust linear matrix inequalities with polynomial dependence on the uncertainty.

- The polynomial base used for calculating optimal kinship polynomials can lead to numerical difficulty if the degree of the desired kinship function is large. Thus, it is of interest to further develop similar theory and to improve the computational efficiency by using different polynomial bases.

- In the topic of distributional robustness, a specific class of distribution, as shown in Figure 4.1, is considered as the admissible distributions. In further research, it is of interest to extend the results contributed by this dissertation to other classes of distributions. For example, it is possible that the proposed approach can be applied to the extension of distribution class where symmetry requirement is eliminated.

- It is of interest to consider other polynomial basis to generate bounding polynomial approximations to the indicator function, to improve numerical stability.

- For the randomized algorithm proposed for the hybrid system identification problem, it is desired to have some stochastic analysis on the algorithm. For example, it is of interest to find a bound on the expected number of hit-and-run iterations needed for a given precision level.

- For the deterministic algorithm proposed for the hybrid system identification problem, it is of interest to further reduce the computational cost and to preserve the system structure information in the reformulated optimization problem.

- While the current relaxation technique, i.e., moment based methods and sum-of-square, used for polynomial optimization have been demonstrated to be successful and powerful for tackling and solving many difficult problems in system and control, the computational cost for solving the relaxations increases very fast if the problems are complex. In many problems, the structure of the reformulated polynomial optimization has very specific structures. Some of them, sparsity for example, are utilized

in the methods proposed in this dissertation to reduce the computational cost. Yet the structure information is far from being fully utilized. Thus, it is of great interest to find more computationally efficient methods by using the structure information.

Appendix A

Proof of Results in Chapter 3

A.1 Introduction to Smolyak Formulae

For every $j = 1, \ldots, d$, consider a sequence of one-dimensional quadrature formulae of increasing precision

$$\mathsf{Q}_j^{(i)}[g] \quad \left(\doteq \mathsf{Q}_{j,N_i}[g]\right) \quad = \sum_{k=1}^{N_i} \omega_{j,k}^{(i)}\, g(\theta_{j,k}^{(i)}), \quad i = 1, 2, \ldots \tag{A.1}$$

with nodes $\Theta_j^{(i)} \doteq \left\{\theta_{j,1}^{(i)}, \ldots, \theta_{j,N_i}^{(i)}\right\}$ and weights $\omega_{j,1}^{(i)} \cdots \omega_{j,N_i}^{(i)}$. Note that the number of nodes N_i in (A.1) is assumed to be a function of the precision index i. In order to apply our construction, we consider special sequences of quadrature formulae that should satisfy the following definition.

Definition A.1 (Smolyak sequences) *A sequence of quadrature formulae of the type (A.1) is said to be a* Smolyak sequence *if it satisfies the following properties*

(a) ***Nested nodes.*** *The nodes of the quadrature formula $\mathsf{Q}_j^{(i)}$ should contain all the nodes used by $\mathsf{Q}_j^{(i-1)}$; i.e., one has $\Theta_j^{(i-1)} \subset \Theta_j^{(i)}$, $i = 1, 2, \ldots$;*

(b) ***Precision.*** *The QF with precision index i should be exact for all polynomials of degree up to $2i-1$, that is $\nu_{i,j} \doteq \deg(\mathsf{Q}_j^{(i)}) \geq 2i-1$, $i = 1, 2, \ldots$;*

(c) ***Initial condition.*** *The quadrature formula* $\mathsf{Q}_j^{(1)}$ *is a one node formula, i.e.* $N_{1,j} = 1$ *and* $\mathsf{Q}_j^{(1)}[g] = 2g(0)$.

These sequences of quadrature formulae are at the basis of the method proposed by Smolyak in 1963 [74] for the construction of cubature rules with a low number of points. Particular sequences studied in the literature satisfying Definition A.1 are the Clenshaw-Curtis (CC) sequences, which are nested CC QFs (see [98]) and Kronrod-Patterson (KP) formulae, that represent particular nested version of classical Gauss formulae (see [99]). For a general overview on these sequences the reader is referred to the paper [100], where a so-called *Petras-delayed* (Pd) sequence is also presented.

These sequences are used to construct the so-called *Smolyak quadrature formula*

$$\mathsf{S}_d^{(\ell)}[g] = \sum_{\ell+1 \le \|\mathbf{i}\|_1 \le \ell+d} (-1)^{\ell+d-\|\mathbf{i}\|_1} \binom{d-1}{\|\mathbf{i}\|_1 - \ell - 1} \left(\mathsf{Q}_1^{(i_1)} \otimes \cdots \otimes \mathsf{Q}_d^{(i_d)}\right)[g], \tag{A.2}$$

where $\mathbf{i} \doteq [i_1 \ \cdots \ i_d]^\top$, $\mathbf{i} \in \mathbb{N}_+^d$ is the vector of precision indices for each dimension.

Looking at the formula, it is evident that it consists of a linear combination of product formulae involving only relatively-low precision QF, chosen so that the interpolation property for $d = 1$ is maintained for $d > 1$. As a consequence, the number of nodes used by $\mathsf{S}_d^{(\ell)}$ is expected to be low. In particular, the following lemma shows that the overall complexity of the Smolyak construction is *polynomial in* d, as long as the sequence $\mathsf{Q}^{(i)}$ satisfies property (iii).

Lemma A.2 (Polynomial complexity) *Assume that the sequence of QF in the Smolyak formula satisfies property (c) in Definition A.1. Then, for fixed* ℓ*, the number of nodes* $\mathsf{N}_d^{(\ell)}$ *depends polynomially in* d*, that is* $\mathsf{N}_d^{(\ell)} = \mathcal{O}\left(d^\ell\right)$. ⋆

Proof. A proof of slightly different versions of this lemma can be found in [101, 100]. □

In particular, if the sequences (CC)–(KP)–(Pd) are used, one can prove that $\mathsf{N}_d^{(\ell)} \approx \frac{2^\ell}{\ell!} d^\ell$. This result is confirmed by the numbers of nodes reported in Table A.1, for the case of Smolyak formulae constructed from Petras-delayed sequences.

Table A.1: Number of nodes $\mathsf{N}_d^{(\ell)}$ of Smolyak formula (using Petras-delayed sequences) for different values of d and ℓ.

$(d,\,\ell)$	3	4	5	6	7
3	39	87	135	207	399
4	81	193	385	641	1,217
5	151	391	903	1,743	3,343
6	257	737	1,889	4,161	8,481
7	407	1,303	3,655	8,975	19,855
8	609	2,177	6,657	17,921	43,137
9	871	3,463	11,527	33,679	87,823
10	1,201	5,281	19,105	60,225	169,185
11	1,607	7,767	30,471	103,247	310,927

As mentioned before, besides exhibiting polynomial dependence on d, Smolyak formulae enjoy also good properties in terms of degree of exactness, as shown in following lemma, which is adapted from [101] (for a formal proof, see also [102]).

