Understanding Probabilistic Uncertainty Using ν-Gap

Anton Nyström

LUND
UNIVERSITY

Department of Automatic Control

MSc Thesis
TFRT-6230
ISSN 0280-5316

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

Printed in Sweden by Tryckeriet i E-huset
Lund 2024

Abstract

System uncertainty constitutes a fundamental restriction on control performance. System models are never perfect and the differences between system and model can be difficult to contend with. Applying a model based controller to the true system with uncertain dynamics can yield unpredictable results which led researchers to produce methods of robust control design. Existing theory on the ν-gap metric provides control performance guarantees given bounds on the metric. However, it does not utilize any further information than the bounds, essentially restricting the set of possible systems into an uncertainty set for which the guarantees apply.

This thesis aims to investigate how additional information about the uncertain system can be leveraged to provide sharper results; specifically, by additionally considering a probability distribution function (PDF) on the uncertainty set. Considering the uncertain system as a random quantity with a known distribution models it as more than simply belonging to some uncertainty set. It also incorporates further knowledge as to where in the set it is more likely to be. As such, this thesis opens up an entirely new perspective on the field of probabilistic uncertainty in control systems using the lens of the ν-gap.

Using the additional information, this thesis provides insight into how the difference between a known model and the uncertain system is characterized as well as the potential effect on control performance. Two expressions for the cumulative distribution function (CDF) of one such difference metric called the chordal distance is derived. With knowledge of this distribution, probabilistic guarantee results of a performance measure called the point-wise generalized stability margin are also produced. Some intermediate results which further illuminate the concepts and their relation to each other are also found. Lastly, a thorough discussion is given on how this field of research could be explored to expand the work started in this thesis.

Key Words: Control system uncertainty, Probabilistic uncertainty, Probabilistic control design, Nu-gap, Stereographic Projection.

Acknowledgements

This thesis was hosted by the Department for Automatic Control at Lund Faculty of Engineering as a master's thesis final project for my engineering degree. The purpose was to explore this new area of research to discern its potential and what interesting questions it could answer.

I would like to express my deepest gratitude to my thesis supervisor Venkatraman Renganathan at the Control Department. Gratitude, not only for his guidance and resolute support throughout this process, but also for laying the crucial groundwork in conceptualizing this research direction and coming up with the idea. Our discussions and regular meetings have been absolutely vital in maintaining a clear focus in the research.

Furthermore, I extend my appreciation to Professor Michael Cantoni at the University of Melbourne, Australia for joining me and Venkatraman in the pursuit of this research and providing keen insight into problems encountered. It has been an honor to discuss these topics with Michael and he has been an instrumental asset during this time.

I am also thankful for the help provided by Johan Lindström at the Department of Mathematical Statistics and my fellow student Oscar Gummesson Atroshi for the invaluable discussions had to better understand the thesis topics.

Lastly, I wish to express my gratitude to my examiner Anders Rantzer and Bo Bernhardsson at the Control Department for enabling this thesis to happen.

Contents

1

Research Motivation

One of the main goals of the field of automatic control is to study how to design robust controllers. A controller is robust if its performance is insensitive to differences between the model system used for the controller synthesis and the actual system subject to control. Several approaches for how to accomplish this goal have been proposed in earlier work including the small gain argument [Desoer and Vidyasagar, 2009], integral quadratic constraints [Megretski and Rantzer, 1997], and $\mathscr{H}_\infty$ control theory [Başar and Bernhard, 2008]. All of these design strategies can provide certain guarantees to the performance of a controller when applied to an uncertain system given that the uncertainty satisfies a set of criteria. As for when these criteria are not met, one must consider alternative methods of guaranteeing performance. In this chapter, this problem will be expanded upon and a practical anchoring to the problem setting shall be provided for the reader. Questions and problems that remain as gaps in the existing theoretical foundation will be discussed. Following this initial discussion, the next chapter will outline the necessary theory for producing the results of the subsequent chapter. Lastly a short discussion of their implications follows. The contents of this thesis has the capacity to open up an entirely new perspective on the field of research capable of extending the possibilities of robust control beyond the previously existing methods ([Desoer and Vidyasagar, 2009; Megretski and Rantzer, 1997; Başar and Bernhard, 2008]) with the referenced theoretical voids as motivation for its importance. In this chapter the essence of this motivation will be explored and lead the reader into an example which puts the problem into a very tangible context.

1.1 Motivating Problem

In this section a typical problem within automatic control related to system identification for control purposes will be laid out. The purpose is to illustrate what issues and questions arise throughout the controller synthesis process and how the theory presented in this thesis will be useful in answering many of these questions. The

example will be brought up and tied back to several times throughout this thesis to further illuminate the results and concepts presented as well as provide an intuitive anchoring of the theory to a concrete setting.

A very common setting that a control engineer will encounter is when the engineer has a system which he or she wishes to design a controller for with appropriate robustness and tracking capabilities. Several methods for linear control design are known including $\mathscr{H}_\infty$ Loop-shaping [McFarlane and Glover, 1992] which can provide robustness guarantees. Other common methods are LQG based design [Kalman, 1960] and Model Predictive Control (MPC) [Rawlings et al., 2017] which are often used to manage tracking capabilities. There are also several methods within stochastic robust control which incorporate stochastic uncertainty in the control design, such as Stochastic Tube MPC [Cannon et al., 2011] and Probabilistic Robust LQR [Rohr et al., 2021] with methods to validate models using probabilistic arguments [Halder, 2014]. Common for all mentioned approaches for controller design is that they are model based and thus require a known model of the system for controller synthesis. However, it is never possible to model the system with no error (for instance by restricting the model space to linear models) and herein lies several major problems. The essence of these problems can be captured nicely by the following questions:

1. If the model dynamics differs from the true system dynamics, can one guarantee that the controller will perform as intended on the true process when it is designed using the model as a substitute?
2. If it is possible to place such guarantees, how do they depend on the magnitude of the differences between the model and the true system?
3. How does one measure the magnitude of such differences when the true system is unknown?

Clearly these are fundamental problems at the base of the entire controller design process which might render the entire process useless if the model of the true system turns out to be inadequate. Without any such performance assurances, using a designed controller directly on the physical system could potentially yield catastrophic consequences. As such, answering these questions before implementation is of very high importance.

The pursuit of the answer to the last two questions has led researchers to come up with several distance metrics between dynamical systems. Specifically for LTI systems, the Gap metric [Zames and El-Sakkari, 1980; Georgiou, 1988; Georgiou and Smith, 1990], and Graph metric [Vidyasagar, 1984] were developed. Building on top of their works, G. Vinnicombe proposed the ν-Gap metric in [Vinnicombe,

1992; Vinnicombe, 2000] having sharper quantitative results than the Gap metric and opening up the door to answering the preceding two questions. Specifically for the SISO case, the ν-Gap metric offers a striking benefit of a direct frequency domain interpretation in terms of the chordal distance between the stereographic projection of the Nyquist plots onto the Riemann sphere which we will explore in Chapter 2.

Further, while analyzing the robustness of feedback systems in [Vinnicombe, 1992], Vinnicombe presents an interesting problem namely,

> how much do we need to know about a system in order to design a feedback compensator that leaves the closed loop insensitive to what we don't know?