Lemma A.3 (DoE of Smolyak formulae) *Assume that the sequence of QF used for constructing the Smolyak formula* (A.2) *satisfies property (b) in Definition A.1. Then* $\deg(\mathsf{S}_d^{(\ell)}) \geq 2\ell + 1$. $\star$

Note that Lemma A.3 states that the degree of exactness of $\mathsf{S}_d^{(\ell)}$ is *at least* $2\ell+1$. Indeed, it can be shown, [101, 100], that better results hold for the sequences (CC)–(KP)–(Pd).

Let us denote by Θ_k and w_k, $k = 1, \ldots, \mathsf{N}_d^{(\ell)}$ the (pre-computed) nodes and weights of Smolyak formula. Then, this formula can be written in a simple $\mathsf{N}_d^{(\ell)}$-points cubature form (3.22). Hence, as long as these nodes and weighs are computed *a priori* for given ℓ, d, formula (A.2) represents a multidimensional QF with a number of nodes[1] which increases polynomially with respect to d.

A.2 Proof of Proposition 3.9

If (3.13)-(3.16) holds, then $p(y)$ is convex and $p'(y)$ is non-decreasing in $[-1, \infty)$ since $p''(y)$ is non-negative. As $p'(y)$ is non-decreasing, $p'(y) \geq p'(-1) = 0$ for $y \in [-1, \infty)$. Hence, $p(y)$

[1] In computing the weights w_k, some of them may turn out to be zero, therefore leading to a formula with less than $\mathsf{N}_d^{(\ell)}$ nodes.

is also non-decreasing. Hence, $p(y)$ is non-negative since $p(y) \geq p(-1) = 0$ for $y \in [-1, \infty)$. Therefore, $p(y)$ is a kinship function according to Definition 3.4.

On the other hand, let the polynomial $p(y)$ of degree ϱ be an optimal polynomial kinship function, i.e., a solution of problem (3.11). By Definition 3.4, we have a) $p(0) = 1$; b) $p''(y) \geq 0$ for $y \in [-1, \infty)$ and c) $p(y) \geq 0$, $p'(y) \geq 0$, for $y \in [-1, \infty)$. Then, it can be seen that (3.13) and (3.16) are immediately satisfied. Moreover, if $p(-1) = t > 0$, then $\hat{p}(y) \doteq (p(y) - t)/(1 - t)$ is also a kinship function and

$$\int_{-1}^{0} \hat{p}(y)dy - \int_{-1}^{0} p(y)dy = \frac{t}{1-t} \int_{-1}^{0} (p(y) - 1)dy < 0,$$

which contradicts that $p(y)$ is the solution of the problem. Thus, we must have $p(-1) = 0$. Similarly, if $p'(-1) = t > 0$, define $\hat{p}(y) \doteq p(y) - t(1 + y) + t(1 + y)^{\varrho}$. It is not difficult to check that $\hat{p}(y)$ is a kinship function since it satisfies (3.13)-(3.16), i.e.,

$$\begin{aligned}
&\hat{p}(0) = p(0) = 1;\\
&\hat{p}'(-1) = p'(-1) - t = 0;\\
&\hat{p}(-1) = p(-1) = 0;\\
&\hat{p}''(y) = p''(y) + \varrho(\varrho - 1)t(1 + y)^{\varrho-2} \geq 0.
\end{aligned}$$

However, we have

$$\int_{-1}^{0} \hat{p}(y)dy - \int_{-1}^{0} p(y)dy = -\frac{t(\varrho - 1)}{2(\varrho + 1)} < 0,$$

which causes contradiction. Hence, $p'(-1) = h'(0) = 0$. Therefore, an optimal kinship polynomial function must satisfy (3.13)-(3.16). This concludes the proof. □

A.3 Proof of Theorem 3.12

First we show that, for sufficient large ϱ, the optimum x^* is always a feasible point for problem $(PO\kappa_{\varrho})$. Note that $f(x^*, q)$ is a polynomial, and that the roots of a non-trivial polynomial (with degree at least 1) is a null set in $\mathcal{Q}$ with respect to the measure induced

by the bounded pdf $\mu(q)$. Thus,

$$\lim_{\varrho\to\infty}\int_{\mathcal{Q}}\kappa_{\varrho}[f(x^*,q)]\mu(q)\mathrm{d}q \quad = \lim_{\varrho\to\infty}\int_{\{q\in\mathcal{Q}:f(x^*,q)<0\}}\kappa_{\varrho}[f(x^*,q)]\mu(q)\mathrm{d}q.$$

To this end, note that, for $y<0$, $\kappa_{\varrho}(y)$ is always bounded between 0 and 1, and $\int_{\mathcal{Q}}\mu(q)$ is also bounded. Hence, by the well-known dominated convergence theorem [103], one can exchange the order of limit and integral, obtaining

$$\lim_{\varrho\to\infty}\int_{\{q\in\mathcal{Q}:f(x^*,q)<0\}}\kappa_{\varrho}[f(x^*,q)]\mu(q)\mathrm{d}q \quad = \int_{\{q\in\mathcal{Q}:f(x^*,q)<0\}}\lim_{\varrho\to\infty}\kappa_{\varrho}[f(x^*,q)]\mu(q)\mathrm{d}q.$$

On the other hand, consider the particular polynomial kinship function $(1+y)^{\varrho}$ (which is not optimal). Then, by the definition of optimality, $\kappa_{\varrho}(\cdot)$ satisfies

$$\int_{-1}^{0}\kappa_{\varrho}(y)\mathrm{d}y \leq \int_{-1}^{0}(1+y)^{\varrho}\mathrm{d}y = \frac{1}{\varrho+1}$$

from which it immediately follows that

$$\lim_{\varrho\to\infty}\int_{-1}^{0}\kappa_{\varrho}(y)\mathrm{d}y = 0,$$

because we are integrating positive functions. Then, since $\kappa_{\varrho}(y)$ is non-negative, $\lim_{\varrho\to\infty}\kappa_{\varrho}(y) = 0$ almost everywhere on $[-1,0]$. Hence, as $\kappa_{\varrho}(y)$ is a non-decreasing function, it follows that the limit $\lim_{\varrho\to\infty}\kappa_{\varrho}(y)$ vanishes everywhere except at the point $y=0$. We conclude

$$\lim_{\varrho\to\infty}\int_{\mathcal{Q}}\kappa_{\varrho}[f(x^*,q)]\mu(q)\mathrm{d}q = 0,$$

which implies that the optimization problem $(PO\kappa_{\varrho})$ is always feasible for any $\epsilon>0$ for sufficient large ϱ.

Second, it is proven that the solution of problem $(PO_{\kappa_{\varrho}})$, x^*_{ϱ}, converges to the feasible set of problem (RO), which is given by

$$\mathcal{X}_f \doteq \{x\in\mathcal{X} : f(x,q)\leq 0 \text{ for all } q\in\mathcal{Q}\}.$$

Proceeding by contradiction, assume that this in not true. Then, there should exist $\delta > 0$ and a sub-sequence $\{\varrho_k\}$ such that for any k, $f(x^*_{\varrho_k}, q_k) > \delta$ for some q_k. Since $\mathcal{X}$ is compact and $f(x, q)$ is continuous in x, the coefficients of $f(\cdot, q)$ are uniformly bounded. Also note that $\mu(q)$ is positive over the compact set $\mathcal{Q}$. Therefore, there exists $\eta > 0$ such that, for all k,

$$\int_{\left\{q \in \mathcal{Q}: f(x^*_{\varrho_k}, q) > \frac{\delta}{2}\right\}} \mu(q) dq > \eta.$$

Thus,

$$\lim_{k\to\infty} \int_{\mathcal{Q}} \kappa_{\varrho_k}[f(x^*_{\varrho_k}, q)]\mu(q)\mathrm{d}q \geq \eta \lim_{k\to\infty} \inf_{y > \frac{\delta}{2}} \kappa_{\varrho_k}(y).$$

On the other hand, $\kappa'_\varrho(y)$ is increasing due to convexity. Thus,

$$\lim_{\varrho\to\infty} \kappa'_\varrho(0) \geq \lim_{\varrho\to\infty} \frac{\kappa_\varrho(0) - \kappa_\varrho(-h)}{h} = \frac{1}{h}$$

for arbitrary $h \in (0, 1)$. Hence, $\lim_{\varrho\to\infty} \kappa'_\varrho(0) = \infty$, which implies $\lim_{\varrho\to\infty} \kappa_\varrho(y) = \infty$ for $y > 0$. In turn, this would imply

$$\lim_{k\to\infty} \int_{\mathcal{Q}} \kappa_{\varrho_k}[f(x^*_{\varrho_k}, q)]\mu(q)\mathrm{d}q = \infty.$$

Then the optimization problem (3.10) would have no solution for some large ϱ, which is in contradiction with what we proved in the first part. Thus, we conclude that x^*_ϱ must converge to the feasible set of problem (RO).

Finally, we prove that x^*_ϱ indeed converges to x^*. In the first part, we proved that, for ϱ sufficient large, the point x^* is always feasible for (PO_{κ_ϱ}). Then, we have $c^\top x^*_\varrho \leq c^\top x^*$. On the other hand, in the second part we proved that x^*_ϱ converges to the feasible set of problem (RO). Formally, this means that there exist points $x_{f,\varrho}$ belonging to $\mathcal{X}_f$, such that x^*_ϱ gets arbitrary close to $x_{f,\varrho}$ as ϱ increases. Hence, for any $\epsilon > 0$, there should exist a R_ϵ such that

$$|c^\top x^*_\varrho - c^\top x_{f,\varrho}| \leq \epsilon \text{ for } \varrho > R_\epsilon. \tag{A.3}$$

Moreover, since $x_{f,\varrho}$ is feasible for the robust problem (RO), and x^* is the solution of (RO), it follows that $c^\top x^* \leq c^\top x_{f,\varrho}$. This, together with (A.3) implies

$$c^\top x^* \geq c^\top x_\varrho^* \geq c^\top x_{f,\varrho} - \epsilon \geq c^\top x^* - \epsilon.$$

As ϵ is arbitrary, we have

$$\lim_{\varrho \to \infty} c^\top x_\varrho^* = \lim_{\varrho \to \infty} c^\top x_{f,\varrho} = c^\top x^*.$$

Finally, since x^* is assumed unique and $x_{f,\varrho}$ is feasible for problem (RO), it should hold that $\lim_{\varrho \to \infty} x_{f,\varrho} = x^*$. This, in turn, implies that

$$\lim_{\varrho \to \infty} x_\varrho^* = x^*,$$

thus concluding the proof. □

Appendix B

Proof of Results in Chapter 4

B.1 Proof of Theorem 4.7

To prove the theorem, we first need to define a subclass of the admissible distribution class $\mathcal{F}$. That is, given a vector $t \in [0,1]^d$, the corresponding *truncated uniform distribution* $u_t(q)$ is defined as the following.

Definition B.1 *Given* $t = (t_1, t_2, \ldots, t_d) \in [0,1]^d$, $u_t(q) \doteq u_1(q_1) \cdots u_d(q_d)$ *is called a truncated uniform distribution if* $u_k(q_k)$ *satisfies the following conditions for* $k = 1, \ldots, d$.

(1) if $t_k > 0$, $u_k(q_k) = \frac{1}{2t_k}$ *for* $q_k \in [-t_k, t_k]$;

(2) if $t_k = 0$, $u_k(q_k) = \delta(q_k)$;

(3) $u_k(q_k) = 0$ *otherwise.*

$\star$

Then, the class of all truncated uniform distributions is defined as

$$\mathcal{U}_T \doteq \{u_t : t \in [0,1]^d\}.$$

Examples of truncated uniform distributions are illustrated in Figure B.1.

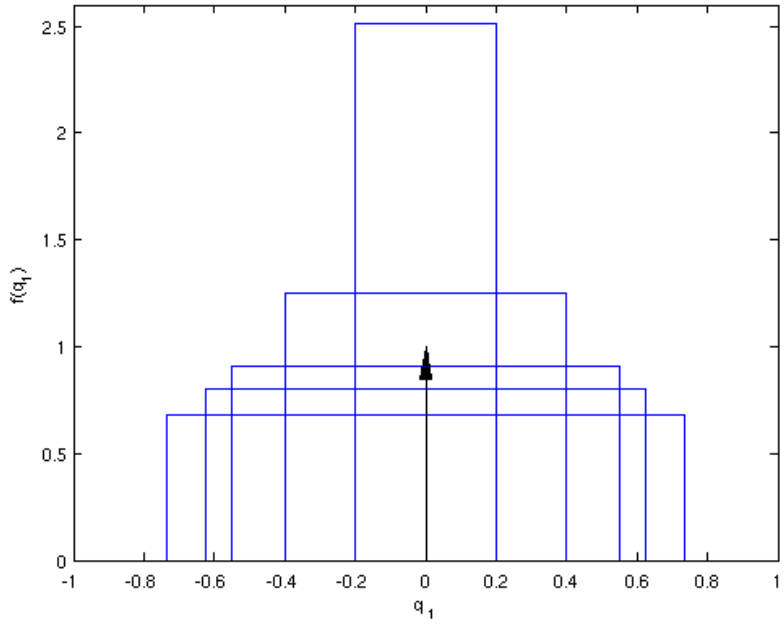

Figure B.1: Truncated Uniform Probability Density Functions

Now, we are ready to introduce the following lemma, which is a small modification of the so-called *Truncated Uniformity Principle* in [22, 27]. The proof is very similar and omitted here.