The following robust stability result that answers this question from [Vinnicombe, 1993] using the ν-Gap metric is re-stated here without proof.

PROPOSITION 1
(From [Vinnicombe, 1993]) Given a nominal continuous time LTI system $\bar{P}$, and nominal feedback compensator $\bar{C}$, let

$$b_{\bar{P},\bar{C}} := \begin{cases} \left\| H(\bar{P},\bar{C}) \right\|_\infty^{-1}, & \text{if } H(\bar{P},\bar{C}) \text{ is stable} \\ 0, & \text{otherwise,} \end{cases} \tag{1.1a}$$

where

$$H(\bar{P},\bar{C}) := \begin{bmatrix} \bar{P} \\ I \end{bmatrix} (I - \bar{C}\bar{P})^{-1} \begin{bmatrix} -\bar{C} & I \end{bmatrix}, \tag{1.1b}$$

and $\|\cdot\|_\infty$ denotes the $\mathscr{H}_\infty$ norm. Then, any controller $\bar{C}$ that stabilises $\bar{P}$ and achieves $b_{\bar{P},\bar{C}} > \alpha$ stabilises the set of systems

$$\{P : \delta_\nu(P,\bar{P}) \leq \alpha\} \tag{1.2}$$

and

$$b_{P,\bar{C}} \geq b_{\bar{P},\bar{C}} - \delta_\nu(P,\bar{P}), \tag{1.3}$$

where $\delta_\nu(P,\bar{P})$ denotes the ν-gap between P and $\bar{P}$ and will be defined in Chapter 2.

Here $b_{\bar{P},\bar{C}}$ is a generalized stability margin where a larger value implies better stability margins. With this the three questions posed above have an answer for applicable

cases, and further still the fourth question posed by Vinnicombe can be partially answered using this. However, the answer is not exhaustive and some problems still remain. Specifically, calculating the ν-gap in (1.3) to be able to place guarantees on the performance of the designed controller requires knowledge of the true process, something we specifically stated is unrealistic. As such the only use we have of Proposition 1 is to specify a set of systems (1.2) which when controlled by $\bar{C}$ will satisfy a performance requirement specified by (1.3). The answer to the question posed by Vinnicombe about how much is needed to know to design a robust compensator would be that we need to know if the model belongs to the set (1.2) or not. This binary answer lacks nuance and fails to capture any potential prior knowledge by specifying where in the set (1.2) it might be by some measure.

One way of respecting the lack of precise knowledge of the true process while also capturing any prior knowledge the control engineer might have of the process is to consider the true system as a random quantity. The prior knowledge can then be expressed in terms of a probability distribution over the set of models. This opens the possibility to consider the results in Proposition 1 from a new perspective using more information about the true system and generating sharper results. The following section explores a practical problem where such prior knowledge is present and provides an example where the theory presented in this thesis would be useful.

The example problem

One of many ways of constructing a model for a system is to use system identification methods on measured data, a concept detailed well in [Ljung et al., 2021]. Input and output data is collected from measuring the true system. Then the data can be used to estimate parameters in order to generate a nominal model satisfying a specific predetermined model structure capturing information such as the model order and any delay present. The nominal model can then be validated against a separate data set. Since the data collected is often prone to random variations the dataset will constitute a random sample of the behavior of the studied system. Consequently, the estimated parameters will have some inherent uncertainty. Generally, a larger sample size of the system will lead to a decrease in variance of the estimations and subsequently a more accurate model, given an accurate parametrization of the true system and a bias free sampling. This data collection process can for certain processes be highly time intensive and costly as it necessitates measuring the physical process for a large number of iterations or a long time using different input. For active industrial processes, this entails downtime in production and could also cause wear and tear on the process, actuators and sensors involved. Thus, the less data needed to be collected, the better.

Suppose that we from prior experience know that a continuous time LTI model with no zeros and a triple pole models a process well. Therefore, we say with high

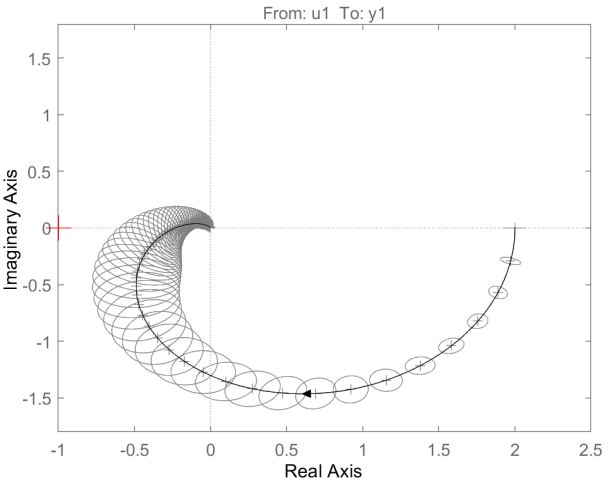


Figure 1.1 Nyquist plot of the identified system along with uncertainty regions corresponding to 5 standard deviations at every fifth frequency.

certainty that the process is modeled well by a transfer function of the form

$$P(s) = \frac{K}{(1+\tau s)^3}, \tag{1.4}$$

where $K, \tau \in \mathbb{R}$ are the parameters that are estimated from data collected as described. Since physical system outputs are subject to measurement noise, uncertainty is introduced into the identification process. In Figure 1.1 one such estimation of the example process Nyquist plot is shown where a random binary sequence was generated as input to a chosen nominal model $\bar{P}(s)$, obtained using $K = 2$, $\tau = \frac{1}{10}$ in (1.4). Gaussian white noise was added to the output representing measurement noise before estimating the parameters. In the figure we can also see point-wise 5 standard deviation confidence regions on a frequency grid of interval length $5\,\mathrm{rad/s}$.

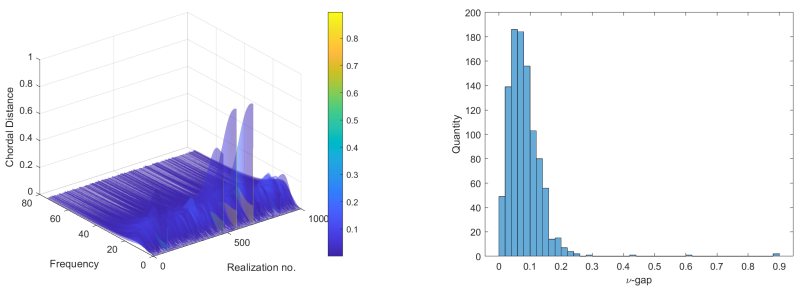
Here, we can see the results of the variability in the parameter estimation mentioned earlier. For every frequency, a region around the estimated point now corresponds to a region where the Nyquist plot of the true system at that frequency is very likely to be given that the model parametrization accurately reflects the true system (in this case well above 99%). Note that for this example, the gain uncertainty at the phase crossover frequency suggest that had the system been identified from a different data sample, a model with a very different gain margin could very well have been identified. This implies a model with different closed loop behavior and as such different controller performance.