Lemma B.2 (Truncated Uniformity Principle) Given any continuous function $g(q)$, it follows that

$$\sup_{f\in\mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{f\in\mathcal{U}_T} \mathcal{E}_f[g(q)].$$

□

Since every well-defined semi-algebraic function $g(q) \in \mathcal{A}$ is continuous on $[-1,1]^d$, it follows that

$$\sup_{f\in\mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{t\in[0,1]^d} \frac{1}{2^d t_1\cdots t_d} \int_{-t_1}^{t_1}\cdots\int_{-t_d}^{t_d} g(q_1,\ldots,q_d)dq_d\cdots dq_1.$$

To prove the uniformity principle, we need only to prove that the above integral takes supremum at $t_i = 1$ for any i when all the other $t_j, j \neq i$ are fixed. Without loss of generality, we assume that $t_2, \ldots, t_d$ are fixed and define

$$h(\xi) \doteq \frac{1}{2^{d-1} t_2 \cdots t_d} \int_{-t_2}^{t_1} \cdots \int_{-t_d}^{t_d} g(\xi, q_2, \ldots, q_d) dq_d \cdots dq_2,$$

and

$$\tau_1(x, q) \doteq g(x, q_2, \ldots, q_d) + g(-x, q_2, \ldots, q_d).$$

Note that

$$\phi(t_1) \doteq \mathcal{E}_{u_t}[g(q)]) = \frac{1}{2t_1} \int_{-t_1}^{t_1} h(\xi) d\xi = \frac{1}{2t_1} \int_0^{t_1} (h(\xi) + h(-\xi)) d\xi.$$

Now we prove that if $\tau_1(x, q_2, \ldots, q_d)$ is non-decreasing for any fixed $-1 \leq q_i \leq 1, i = 2, \ldots, d$, $\phi(t_1)$ takes its supremum at $t_1 = 1$. In fact, if this assumption is true, $h(\xi) + h(-\xi)$ is non-decreasing on $\xi \in [0, 1]$. Hence, the average, $\phi(t_1)$, is maximized when t_1 is maximal; i.e., $t_1 = 1$ maximizes $\phi(t_1)$.

To this end, we have proven that the uniformity principle holds if for any $1 \leq k \leq d$, $\tau_k(x, q)$ is non-decreasing for $x \in [0, 1]$ when $q_2, \ldots, q_d$ are fixed. This is equivalent to the condition that $\omega_k(x_1, x_2, q) \geq 0$ for all $q \in [-1, 1]^d$ and for all (x_1, x_2) that $0 \leq x_1 \leq x_2 \leq 1$. Note that $\omega_k(x_1, x_2, q)$ belongs to $\mathcal{A}$. Hence, it has a BSAL. According to Proposition 8 in [83], positivity of $\omega_k(x_1, x_2, q)$ is equivalent to the optimization problem (4.9).

B.2 Proof of Lemma 4.10

Denote the joint p.d.f. of (q, z) by $\tilde{f}_{(q,z)}(q, z)$ and define its marginal density function on q as $f_q(q) \doteq \int \tilde{f}_{(q,z)}(q, z) dz$. Moreover, define a special class of probability density functions $\tilde{\mathcal{F}}$ supported on $\mathcal{K}$ as

$$\tilde{\mathcal{F}} \doteq \{\tilde{f}_{(q,z)}(q, z) : f_q(q) \in \mathcal{U}_t; \tilde{f}_{(q,z)}(q, z) = 0 \text{ for } (q, z) \notin \mathcal{K}\}.$$

Then, for every $\tilde{f}_{(q,z)}(q,z) \in \tilde{\mathcal{F}}$, the marginal $f_q(q)$ is a truncated uniform distribution, i.e., $f_q(q) = u_t(q)$ for some $t \in [0,1]^d$. Hence, the polynomial dependence introduced in the definition of $\mathcal{K}$ guarantees that, for any $\tilde{f} \in \tilde{\mathcal{F}}$,

$$\mathcal{E}_{\tilde{f}}[z_m] = \mathcal{E}_{u_t}[g(q)].$$

By the truncated uniformity principle introduced in Lemma B.2, i.e.,

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{f \in \mathcal{U}_T} \mathcal{E}_f[g(q)],$$

it follows that

$$\sup_{\tilde{f}_{(q,z)} \in \tilde{\mathcal{F}}} \mathcal{E}_{\tilde{f}}[z_m] \leq \sup_{f \in \mathcal{U}_T} \mathcal{E}_f[g(q)] = \sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)].$$

On the other hand, for every fixed $f_q(q) \in \mathcal{U}_t$, there exists a total distribution $\tilde{f}_{(q,z)} \in \tilde{\mathcal{F}}$ with its marginal being $f_q(q)$. Moreover, it is *uniquely* determined by the polynomial dependency that defines $\mathcal{K}$. Hence,

$$\sup_{f \in \mathcal{U}_T} \mathcal{E}_f[g(q)] \leq \sup_{\tilde{f}_{(q,z)} \in \tilde{\mathcal{F}}} \mathcal{E}_{\tilde{f}}[z_m].$$

Therefore,

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{\tilde{f}_{(q,z)} \in \tilde{\mathcal{F}}} \mathcal{E}_{\tilde{f}}[z_m]. \tag{B.1}$$

Now one can see that to compute the worst-case expected value of $g(q)$, it suffices to characterize the moment set of all possible p.d.f. that belong to $\tilde{\mathcal{F}}$, i.e.,

$$\sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)] = \sup_{\Gamma_q m \in \mathcal{M}_q^{\mathcal{U}_T}, m \in \mathcal{M}_{\mathcal{K}}} m_{0\cdots 1}, \tag{B.2}$$

where $m_{0\cdots 1}$ is an element in the moment matrix that corresponds to the expected value of

z_m. Thus, to finish the proof of this lemma, it suffice to prove that

$$\sup_{\Gamma_q m \in \mathcal{M}_q^{\mathcal{U}_T}, m \in \mathcal{M}_{\mathcal{K}}} m_{0\cdots 1} \geq \sup_{\Gamma_q m \in \mathrm{conv}(\mathcal{M}_q^{\mathcal{U}_T}), m \in \mathcal{M}_{\mathcal{K}}} m_{0\cdots 1},$$

since the other direction of the above inequality is alway preserved. In fact, for any m that $\Gamma_q m \in \mathcal{M}_q^{\mathcal{U}_T}$, the corresponding marginal density distribution $f_q(q)$ satisfies $f_q \in \mathcal{F}$. Hence, the polynomial dependency in the definition of $\mathcal{K}$ guarantees that

$$\sup_{m \in \mathrm{conv}(\mathcal{M}^{\mathcal{U}_t}) \cap \mathcal{M}_{\mathcal{K}}} m_{0\cdots 1} \leq \sup_{f \in \mathcal{F}} \mathcal{E}_f[g(q)].$$

This, combing with (B.2), concludes the proof.

B.3 Proof of Theorem 4.12

To prove this theorem, we first introduce the following lemma.