What this example illustrates is precisely the implications of the questions posed earlier. Here, we do not know what the true system is but are fairly certain it belongs to some set of models in proximity to an estimated nominal model with respect to their Nyquist plots. Furthermore, we have some estimate of the probability distribution over that set of models since the set itself was determined using the point-wise distribution of the model estimation and a threshold to close the set, in this case 5 standard deviations. It is now interesting to consider the following:

> If a controller is designed using the identified nominal model, is there any way of taking the estimated probability distribution into account to leave the control of the true system invariant to the model estimation uncertainty?

In essence this would involve answering the following question which is the main goal of this research direction.

QUESTION 1
Given a nominal model and a probability distribution with support on some set modeling the true system, what is the probability that a controller designed using the nominal model will be satisfactory by some performance threshold on a randomly chosen system in the set?



(a) Realizations of chordal distance to the nominal model across frequency and 1000 trials.

(b) The histogram of the ν-gap between the nominal model and 1000 independent trials.

Figure 1.2 Simulation results of the example problem showing how randomly sampled true systems yield realizations of the chordal distance over frequency to the left as well as realizations of the ν-gap to the right, calculated as the maximum chordal distance for each realization.

The following simulation illustrates how this question could be answered. We simulated the system identification procedure described earlier for $N = 1000$ indepen-

dent trials (with different randomly generated inputs and measurement noise) to identify N systems of the form (1.4). The results shown in Figures 1.2a and 1.2b indicate that the point-wise chordal distance can be formulated as a random quantity with a normalized histogram being an estimate of the distribution of the ν-gap for this case. With a known distribution of the ν-gap an examination of (1.3) could yield performance bounds in a probabilistic sense.

With an answer to Question 1 the realm of queries opens up and future work could focus on answering any of the following questions:

1. Given a tolerable uncertainty set size and requirements on the distribution restricting a large likelihood to a certain subset, how much data needs to be collected to achieve that level of certainty?

2. If the distribution and support set are derived from data collection as with the example problem, how does different identification methods map to different distributions and sets?

1.2 Summary and Disposition

In this chapter, a context to this thesis has been provided, and the missing pieces of existing probabilistic control theory have been identified. A thorough motivation and initial problem exploration have been provided along with a practical example which illustrates the problem well. Clearly, there are many unanswered questions and in the pursuit of answering some of them, we need to develop a sturdy theoretical foundation. In the following chapter, existing theory and background will be outlined to provide this foundation for the derivations that will follow in an attempt to answer some important questions described previously. The following chapter will also include a description of the limitations of this thesis. It turns out that answering Question 1 is rather difficult in terms of analytically determining the probability distribution of the ν-gap. This thesis will therefore consider the corresponding point-wise frequency measure as a precursor to future work. This involves specifying both the mentioned model distribution and support set for every frequency and analyzing the implications on control performance for a fixed frequency.

Following the chapter discussing background and limitations, a step back will be taken again to discuss the problem setting more clearly using the information in the background providing further understanding of its relevance and intricacies. With the problem formulation and its context to the theoretical background crystallized, the subsequent chapter will outline all theoretical and numerical results along with their derivations. Lastly, a chapter regarding conclusions and discussion will delineate the implications of the results and the doors they open up in terms of future

research. All plots can be reproduced using the code publicly available on [Nyström, 2024].

It should be noted that a subset of the results of this thesis has been submitted to the 2024 CDC conference in the paper *Stereographic Projection of Probabilistic Frequency-Domain Uncertainty* by myself Anton Nyström, Venkatraman Renganathan and Michael Cantoni [Nyström et al., 2024].

2

Background

In this chapter, all the theoretical foundation relevant for the results produced in this thesis will be laid out and explained. Throughout the chapter each concept and its corresponding notation are introduced.

2.1 Notations and Basics

In order to maintain a suitable scope for the included background, the reader is expected to understand some basics in control theory including Nyquist plots and the Nyquist stability criterion. Furthermore, the reader is expected to have a basic understanding about control design methods and concepts regarding stability, robustness and control performance. While perhaps not strictly necessary, such knowledge will be helpful in fully understanding the content of this thesis along with the implications of the results. Nevertheless, this section will briefly review some of the basic concepts without extensive detail and introduce some notations which will be used throughout this thesis.

Notations

The set of real numbers, integers and the natural numbers are denoted by $\mathbb{R}, \mathbb{Z}$, and $\mathbb{N}$ respectively. The space of complex numbers is denoted by $\mathbb{C}$ and j represents the imaginary unit. Later we will see that we also use this notation for the extended complex plane $\mathbb{C} \cup \{\infty\}$. The real and imaginary parts of the complex number $z \in \mathbb{C}$ is denoted by $\Re(z)$ and $\Im(z)$ respectively. Given $z \in \mathbb{C}$, we denote its complex conjugate as $z^\star \in \mathbb{C}$ and its modulus by $|z|$. For convenience, we will in certain instances write parameterized multi-variable functions as follows $f(x(t),\, y(t)) = f(x,y)(t)$. Throughout this thesis, variations of P will be used to denote systems and their transfer functions, with either an implicit dependence on $s \in \mathbb{C}$ for transfer functions or $j\omega,\ \omega \in \mathbb{R}$ for frequency responses. The same variations of P will also sometimes be used to denote complex numbers. In such cases the reader is

encouraged to think of $P \in \mathbb{C}$ as instances of the transfer function P for a fixed argument s or ω.

The probability space is defined using a triplet $(\Omega, \mathscr{F}, \mathbf{P})$, where $\Omega, \mathscr{F}$, and $\mathbf{P}$ denote the sample space, event space and the probability function respectively with $\mathbf{P} : \Omega \to [0,1]$. A real random vector $\mathsf{x} \in \mathbb{R}^n$ following a probability density function f_{x} is denoted by $\mathsf{x} \sim f_{\mathsf{x}}$ and its CDF is denoted by F_{x}. Let $x \in \mathbb{R}^n$ be a realization of x. A complex random variable p on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ can be considered as a random vector in $\mathbb{R}^2$ with real and and imaginary parts in each dimension.

Control Fundamentals

This subsection deals with the fundamental control concepts including spaces of transfer functions, notions of stability and robustness as well as theory and known results around these concepts.

Spaces and Norms. In order to properly discuss stability and robustness a thorough restriction of the spaces of the systems and their input and output is required. In this section, the relevant spaces will be defined and explained.

Let $\mathbf{R}(s)$ denote the set of proper rational functions in $s \in \mathbb{C}$ with real coefficients. The Hardy space consisting of transfer functions of stable LTI continuous time systems is denoted by $\mathscr{H}_\infty$ and is equipped with the norm

$$\|P\|_{\mathscr{H}_\infty} = \sup_{\mathrm{Re}(s)\geq 0} |P(s)|. \tag{2.1}$$

Also let $\mathbf{R}\mathscr{H}_\infty := \mathbf{R}(s) \cap \mathscr{H}_\infty$ to denote the set of proper rational functions whose poles are in the open left half-plane.

For every $p(s) \in \mathbf{R}\mathscr{H}_\infty$, we define its norm as

$$\|p\| := \sup_{\omega} |p(j\omega)| = \sup_{\mathrm{Re}(s)\geq 0} |p(s)|. \tag{2.2}$$

Then, a distance between two functions $p(s), q(s) \in \mathbf{P}(s)$ can be given by $\|p-q\|$ which is induced by the norm defined in (2.2).

The Nyqust plot and Nyquist Stability Criterion. In this section, some fundamental control theory which will be useful to understand in order to fully grasp the contents of this thesis will be briefly outlined. This includes well known and widely used tools and results such as the Nyquist plot, where details can be found in [Hägglund, 2021], and the Nyquist Stability Criterion outlined in [Åström and Murray, 2021; Pates, 2021]. As mentioned in the chapter introduction, the reader is expected to have an earlier understanding of these topics and hence those concepts shall be restated here with little details.

One instance of a Nyquist plot has been shown already in Figure 1.1 and constitutes a useful tool for analysis of LTI systems. It consists of a curve in the complex plane of the frequency response of a system parameterized by a complex frequency. As such, for a SISO system P, its frequency response $P(j\omega)$ is plotted in the complex plane for all $\omega \in \mathbb{R}$. The plot shown in Figure 1.1 only displays one half of the plot corresponding to positive ω. However, for real rational frequency responses, the negative half will always be the conjugate of the positive, resulting in a plot mirrored with respect to the real axis. The Nyquist plot is therefore equivalent to another common tool for visualization in the Bode magnitude and phase plots since every point on the curve has a modulus and an argument equal to the magnitude and phase at that frequency. Thus, using the Nyquist plot, one can investigate a multitude of important quantities including gain and phase margin, bandwidth and crossover frequency to get closed-loop inferences using only open-loop characteristics.

One central use of the Nyquist plot uses the Cauchy Argument Principle to draw conclusions about the system behaviour in closed loop, a result called the Nyquist Stability Criterion which can be found in [Åström and Murray, 2021; Pates, 2021]. If we let P_o be the open loop frequency response and P_{cl} be the frequency response of the closed loop system under negative unit feedback, then the Nyquist Stability Criterion states that

$$wno(1+P_o) = \eta(P_o) - \eta(P_{cl}) \tag{2.3}$$

where *wno* corresponds to the winding number evaluated along the standard Nyquist contour and $\eta(\cdot)$ denotes the number of unstable poles. As such the criterion states that the difference in the number of open and closed loop unstable poles is equal to the number of anti-clockwise encirclements of the point -1 by the open loop Nyquist plot. We shall see that this condition will be an integral part in how the ν-gap is defined but will not be explored fully in this thesis.

Probability Theory

Since this thesis deals with probabilistic uncertainty, we will require some results from probability theory. Useful notions and known results from this category will be stated here and used throughout the remainder of the thesis in derivations of the main results.

The main result we will find use for regards how mappings between random variables affect distributions. More specifically we will use the following known result. Let x,y be random variables belonging to the joint distribution $f_{\mathsf{xy}}(x,y)$, and $g_1(\mathsf{x},\mathsf{y}), g_2(\mathsf{x},\mathsf{y})$ be continuous differentiable functions. Then define two new random variables,

$$\begin{cases} \mathsf{z} := g_1(\mathsf{x},\mathsf{y}), \\ \mathsf{w} := g_2(\mathsf{x},\mathsf{y}). \end{cases} \tag{2.4}$$

Then, the joint density $f_{\mathsf{zw}}(z,w)$ can be expressed in terms of f_{xy}. That is,

$$f_{\mathsf{zw}}(z,w) = \sum_i \frac{1}{|J(x_i(z,w),y_i(z,w))|} f_{\mathsf{xy}}(x(z,w),y(z,w)) \tag{2.5}$$

where J denotes the Jacobian determinant of the transformation defined by (2.4) and x_i, y_i denote solutions to (2.4) for fixed z,w (see chapter 6 in [Papoulis and Pillai, 2002]).

The following is another useful result which will be used in this thesis. Define the random variable,

$$\mathsf{Q} = \frac{\mathsf{z}}{\mathsf{w}}$$

Then for non-negative random variables z,w with joint density $f_{\mathsf{zw}}(z,w)$, the following holds true for the cumulative distribution of Q (see Chapter 5 in [Papoulis and Pillai, 2002]):

$$F_{\mathsf{Q}}(q) := \int_0^\infty \int_0^{wq} f_{\mathsf{zw}}(z,w)\, dz\, dw, \quad q \geq 0. \tag{2.6}$$

2.2 Preliminaries

One of the stated questions in the previous chapter regarded how one measures the magnitude of a difference between two systems. In this section, some useful metrics on the system spaces defined above will be explored along with related results on how they can relate to robustness criteria set by the control engineer.

The Stereographic Projection and Chordal Distance

Before properly defining the metric most relevant in this thesis, the definition of the stereographic projection is necessary. First, consider the Riemannian sphere $\mathfrak{R} \subset \mathbb{R}^3$ as a sphere with unit diameter centered in $z = \frac{1}{2}$ and its south pole tangent to the x-y plane. As such we can characterize the sphere as follows

$$\mathfrak{R} = \left\{ (x,y,z) \in \mathbb{R}^3 \ : \ x^2 + y^2 + \left(z - \frac{1}{2}\right)^2 = \frac{1}{4} \right\}. \tag{2.7}$$

Note that this is a surface in $\mathbb{R}^3$ and as such a 2D structure. Therefore we can also express the set in terms of spherical coordinates using the point $(x,y,z) = \left(0,0,\frac{1}{2}\right)$ as its origin. The equivalent coordinates would then be $\left(\frac{1}{2},\theta,\varphi\right)$ where θ and φ are the polar and azimuthal angles respectively. Thus, we have the alternative representation

$$\mathfrak{R} = \left\{ (r,\theta,\varphi) \in \mathbb{R}^3 \ : \ r = 1/2,\ \theta \in [0,\pi],\ \varphi \in [0,2\pi] \right\} \tag{2.8}$$

which will be leveraged later in this thesis.

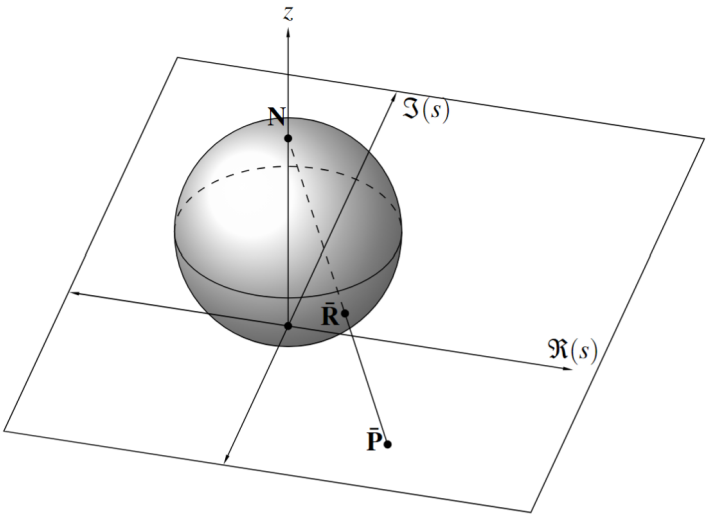


Figure 2.1 Illustration of the stereographic projection mapping (2.9).