Lemma B.3 There exists a surjective linear transformation $\Pi : \mathcal{M}_{[-1,1]^d} \to \mathrm{conv}(\mathcal{M}^{\mathcal{U}_T})$ that

$$\Pi(m) = \Phi m,$$

where Φ is a diagonal matrix with elements

$$\phi_{\mathbf{i},\mathbf{i}} = \begin{cases} \frac{1}{(i_1+1)\cdots(i_d+1)} & \text{, if } i_1, \ldots, i_d \text{ are even;} \\ 0 & \text{, otherwise.} \end{cases}$$

Proof: For any moment sequence $m^t \doteq \{m^t_{\mathbf{i}}\}_{\mathbf{i}=1}^{\infty} \in \mathcal{M}^{\mathcal{U}_T}$, we have

$$\begin{aligned} m^t_{\mathbf{i}} &= \mathcal{E}_{u_t}(q^{\mathbf{i}}) \\ &= \frac{1}{2^d t_1 \cdots t_d} \int_{-t_d}^{t_d} \cdots \int_{-t_1}^{t_1} q_1^{i_1} \cdots q_d^{i_d} dq_1 \cdots dq_d \\ &= \frac{(t_1^{i_1+1} - (-t_1)^{i_1+1}) \cdots (t_d^{i_d+1} - (-t_d)^{i_d+1})}{(i_1+1) \cdots (i_d+1) t_1 \cdots t_d} \\ &= \begin{cases} \frac{t_1^{i_1} \cdots t_d^{i_d}}{(i_1+1)\cdots(i_d+1)} & \text{, if } i_1, \ldots, i_d \text{ are even;} \\ 0 & \text{, otherwise} \end{cases} \end{aligned}$$

Therefore, m^t can be represented by a (surjective) linear transformation of the moment sequence of a Dirac distribution on $[0,1]^d$. Hence, for any $\bar{m}^t \in \text{conv}(\mathcal{M}^{\mathcal{U}_T})$, it also can be represented by the linear transformation of moments for a convex combination of Dirac distributions in $[0,1]^d$, which is in the set $\mathcal{M}_{[-1,1]^d}$; e.g., see [104, 7]. □

Since the linear transformation is surjective, the above lemma implies that the convex hull $\text{conv}(\mathcal{M}^{\mathcal{U}_T})$ is indeed identical to the linear transformation, denoted by $\Pi\mathcal{M}_{[-1,1]^d}$, of the set $\mathcal{M}_{[-1,1]^d}$. Moreover, Lemma 4.10 guarantees that the worst-case expected value of $g(q)$ remains the same if we substitute $\mathcal{M}^{\mathcal{U}_T}$ by its convex hull. By combing these lemmas, we conclude the proof of Theorem 4.12.

B.4 Proof of Theorem 4.14

According to the known Stone-Weierstrass approximation theorem [85], the continuous function g can be approximated uniformly by polynomials, i.e. for any $\epsilon > 0$, there is a polynomial p that $\|p - g\|_\infty < \epsilon$. Hence,

$$|\mathcal{E}_f p - \mathcal{E}_f g| \leq \int_{\mathcal{Q}} |(p-g)f| \leq \|p-g\|_\infty \cdot \|f\|_1 < \epsilon,$$

where $\|\cdot\|_\infty$ and $\|\cdot\|_1$ denote $\mathcal{L}_\infty$ norm and $\mathcal{L}_1$ norm of functions defined on $\mathcal{Q}$, respectively. Moreover, suppose $\sup_{f\in\mathcal{F}} \mathcal{E}_f p = \mathcal{E}_{f_1} p$ and $\sup_{f\in\mathcal{F}} \mathcal{E}_f g = \mathcal{E}_{f_2} g$, then

$$\begin{aligned} \mathcal{E}_{f_1} g \;\leq\; \mathcal{E}_{f_2} g \;&\leq\; \mathcal{E}_{f_2} p + \epsilon \leq \mathcal{E}_{f_1} p + \epsilon \\ &\leq \mathcal{E}_{f_1} g + 2\epsilon. \end{aligned}$$

Hence,

$$\begin{aligned} \left| \sup_{f\in\mathcal{F}} \mathcal{E}_f p - \sup_{f\in\mathcal{F}} \mathcal{E}_f g \right| \;&=\; |\mathcal{E}_{f_1} p - \mathcal{E}_{f_1} g + \mathcal{E}_{f_1} g - \mathcal{E}_{f_2} g| \\ &<\; \epsilon + 2\epsilon = 3\epsilon. \end{aligned}$$

□

Appendix C

Proof of Results in Chapter 5

C.1 Proof of Theorem 5.9

First note that there exists some unit vector x and admissible noise e such that

$$V_s(\mathbf{r}, e)x = 0$$

if and only if

$$\sum_{i=1}^{L} x^T v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)x = 0. \tag{C.1}$$

Since $v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)$ is positive semi-definite, it is equivalent to finding a unit vector x^* that the minimum of the following problem

$$\begin{aligned} \min_{e} \quad & (x^*)^T \left[\sum_{i=1}^{L} v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e) \right] x^* \\ \text{s.t.} \quad & e \in K \end{aligned} \tag{C.2}$$

is zero, where K is defined as

$$K \doteq \{e : p_k \doteq \bar{e}^2 - e_k^2 \geq 0, k = 1, \ldots, L + n\}.$$

Note that the objective function in (C.2) is a SOS in terms of e for any fixed x, hence, it can be represented in the form of (2.17). Moreover, the running intersection property is satisfied for (C.2). Therefore, by Theorem 2.11, it follows that the minimum of the following problem

$$\begin{aligned} \min_{m} \quad & (x^*)^T Q(m) x^* \\ \text{s.t.} \quad & M_N(m, I_k) \succeq 0, k = 1, \ldots, L, \\ & M_{N_i}(p_i m, I_{k(i)}) \succeq 0, i = 1, \ldots, L+n, \end{aligned} \tag{C.3}$$

is zero, where $N = s$, $N_i = s-1$ for all i, and Q is defined in (5.11). Since $v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)$ is positive semi-definite, $Q(m)$ is positive semi-definite as well.[1] Hence, we have $(x^*)^T Q(m) x^* = 0$ for some unit vector x^* if and only if the matrix $Q(m)$ is rank deficient.

Conversely, if the rank condition holds, one can always extract a moment sequence m^* corresponding to a probability measure with point support, such that $Q(m^*)$ is rank deficient. Hence, there is an admissible noise e that $\sum_{i=1}^{L} v_s(\mathbf{r}_i, e)^T v_s(\mathbf{r}_i, e)$ is rank deficient. This implies the Veronese matrix is rank deficient for the same noise e, which concludes the proof. □

[1] Note that $Q(m)$ is p.s.d. since m satisfies the LMI constraints in (C.3). This comes directly from the fact that the finite moment condition is the dual formulation of the related SOS problem.

Bibliography

[1] Boyd, S., Ghaoui, El, Feron, E., and Balakrishnan, V., *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *SIAM Studies in Applied Mathematics*, SIAM, Philadelphia, PA, ISBN 0-89871-334-X, 1994

[2] Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V., *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, 1994

[3] Ben-Tal, A. and Nemirovski, A., "Robust Convex Optimization," *Mathematics of Operations Research*, volume 23, pp. 769–805, 1998

[4] Blondel, Vincent D. and Tsitsiklis, John N., "A survey of computational complexity results in systems and control," *Automatica*, volume 36, no. 9, pp. 1249–1274, ISSN 0005-1098, 2000

[5] Murty, K.G. and Kabadi, S.N., "Some NP-complete problems in quadratic and nonlinear programming," *Mathematical Programming*, volume 39, pp. 117–129, 1987

[6] Parrilo, P.A., *Structural semidefinite programs and semialgebraic geometry methods in robustness and optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, California, 2000

[7] Lasserre, Jean B., "Global optimization with polynomials and the problem of moments," *SIAM J Optim*, volume 11, no. 3, pp. 796–817, ISSN 1052-6234, 2001

[8] Lavaei, J. and Aghdam, A.G., "Robust Stability of LTI Systems Over Semialgebraic Sets Using Sum-of-Squares Matrix Polynomials," *Automatic Control IEEE Transactions on*, volume 53, no. 1, pp. 417–423, ISSN 0018-9286, 2008

[9] Henrion, D., Arzelier, D., Peaucelle, Dimitri, and Lasserre, J.-B., "On parameter-dependent Lyapunov functions for robust stability of linear systems," *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 1, ISSN 0191-2216, Dec.