Before defining the stereographic projection, a short discussion of the extended complex plane $\mathbb{C} \cup \{\infty\}$ is also required. As we shall see the range of the projection is the entire complex plane including the point at infinity. However, the usage of the complex plane in this thesis is always with respect to a system frequency response function. Furthermore, as we will see later, we have restricted our study to proper systems in $\mathscr{RH}_\infty$. Therefore, we will always have systems of finite open loop gain and thus never approach infinity. As such we will for the purpose of convenience and brevity view the complex plane and extended complex plane as interchangeable and denote them both with $\mathbb{C}$.

We are now ready to define the stereographic projection as a mapping from the Riemannian sphere onto the extended complex plane $\mathbb{C}$. The mapping originates from letting the complex plane coincide with the x-y plane, making $\mathfrak{R}$ rest its south pole on the origin of the complex plane. Then any point $\bar{R} \in \mathfrak{R}$ maps onto the point $\bar{P} \in \mathbb{C}$ at which a straight line passing through $\bar{R}$ and the north pole intersects the complex plane. The point at the north pole maps to infinity [Vinnicombe, 2000]. An illustration of the mapping can be seen in Figure 2.1. Formally, we have the following definition where $\bar{R} = (x, y, z) \in \mathfrak{R}$ in accordance with (2.7). We remind the reader of the notation convention that $\bar{R}$ does not constitute any complex conjugate, but simply a known point.

$$\phi : \mathfrak{R} \to \mathbb{C}, \tag{2.9}$$

$$\phi(x,y,z) = \frac{x}{1-z} + \frac{y}{1-z} j \tag{2.10}$$

Conversely, we define its inverse $\phi^{-1} : \mathbb{C} \to \mathfrak{R}$ as

$$\phi^{-1}(P) = \left(\frac{\Re(P)}{1+|P|^2}, \frac{\Im(P)}{1+|P|^2}, \frac{|P|^2}{1+|P|^2} \right) \tag{2.11}$$

which we will see more use of later. We shall stay consistent with the notation used previously where $\bar{P} = \phi(\bar{R})$.

With these prerequisites, we are now ready to define the last quantity before arriving at the metrics. We define the point-wise chordal distance $\kappa : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ between two points $P, \bar{P} \in \mathbb{C}$ as follows:

$$\kappa(\bar{P}, P) = \frac{|\bar{P} - P|}{\sqrt{1+|P|^2}\sqrt{1+|\bar{P}|^2}}. \tag{2.12}$$

This definition coincides with the Euclidean distance between $\phi^{-1}(P)$ and $\phi^{-1}(\bar{P})$ making the naming appropriate. A reminder is in order that this definition of the chordal distance regards points in $\mathbb{C}$ but should be thought of as instances of a frequency response of a system at a certain frequency. As such, we can see the chordal distance as a function of frequency since both its arguments can be interpreted as functions of frequency. More rigorously for two SISO systems with transfer functions $P(s)$ and $\bar{P}(s)$ we find the point-wise chordal distance for each frequency $s = j\omega$ according to (2.12) where $P = P(\omega), \bar{P} = \bar{P}(\omega)$ [Vinnicombe, 2000].

For reasons that will become apparent, the SISO system specific definition (2.12) is sufficient for our purposes, but it should be noted that in the referenced literature [Vinnicombe, 2000] the quantity is defined with respect to MIMO systems. However, it is shown that (2.12) is the result of that definition applied to SISO systems which is why it is defined in this way here. When conceptually discussing the results later in this thesis, the existence of κ with respect to MIMO systems will be leveraged without stating precisely how it is defined.

Metrics on Spaces of LTI Dynamical Systems

In this section, the ν-gap metric will be defined and discussed. Before arriving at the main topic, some initial context with respect to other metrics will be outlined. The ν-gap builds on prior work and work done on other metrics. One branch of

research focuses on the development of the Graph-metric [Vidyasagar, 1984] and theory related to it. However, the main prior work leading up to the ν-gap was the introduction of the Gap-metric [Zames and El-Sakkari, 1980; Georgiou and Smith, 1990; Georgiou, 1988] and surrounding results. The two metrics have led to their own respective sub branches of research providing a more nuanced picture of other options of metrics and results derived from them. Exactly what these results are will not be detailed in this thesis and we will instead delve deeper into the main metric of interest, the ν-gap[Vinnicombe, 2000; Vinnicombe, 1993; Vinnicombe, 1992]. As will become apparent when discussing the limitations of this paper, the ν-gap will not be used in the production of theoretical results, where we will instead study its point-wise in frequency counterpart. Despite this, the definition is still important to discuss to provide better context to the importance of the results produced and to the simulation results shown in Figure 1.2.

The ν-gap, denoted by δ_ν is another metric on the space of systems. Again, the general definition holds for MIMO systems, but with restriction to SISO systems, we will only consider the following definition of the metric. Consider two systems with transfer functions $P(s), \bar{P}(s)$. As per the previous section, we can consider the point-wise chordal distance as a function of frequency,

$$\kappa(P(j\omega), \bar{P}(j\omega)).$$

The ν-gap is then defined by the following,

$$\delta_\nu(P, \bar{P}) = \begin{cases} \sup_\omega \kappa(P(j\omega), \bar{P}(j\omega)), & \text{if } wno(1 + \bar{P}^\star P) + \eta(P) + \eta(\bar{P}) \neq 0 \\ 1, & \text{otherwise} \end{cases} \tag{2.13}$$

where *wno* and $\eta(\cdot)$ again correspond to the winding number and number of unstable poles respectively. A small remark is in order here as the winding number condition in (2.13) formally is also accompanied by the condition that $det(I + \bar{P}^* P) \neq 0$, $\forall \omega \in \mathbb{R}$ in the MIMO case. However, for the SISO case this simply entails that the Nyquist plot of $P^* \bar{P}$ is not identically equal to the point -1 on the real axis for all frequencies. This edge case is wholly uninteresting which is why we have omitted this condition from (2.13) entirely.

Generalized Stability Margin $b_{P,C}$ and ν-gap Duality

The usefulness of the ν-gap comes from its duality with a generalized stability margin measure $b_{P,C}$. Once again, this measure will not be directly used in this thesis due to the point-wise restriction to a frequency, however it is still useful to consider and there is similarly a point-wise notion which will be more thoroughly involved in the thesis.

For any SISO system P and compensator C, using unit feedback we obtain the pair $[P, C]$. It is well known that for a standard linear feedback system interconnection

with disturbances and measurement noise the Gang of Four matrices [Åström and Murray, 2021] are the four interesting closed-loop transfer functions to study. Together, they describe the system dynamics from all inputs including reference value, control signal, noise, and disturbance to the output. We can obtain the four transfer functions as the elements of the matrix

$$H(P,C) := \frac{1}{1-PC}\begin{bmatrix} P \\ 1 \end{bmatrix}\begin{bmatrix} -C & 1 \end{bmatrix} = \begin{bmatrix} \frac{-PC}{1-PC} & \frac{P}{1-PC} \\ \frac{-C}{1-PC} & \frac{1}{1-PC} \end{bmatrix}. \tag{2.14}$$

For a system to be robust, we need all of these to be bounded since if one of them is unbounded there will exist input signals that cause the system to go unstable. The concept of the generalized stability margin $b_{P,C}$, which can be found in [Vinnicombe, 2000], follows this logic and is defined as follows:

$$b_{P,C} = \begin{cases} \|H(P,C)\|_\infty^{-1}, & \text{if } [P,C] \text{is stable.} \\ 0, & \text{otherwise} \end{cases}. \tag{2.15}$$

Thus a small value for $b_{P,C}$ implies a large norm and as such at least one of the gang of four being very large which is commensurate with our notion of poor robustness. The converse is also true.

Some important known properties of this measure is that $b_{P,C} \in [0,\ 1]$ and that lower bounds on it can be translated into upper bounds on the sensitivity and complementary sensitivity functions. Thereby, bounds on the measure $b_{P,C}$ provides a connection to the traditional loop shaping procedure of robust control design referenced earlier in this thesis. More specifically, if P already has a desirable loop shape, then $b_{P,C}$ provides a bound on the degradation of that loop shape by C at high and low gains. If P does not have a desirable loop shape, there are weighting techniques that can be used to shape the initial loop shape before selecting a compensator C which can be found in [Vinnicombe, 2000]. In turn, a desirable loop shape is associated with desirable performance, which is measured using $b_{P,C}$. Exactly what these bounds are will not be stated here, but the knowledge of their existence helps in understanding the implications of the results found in this thesis and how further study in this field would be useful.

The existence of a useful duality between the ν-gap and $b_{P,C}$ was already stated in Proposition 1 and Equation (1.3). It is restated here with the preceding context providing clearer understanding. For two systems P and $\bar{P}$ and a compensator $\bar{C}$ which stabilizes $\bar{P}$ the following two inequalities always hold:

$$b_{P,\bar{C}} \geq b_{\bar{P},\bar{C}} - \delta_\nu(P,\bar{P}), \tag{2.16}$$

The inequality also implies the following two statements as stated in [Vinnicombe, 2000].

1. Given a plant and compensator $\bar{P}, \bar{C} \in \mathbf{R}$, and a number β, then: $[P, \bar{C}]$ is stable for all plants P satisfying $\delta_\nu(P, \bar{P}) \leq \beta$ if and only if $b_{\bar{P},\bar{C}} > \beta$.

2. Given two plants $\bar{P}, P \in \mathbf{R}$, and a number $\beta < \sup_{\bar{C}} b_{\bar{P},\bar{C}}$, then: $[P, \bar{C}]$ is stable for all compensators $\bar{C}$ satisfying $b_{\bar{P},\bar{C}} > \beta$ if and only if $\delta_\nu(P, \bar{P}) \leq \beta$.

The two statements provide guarantees with respect to how large an uncertainty set can be given a performance requirement as well as what performance can be guaranteed given a size of the uncertainty set. However, as stated these bounds are not sharp and lack incorporation of further knowledge of the involved systems.

As mentioned above, due to the restrictions on this thesis, we are interested in the corresponding point-wise version of $b_{P,C}$. For two points $z_1, z_2 \in \mathbb{C}$ we define

$$\rho(z_1, z_2) = \bar{\sigma}^{-1}\left(\frac{1}{1 - z_1 z_2}\begin{bmatrix} z_1 \\ 1 \end{bmatrix}\begin{bmatrix} -z_2 & 1 \end{bmatrix}\right) = \frac{1 - z_1 z_2}{\bar{\sigma}\left(\begin{bmatrix} -z_1 z_2 & z_1 \\ -z_2 & 1 \end{bmatrix}\right)} \tag{2.17}$$

as the point-wise generalised stability margin [Vinnicombe, 2000]. Here $\bar{\sigma}(\cdot)$ corresponds to the largest singular value. The intended application of this definition is for when z_1 and z_2 correspond to the frequency responses $P(j\omega)$, $C(j\omega)$ at a certain fixed frequency. With this perspective it holds that if $[P, C]$ is stable, we find

$$b_{P,C} = \inf_{\omega} \rho(P(j\omega), C(j\omega)).$$

We also know that the point-wise inequality corresponding to (2.16) also holds for ρ and κ [Vinnicombe, 2000]. With the notation $\rho_{P,C} := \rho(P(j\omega), C(j\omega))$ we have

$$\rho_{P,\bar{C}} \geq \rho_{\bar{P},\bar{C}} - \kappa(P(j\omega), \bar{P}(j\omega)). \tag{2.18}$$

Once again, both b and ρ are originally defined with respect to MIMO systems in the referenced material [Vinnicombe, 2000] whereas the definitions given here are equivalent when considering SISO systems, the reason for which will be apparent in the following section.

2.3 Limitations

The limitations of this thesis have been mentioned previously. For the sake of clarity we collect all restrictions made in this thesis here.

As stated in the end of the previous chapter, this thesis will be limited to the study of probability distributions of the point-wise distance between systems since it turns out that an analytic derivation of the distribution of ν-gap is difficult without a multitude of highly restrictive assumptions and simplifications. Furthermore, this thesis will be limited to study SISO LTI systems, hence the insistence on SISO definitions above. More specifically, we will restrict our study to the set of SISO systems with transfer functions in $\mathbf{R}\mathscr{H}_\infty$, meaning stable and proper rational transfer functions. However, it is not without consideration of how to conceptually generalize the results to the MIMO case and unstable systems. Some comments will be made on this in later parts of the thesis.

2.4 Summary

Throughout this chapter we have seen some necessary basic concepts and notations introduced along with more advanced preliminaries which are necessary to understand before embarking on producing the results of this thesis. Furthermore, a clearly defined limitation boundary has been drawn to crystallize what will and will not be covered by this thesis. In the following chapter, the problem context will be brought back into focus and discussed using the theoretical foundation and limitations laid in this chapter. This will lead to a more cohesive problem formulation which in turn guides us into the results with more clarity.

3

Further Motivation and Problem Formulation

Before embarking on the results of this thesis, we shall use our newfound understanding of the theory presented in the previous chapter to gain a better understanding of the problem setting. In this section, another simple second motivating example which showcases the idea of the problem in its most basic format is presented and is then followed up by a description of the full general problem setting.

3.1 A Simplified Illustrative Example

Let us consider the simple scalar real case ($P \in \mathbb{R}$). Now consider a circle $\mathscr{C}$ with unit diameter tangent with respect to the real number line at its south pole ($\mathscr{C}$ is the analogue of Riemann sphere in the complex setting defined earlier in the thesis). The line that connects the north pole of the circle, $\mathscr{N}$, and the points on $\mathbb{R}$ intersects the circle precisely at one point and constitutes the Stereographic Projection $\phi : \mathscr{C} \to \mathbb{R}$ as discussed. Now consider the case when the point $P \in \mathbb{R}$ is random and let $\mathbf{S} \subset \mathbb{R}$ denote the compact support set of the distribution f_{P} governing the uncertainty in P. Further, assume that a nominal value denoted by $\bar{P} \in \mathbf{S} \subset \mathbb{R}$ is known apriori. Note that, $P \sim f_{\mathsf{P}}$ can take any value in $\mathbf{S}$. Then, the distance between the projected values of nominal value $\bar{P}$ and any realization P on the circle $\mathscr{C}$ becomes random as well. A simple illustration is shown in Figure 3.1. Note that the angle $\alpha := \measuredangle P\mathscr{N}\bar{P}$ is a random quantity determined as a function of P and $\bar{P}$ and hence the chordal distance between $\bar{R} := \phi^{-1}(\bar{P})$ and $R := \phi^{-1}(P)$ given by $|\sin(\alpha)|$ turns out to be random as well. In such a setting, we would want to study the distribution of the perturbation from $\bar{P}$ needed to realize any $P \in \mathbf{S}$.