[10] Prajna, S., Papachristodoulou, A., and Wu, Fen, "Nonlinear control synthesis by sum of squares optimization: a Lyapunov-based approach," *Control Conference, 2004. 5th Asian*, volume 1, July

[11] Barmish, B.R., *New Tools for Robustness of Linear Systems*, MacMillan, New York, 1994

[12] Barmish, B.R. and Kang, H.I., "A survey of extreme point results for robustness of control systems," *Automatica*, volume 29, pp. 13–35, 1993

[13] El Ghaoui, L. and Lebret, H., "Robust solutions to uncertain semidefinite programs," *SIAM J Optim*, volume 9, no. 1, pp. 33–52, 1998

[14] Ben-Tal, A. and Nemirovski, A., "On Tractable Approximations of Uncertain Linear Matrix Inequalities Affected by Interval Uncertainty," *SIAM Journal on Optimization*, volume 12, no. 3, pp. 811–833, 2002

[15] Calafiore, G. and Dabbene, F., editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, Springer-Verlag, London, 2006

[16] Tempo, R., Calafiore, G., and Dabbene, F., *Randomized Algorithms for Analysis and Control of Uncertain Systems*, Communications and Control Engineering Series, Springer-Verlag, London, 2005

[17] Calafiore, G. and Polyak, B.T., "Stochastic Algorithms for Exact and Approximate Feasibility of Robust LMIs," *IEEE Transactions on Automatic Control*, volume 46, pp. 1755–1759, 2001

[18] Polyak, B.T. and Tempo, R., "Probabilistic Robust Design with Linear Quadratic Regulators," *Systems Control Letters*, volume 43, pp. 343–353, 2001

[19] Calafiore, G. and Campi, M.C., "Sampled Convex Programs and Probabilistically Robust Design," *Probabilistic and Randomized Methods for Design under Uncertainty*, Springer-Verlag, London, pp. 161–188, 2006

[20] Alamo, T., Tempo, R., and Luque, A., "On the Sample Complexity of Probabilistic Analysis and Design Methods," *Perspectives in Mathematical System Theory, Control, and Signal Processing*, edited by S. Hara, Y. Ohta, and J. C. Willems, Springer Berlin / Heidelberg, pp. 39–50, 2010

[21] Alamo, T., Tempo, R., and Camacho, E.F., "Randomized Strategies for Probabilistic Solutions of Uncertain Feasibility and Optimization Problems," *IEEE Transactions on Automatic Control*, volume 54, no. 11, pp. 2545 –2559, 2009

[22] Barmish, B.R. and Lagoa, C.M., "The Uniform Distribution: A Rigorous Justification for its use in Robustness Analysis," *Mathematics of Control Signals and Systems*, volume 10, pp. 203–222, 1997

[23] Bai, E.-W., Tempo, R., and Fu, M., "Worst-Case Properties of the Uniform Distribution and Randomized Algorithms for Robustness Analysis," *Mathematics of Control Signals and Systems*, volume 11, pp. 183–196, 1998

[24] Kettani, H. and Barmish, B.R., "A new Monte Carlo circuit simulation paradigm with specific results for resistive networks," *IEEE Trans Circuits Syst I*, volume 53, no. 6, pp. 1289–1299, ISSN 1549-8328, 2006

[25] Ross, S.R. and Barmish, B.R., "A Sharp Estimate for the Probability of Stability for Polynomials With Multilinear Uncertainty Structure," *IEEE Trans Automat Contr*, volume 53, no. 2, pp. 601–606, March 2008

[26] Lagoa, C., Dabbene, F., and Tempo, R., "Hard Bounds on the Probability of Performance with Application to Circuit Analysis," *IEEE Transactions on Circuits and Systems I*, volume to appear, 2008

[27] Barmish, B. R., "A probabilistic robustness result for a multilinearly parameterized H_∞ norm," *Proc. American Control Conference*, 28–30 June 2000

[28] Barmish, B.R., Lagoa, C.M., and Tempo, R., "Radially truncated uniform distributions for probabilistic robustness of control systems," *Proceedings of the American Control Conference*, 1997

[29] Zhang, J., Lagoa, C.M., and Barmish, B.R., "Probabilistic robustness: an RLC circuit realization of the truncation phenomenon," *Proc. American Control Conference*, 1998

[30] Paoletti, S., Juloski, A., Ferrari-trecate, G., and Vidal, R., "Identification of hybrid systems: a tutorial," *European Journal of Control*, volume 13, no. 2, pp. 123–130, 2007

[31] Ma, Yi and Vidal, René, "Identification of Deterministic Switched ARX Systems via Identification of Algebraic Varieties," *Hybrid Systems: Computation and Control*, 2005

[32] Vidal, R., Soatto, S., Ma, Y., and Sastry, S., "An algebraic geometric approach to the identification of a class of linear hybrid systems," *Proceedings of the IEEE Conference on Decision and Control*, 2003

[33] Roll, J., Bemporad, A., and Ljung, L., "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, volume 40, no. 1, pp. 37 – 50, ISSN 0005-1098, 2004

[34] Bemporad, A., Garulli, A., Paoletti, S., and Vicino, A., "A bounded-error approach to piecewise affine system identification," *IEEE Trans Automat Contr*, volume 50, no. 10, pp. 1567–1580, 2005

[35] Ozay, N., Sznaier, M., Lagoa, C., and Camps, O., "A sparsification approach to set membership identification of a class of affine hybrid systems," *Proceedings of the IEEE Conference on Decision and Control*, 2008

[36] Ozay, N., Lagoa, C., and Sznaier, M., "Robust identification of switched affine systems via moments-based convex optimization," *Proceedings of the IEEE Conference on Decision and Control*, 2009

[37] Ozay, N., Sznaier, M., Lagoa, C., and Camps, O., "GPCA with denoising: a moment-based convex approach," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010

[38] Lasserre, Jean B., "Global optimization with polynomials and the problem of moments," *SIAM J Optim*, volume 11, no. 3, pp. 796–817, ISSN 1052-6234, 2001

[39] Lasserre, Jean B., "Convergent SDP-Relaxations in Polynomial Optimization with Sparsity," *SIAM J on Optimization*, volume 17, no. 3, pp. 822–843, ISSN 1052-6234, 2006

[40] Waki, H., Kim, S., Kojima, M., and Muramatsu, M., "Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with Structured Sparsity," *SIAM Journal on Optimization*, volume 17, no. 1, pp. 218–242, 2006

[41] Sarason, D., "Generalized interpolation in H_∞," *Trans Amer Math Soc*, volume 127, pp. 179–203, 1967