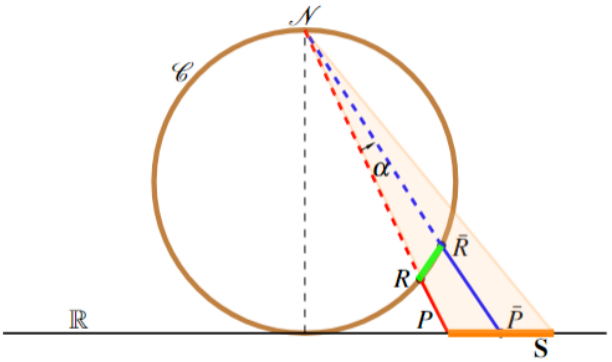


Figure 3.1 Stereographic projection on $\mathbb{R}$ is illustrated here. The uncertainty in point P is depicted as an orange interval $\mathbf{S} \subset \mathbb{R}$. Both the nominal point $\bar{P}$ and its projection $\bar{R}$ are shown in blue color. A realization of the random point $P \in \mathbf{S}$ and its inverse projection R are shown in red color. Since P is random, the angle $\alpha := \measuredangle P\mathcal{N}\bar{P}$ and hence the corresponding chordal distance line in green color are random as well.

3.2 The General Problem Setting

In this section, the setting from Figure 3.1 will be generalized to be commensurate with Figure 2.1 and will illuminate the conceptual framework which will follow throughout the rest of the thesis.

Consider the case where we have a known system $\bar{P}$ and one random system P such that for every frequency ω, $\bar{P}(j\omega) \in \mathbb{C}$ is known and $P(j\omega) \in \mathbb{C}$ is a random variable with known distribution $f_P(j\omega)$ on a compact support set $\mathbf{S}(j\omega)$ which governs the uncertainty of P. Note that, $P(j\omega) \sim f_P(j\omega)$ can take any value in $\mathbf{S}(j\omega)$. Then, similarly to the previous section, the distance between the projected values of nominal value $\bar{P}(j\omega)$ and any realization $P(j\omega)$ on the Riemann sphere $\mathfrak{R}$ becomes random as well. An illustration is shown in Figure 3.2, where the concept illustrated in Figure 3.1 is incorporated into Figure 2.1. In the figure, the dependence on $j\omega$ is omitted for brevity and both the distribution of $P(j\omega)$ over $\mathbf{S}(j\omega)$ and of $R := \phi^{-1}(P)$ on $\mathfrak{R}_S(j\omega) := \phi^{-1}(\mathbf{S}(j\omega))$ is exemplified using a red and green color gradient respectively. A blue Nyquist curve for the nominal model $\bar{P}$ is also included to further indicate that the point $\bar{P}(j\omega)$ is a point on a known Nyquist curve. Also note the close connection to Figure 1.1 where a single one of the ellipses along the curve has now been isolated and projected onto the Riemann sphere using the inverse stereographic projection. Once again we are interested in studying the distribution of the perturbation from $\bar{P}(j\omega)$ to realize any $P(j\omega) \in \mathbf{S}(j\omega)$ in terms of the chordal distance κ. For ease of notation the dependence on $j\omega$ will be omitted in the following section. However do note that the frequency dependence remains and will be revisited in discussion in the subsequent chapter.

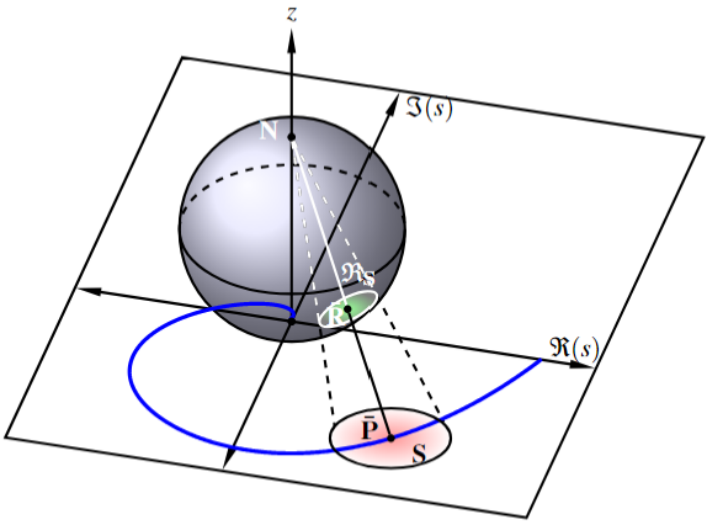


Figure 3.2 A conceptual understanding of how the problem of point-wise random system perturbation is illustrated here. The blue curve constitutes the Nyquist curve of the known nominal system $\bar{P}$ and the set $\mathbf{S}$ is the compact support set of distribution f_P which is exemplified by the red color gradient. The Riemann sphere with the corresponding inversely projected point $\bar{R}$ and support set $\mathfrak{R}_S$ with distribution f_R exemplified by the green color gradient is also shown.

3.3 Summary

With an initial problem description in the research motivation chapter, and theoretical development and a clear limitation in the background chapter we were here able to connect everything using two more examples into one cohesive understanding of the problem. We are now fully ready to present the results of this thesis in the following chapter answering Question 1 and finding important results along the way. The results that are relevant within the set limitations will be presented along with derivations and proofs. Additionally, two numerical examples will also be provided to illustrate the findings. In the subsequent chapter, a deeper discussion of the implications of the results will be presented along with a discussion on several possible future works and how this new take on the field of probabilistic robust control could yield fruitful results.

4

Main Results

This chapter contains the main results of this thesis along with ties back to the motivating problem from Section 1.1 in the form of a numerical example commensurate with the original problem setting. An initial result characterizing the mapping between the distribution f_P and f_R described in the previous section is provided. Subsequently, the main result is given from two perspectives providing two alternate ways of characterizing the distribution of the point-wise chordal distance between the known and random system. Additionally, a special case where an analytic solution can be obtained is presented and analyzed. Lastly, a continuation result utilizing the main results to provide insight into the potential control performance degradation is also presented.

4.1 Characterizing the Projected Distribution

From figure 3.2, we brought up the notion of considering the distribution of $R \in \mathfrak{R}$, f_R, with support set $\mathfrak{R}_S$ given a known distribution of P, f_P, with support $\mathbf{S}$. In this section, we will provide a characterization of this transformation of distribution for completeness of understanding. The later derivations are done completely with respect to f_P but could equivalently be considered with respect to f_R illustrated by the green color gradient in Figure 3.2 and as such an understanding of how to translate between the two distributions is important.