[42] Walsh, J. L., *Interpolation and approximation by rational functions in the complex domain*, volume 20, Amer. Math. Soc. Colloquium Publications, 1956

[43] Byrnes, C. I., Georgiou, T. T., and Lindquist, A., "A generalized entropy criterion for Nevanlinna-Pick interpolation: A convex optimization approach to certain problems in systems and control," *IEEE Trans Automat Contr*, volume 45, no. 6, pp. 822–839, 2001

[44] Schur, I., "On power series which are bounded in the interior of the unit circle I and II," *J fur die reine und angewandte Mathematik*, volume 148, pp. 122–145, 1918

[45] Bynes, C., Gusev, S., and Lindquist, A., "A convex optimization approach to the rational covariance externsion problem," *SIAM J Control Optim*, volume 37, no. 1, pp. 211–229, 1998

[46] Inanc, T., Sznaier, M., Parrilo, P. A., and Pena, R. S., "Robust identification with mixed parametric/nonparametric models and time/frequency-domain experiments: theory and an application," *IEEE Trans Contr Syst Technol*, volume 9, no. 4, pp. 608–617, 2001

[47] Parrilo, P.A., Sanchez Pena, R.S., and Sznaier, M., "A parametric extension of mixed time/frequency robust identification," *IEEE Trans Automat Contr*, volume 44, no. 2, pp. 364 –369, 1999

[48] Parrilo, P.A., Sznaier, M., Pena, R., and Inanc, T., "Mixed Time/Frequency-Domain Based Robust Identification," *Automatica*, volume 34, no. 11, pp. 1375 – 1389, 1998

[49] Georgiou, T.T. and Lindquist, A., "Remarks on control design with degree constraint," *IEEE Trans Automat Contr*, volume 51, no. 7, pp. 1150–1156, 2006

[50] Takyar, M.S., Nasiri-Amini, A., and Georgiou, T.T., "Weight Selection in Feedback Design with Degree Constraints," *IEEE Trans Automat Contr*, volume 53, no. 8, pp. 1951–1955, 2008

[51] Cerone, Vito, Piga, Dario, and Regruto, Diego, "Bounding the parameters of linear systems with stability constraints," *Proc. American Control Conference*, 2010

[52] Cerone, Vito, Piga, Dario, and Regruto, Diego, "Set-Membership EIV Identification through LMI Relaxation Techniques," *Proc. American Control Conference*, 2010

[53] Laurent, M., "Sums of squares, moment matrices and optimization over polynomials," *Emerging applications of algebraic geometry*, volume 149, pp. 157–270, 2009

[54] Parrilo, P. A., "Semidefinite programming relaxations for semialgebraic problems," *Math Program*, volume 96, no. 2, Ser. B, pp. 293–320, ISSN 0025-5610, 2003

[55] Calafiore, G., Topcu, U., and El Ghaoui, L., "Parameter estimation with expected and residual-at-risk criteria," *Systems Control Letters*, volume 58, no. 1, pp. 39–46, ISSN 0167-6911, 2009

[56] Ghaoui, L. El, Oks, M., and Oustry, F., "Worst-Case Value-At-Risk and Robust Portfolio Optimization: A Conic Programming Approach," *Operations Research*, volume 51, no. 4, pp. 543–556, 2003

[57] Ben-Tal, A. and Nemirovski, A., "Robust Truss Topology Design via Semidefinite Programming," *SIAM Journal on Optimization*, volume 7, no. 4, pp. 991–1016, 1997

[58] Mutapcic, A., Kim, S.-J., and Boyd, S., "Beamforming With Uncertain Weights," *IEEE Signal Processing Letters*, volume 14, pp. 348–351, May 2007

[59] Ben-Tal, A., Golany, B., Nemirovski, A., and Vial, J. P., "Retailer-Supplier Flexible Commitments Contracts: A Robust Optimization Approach," *Manufacturing Service Operations Management*, volume 7, no. 3, pp. 248–271, ISSN 1526-5498, 2005

[60] Xu, Y., Hsiung, K.-L., Li, X., Nausieda, I., Boyd, S., and Pileggi, L., "OPERA: optimization with ellipsoidal uncertainty for robust analog IC design," *Proceedings of Design Automation Conference*, 2005

[61] Chesi, G., Garulli, A., Tesi, A., and Vicino, A., "Solving quadratic distance problems: an LMI-based approach," *IEEE Transactions on Automatic Control*, volume 48, pp. 200–212, 2003

[62] Lasserre, Jean B., "Robust global optimization with polynomials," *Math Program*, volume 107, no. 1-2, Ser. B, pp. 275–293, 2006

[63] Scherer, C. W. and Hol, C. W. J., "Matrix sum-of-squares relaxations for robust semidefinite programs," *Math Program*, volume 107, no. 1-2, Ser. B, pp. 189–211, ISSN 0025-5610, 2006

[64] Prékopa, András, *Stochastic Programming*, volume 324 of *Mathematics and its Applications*, Kluwer Academic Publishers Group, Dordrecht, ISBN 0-7923-3482-5, 1995

[65] Calafiore, G. and Campi, M.C., "The scenario approach to robust control design," *IEEE Trans Automat Contr*, volume 51, no. 5, pp. 742–753, May 2006

[66] Barmish, B.R., Shcherbakov, P.S., Ross, S.R., and Dabbene, F., "On Positivity of Polynomials: The Dilation Integral Method," *IEEE Transactions on Automatic Control*, volume 54, no. 5, pp. 965–978, ISSN 0018-9286, 2009

[67] Keel, L.H. and Bhattacharyya, S.P., "A linear programming approach to controller design," *Proceedings of the IEEE Conference on Decision and Control*, 1997

[68] Feng, Chao and Lagoa, Constantino M., "Distributional Robustness Analysis for Polynomial Uncertainty," *Proceedings of the IEEE Conference on Decision and Control*, 2009

[69] Boyd, S. and Vandenberghe, L., *Convex Optimization*, Cambridge University Press, Cambridge, ISBN 0-521-83378-7, 2004

[70] Todd, M.J., "Semidefinite optimization," *Acta Numerica*, volume 10, pp. 515–560, 2001

[71] Vandenberghe, L. and Boyd, S., "Semidefinite programming," *SIAM Review*, volume 38, pp. 49–95, 1996

[72] Rivlin, Theodore J., *Chebyshev Polynomials, from approximation theory to algebra and number theory*, Pure and Applied Mathematics (New York), second edition, John Wiley & Sons Inc., New York, 1990

[73] Abramowitz, M. and Stegun, I. A., editors, *Handbook of Mathematical Functions*, Dover, New York, 1970

[74] Smolyak, S.A., "Quadrature and interpolation formulas for tensor productsof certain classes of functions," *Doklady Akademii Nauk SSSR*, volume 4, pp. 240–243, 1963

[75] Gerstner, T. and Griebel, M., "Numerical integration using sparse grids," *Numerical Algorithms*, volume 18, pp. 209–232, 1998

[76] Petras, K., "Fast calculation of coefficients in the Smolyak algorithm," *Numerical Algorithms*, volume 26, pp. 93–109, 2001

[77] Feng, Chao, "The Smolyak formulae package (www.personal.psu.edu/cml18/kinship),"

[78] Boyd, Stephen and Barratt, Craig, *Linear Controller Design: Limits of Performance*, Prentice-Hall, 1991

[79] Elzinga, Jack and Moore, Thomas G., "A central cutting plane algorithm for the convex programming problem," *Math Programming*, volume 8, pp. 134–145, ISSN 0025-5610, 1975

[80] Henrion, Didier, Lasserre, Jean B., and Löofberg, Johan, *GloptiPoly 3: moments, optimization and semidefinite programming*

[81] Boni, O., Ben-Tal, A., and Nemirovski, A., "Robust solutions to conic quadratic problems and their applications," *Optim Eng*, volume 9, no. 1, pp. 1–18, ISSN 1389-4420, 2008

[82] Fama, E. F. and French, K. R., "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, volume 33, no. 1, pp. 3–56, 1993

[83] Lasserre, J.-B. and Putinar, M., "Positivity and optimization for semi-algebraic functions," *ArXiv e prints*, October 2009

[84] Feng, C. and Lagoa, C. M., "Distributional robustness analysis for polynomial uncertainty," *Proc. 48th IEEE Conf. held jointly with the 2009 28th Chinese Control Conf Decision and Control CDC/CCC 2009*, 2009

[85] Timan, A. F., *Theory of approximation of functions of a real variable*, The Macmillan Co., New York, 1963

[86] Caliari, M., Vianello, M., De Marchi, S., and Montagna, R., "Hyper2d: a numerical code for hyperinterpolation on rectangles," *Appl Math Comput*, volume 183, no. 2, pp. 1138–1147, ISSN 0096-3003, 2006

[87] Dabbene, F., Feng, C., and Lagoa, C. M., "Robust and Chance-Constrained Optimization under Polynomial Uncertainty," *Proc. 48th IEEE Conference on Decision and Control*, 2009

[88] Nesterov, Yu., "Squared Functional Systems and Optimization Problems," *High Performance Optimization*, edited by H. Frenk, K. Roos, T. Terlaky, and S. Zhang, chapter 17, Kluwer Academic Publishers, The Netherlands, pp. 405–440, 2000

[89] Feng, C., Lagoa, C. M., and Sznaier, M., "Hybrid System Identification via Sparse Polynomial Optimization," *Proceedings of the American Control Conference*, 2010

[90] Henrion, D. and Lasserre, J. B., "Detecting global optimality and extracting solutions in GloptiPoly," Technical report, LAAS-CNRS, 2005

[91] Fazel, M., Hindi, H., and Boyd, S., "Log-Det Heuristic for Matrix Rank Minimization with Applications to Hankel and Euclidean Distance Matrices," *Proceedings of the American Conference of Control*, 2003

[92] Waki, H., Kim, S., Kojima, M., Muramatsu, M., and Sugimoto, H., "SparsePOP-A Sparse Semidefinite Programming Relaxation of Polynomial Optimization Problems," *ACM Trans Math Softw*, volume 35, no. 2, pp. 1–13, ISSN 0098-3500, 2008

[93] Scherer, C. W. and Weiland, S., *Linear Matrix Inequalities in Control*, Dutch Inst. Syst. Contr. Lecture course notes, 1997

[94] Feng, C., Lagoa, C., Ozay, N., and Sznaier, M., "Hybrid system identification: an SDP approach," *Proc. IEEE Conf. Dec. Contr.*, 2010

[95] Jury, E., *The theory and application of Z-transform method*, Wiley, New York, 1964

[96] Dumitrescu, B., *Positive Trigonometric Polynomials and Signal Processing Applications*, Springer, 2007

[97] Feng, C., Lagoa, C., and Sznaier, M., "Identifying Stable Fixed Order Systems from Time and Frequency Response Data," *Proc. American Control Conference*, 2011

[98] Clenshaw, C.W. and Curtis, A.R., "A method for numerical integration on an automatic computer," *Numerische Mathematik*, volume 2, pp. 197–205, 1960

[99] Patterson, T.N.L., "The optimum addition of points to quadrature formulae," *Mathematics of Computation*, volume 22, pp. 847–856, 1968

[100] Petras, K., "Smolyak cubature of given polynomial degree with few nodes for increasing dimension," *Numerische Mathematik*, volume 93, pp. 729–753, 2003

[101] Novak, E. and Ritter, K., "Simple cubature formulas wigh high polynomial exactness," *Constructive Approximation*, volume 15, pp. 499–522, 1999

[102] Dabbene, Fabrizio, "Solving uncertain convex optimization problems using sparse grids," Technical report, IEIIT-08c, 2008

[103] Rudin, Walter, *Real and Complex Analysis (3rd edition)*, McGraw-Hill Book Co., New York, 1987

[104] Barvinok, Alexander, *A Course in Convexity*, volume 54 of *Graduate Studies in Mathmatics*, American Mathmatical Society, Providence, RI, 2002

Vita

Chao Feng

Chao Feng was born in Hubei Province, China. He received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2003 and 2006, respectively. Then he joined the Ph.D. program in Electrical Engineering at the Pennsylvania State University under supervision of Prof. Constantino Lagoa. His research interests include robust control, distributional robustness analysis, hybrid system identification, and applications of control theory and optimization to practical problems in computer vision and finance. His publications (including those in preparation) during PhD study are listed in the following:

- (In preparation). A Convex Approach to Generalized Fixed Order Interpolation.
- (In preparation). Distributional Robustness Analysis for Nonlinear Uncertainty.
- (In preparation). Moments-based Algorithms for Identification of Switched Affine Systems.
- C. Feng, C. M. Lagoa and M. Sznaier (2012). A Convex Approach to Generalized Fixed Order Interpolation *American Control Conference, Montreal, Canada*
- C. Feng, F. Dabbene and C. Lagoa (2011). A Kinship Function Approach to Robust and Probabilistic Optimization under Polynomial Uncertainty, *IEEE Transactions on Automatic Control, Volume 56, issue 7, July, 2011.*
- C. Feng, C. Lagoa and M. Sznaier (2011). Identifying Stable Fixed Order Systems from Time and Frequency Response Data, *American Control Conference, San Francisco, CA, USA*
- C. Feng, C. Lagoa, N. Ozay and M. Sznaier (2010). Hybrid System Identification: An SDP Approach, *Conference of Decision and Control, Atlanta, GA, USA.*
- C. Feng, C. Lagoa and M. Sznaier (2010). Hybrid System Identification vis Sparse Polynomial Optimization, *American Control Conference, Baltimore, MD, USA.*
- C. Feng and C. Lagoa (2009). Distributional Robustness Analysis for Polynomial Uncertainty, *Conference of Decision and Control, Shanghai, China.*
- F. Dabbene, C. Feng and C. Lagoa (2009). Robust and Chance-Constrainted Optimization under Polynomial Uncertainty, *American Control Conference, St. Louis, MO, USA.*
- F. Dabbene, C. Feng and C. Lagoa (2007). Robust LMIs with Polynomial Dependence On the Uncertainty, *American Control Conference, New York City, NY, USA.*