We shall see that given a fixed frequency ω, a compact set $\mathbf{S} \subset \mathbb{C}$, a deterministic nominal model $\bar{P} \in \mathbf{S}$ and a random variable $P \in \mathbf{S}$ such that $\begin{pmatrix} \Re(P) \\ \Im(P) \end{pmatrix} \sim f_P$ for some two dimensional probability distribution f_P on $\mathbf{S}$, we will be able to uniquely characterize the distribution f_R of $R := \phi^{-1}(P)$.

A short elaboration on the insistence of compactness is also in order as compactness is not needed for the results below to hold true. However, compactness would help

to simplify future work as a closed region around a nominal Nyquist plot means the Nyquist plot of $1+P\bar{P}$ in the winding number condition of (2.13) will be bounded for any P in the compact set and thus facilitates the investigation of the now probabilistic winding number condition. The assumption is also reasonable since we have restricted the study to the set of systems in $\mathscr{RH}_\infty$ which are always bounded and is further commensurate with the discussion on the interchangeability of the complex and extended complex planes. The above would also hold true simply for bounded sets, however another opening provided by compactness is the study of worst-case scenarios by considering the models that are on the boundary of the uncertainty region. Further discussion on these lines of future work will be discussed in the next chapter.

Let $P = \mathsf{x} + j\mathsf{y}$ and parameterize f_P as $f_{\mathsf{xy}}(x,y)$ thereby representing the uncertainty in $P \in \mathbb{C}$ as uncertainty along the real and imaginary axes. We will also leverage the spherical representation of $\mathfrak{R}$ from (2.8) to parameterize f_R as $f_{\mathsf{zw}}(z,w)$. Using this along with the stereographic projection mapping we consider the mapping

$$\begin{cases} \mathsf{z} := f_1(\mathsf{x},\mathsf{y}), \\ \mathsf{w} := f_2(\mathsf{x},\mathsf{y}) \end{cases}$$

from the real random variables $\mathsf{x},\mathsf{y} \in \mathbb{R}$ to the real random variables $\mathsf{z} \in [0,\pi], \mathsf{w} \in (-\pi,\pi]$ corresponding to random polar and azimuthal angles θ and φ. The intention is to use the result (2.5) where the mapping $\begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$ constitutes the composition of the inverse stereographic projection (2.11) and the spherical coordinate transform with fixed radius $\frac{1}{2}$ and center point $(0,0,\frac{1}{2}) \in \mathbb{R}^3$. With the transformation as described and the notation $\mathsf{r} = \sqrt{\mathsf{x}^2+\mathsf{y}^2}$ we obtain

$$\begin{cases} \mathsf{z} = \theta\left(\phi^{-1}(P)\right) = \arccos\left(2\frac{\mathsf{r}^2}{1+\mathsf{r}^2} - 1\right) =: f_1(\mathsf{x},\mathsf{y}), \\ \mathsf{w} = \varphi\left(\phi^{-1}(P)\right) = \begin{cases} \pi, & \text{if } y = 0,\ x < 0 \\ \operatorname{sgn}(y)\arccos\frac{x}{\mathsf{r}} =: f_2(\mathsf{x},\mathsf{y}), & \text{otherwise.} \end{cases} \end{cases} \tag{4.1}$$

Since w should lie in the half-open interval $(-\pi,\pi]$ we have to utilize wraparound to declare that should the mapping result in $-\pi$, which happens for realizations $y = 0, x < 0$, we will instead map it to π. This mapping is a bijection in every point except for the realization $x = y = 0$ for which the corresponding realization w is undefined, since both the coordinate transformation as well as the stereographic projection are bijections when not considering the origin. For every realization x and y this yields the Jacobian matrix

$$\begin{pmatrix} \frac{-2x}{(1+r^2)r} & \frac{-2y}{(1+r^2)r} \\ \frac{-y}{r^2} & \frac{x}{r^2} \end{pmatrix} \tag{4.2}$$

where $r = \sqrt{x^2 + y^2}$, for when the Jacobian is defined. The Jacobian is undefined precisely when $y = 0$, $x \leq 0$. This lets us characterize the distribution by the following equality, when the Jacobian is defined

$$f_{\mathsf{zw}}(z,w) = \frac{1}{|J(x,y)|} f_{\mathsf{xy}}(x,y).$$

Here $J(x,y)$ is the determinant of (4.2) given by

$$J(x,y) = \frac{-2}{r + r^3}$$

and again $x(z,w)$ and $y(z,w)$ are the unique solutions to (4.1) for fixed realizations $\mathsf{z} = z, \mathsf{w} = w$. Again, this holds true for points not on the negative x-axis or the origin. Thus we finally get, using the absolute value in (2.5)

$$f_{\mathsf{zw}}(z,w) = \frac{r + r^3}{2} f_{\mathsf{xy}}(x,y). \tag{4.3}$$

We need to contend with f_{zw} being undefined for z, w such that $x = y = 0$. This is only the case for $z = \pi$ corresponding to the coordinates for the south pole of the Riemann sphere where w is undefined. As stated previously, $y = 0$, $x < 0$ also leads to f_{zw} being undefined. However, in the z, w space the region where it is undefined has measure zero and can as such be worked around when integrating regions which contain it. Further still, when considering the transformation using distribution theory such that the discontinuity results in some terms involving a Kronecker delta distribution in the partial derivative of $\varphi(\phi^{-1}(P)))$ instead of being undefined, all such terms cancel yielding the same result as in (4.3). Since the support set $\mathbf{S}$ is bounded, the case when r approaches infinity need not be contended with as f_{xy} is guaranteed 0 for large enough r.

4.2 Finding the Distribution Function of the Chordal Distance

As per the restrictions of this thesis, we will only consider the SISO and point-wise case. The ν-gap simplifies to the supremum over frequency of the point-wise chordal distance $\kappa(\bar{P}(\omega), P(\omega))$ as per (2.13). Thus, we will now investigate how we can characterize the randomness of the point-wise chordal distance. We shall see that given the setting of $P \sim f_P$ on $\mathbf{S}$ and known $\bar{P}$ from the previous section, we will be able to find the cumulative distribution function $F_\kappa(d)$ expressed in terms of f_P. We will again deal with the general case where $\mathbf{S}$ at a fixed frequency is an arbitrary compact set and f_P is an arbitrary distribution. In this section, we will present two separate perspectives on this general characterization along with a special case of one of them.

Deriving an explicit CDF

In this section, the main result of this thesis will be derived and stated. For clarity, in this section, we will let $\mathbf{S} \subset \mathbb{C}$ remain completely arbitrary and only implicitly affect the results of this section. By considering f_P defined on $\mathbb{C}$ we can capture the implication of $\mathbf{S}$ by defining it to be non-zero only on some introduced $\mathbf{S}$. This allows for $\mathbf{S} = \mathbb{C}$ and the effect of a restriction of $\mathbf{S}$ appears through a selection of f_P.

Recall the definition of the point-wise chordal distance (2.12) where the frequency response $P(j\omega)$ is a random quantity for every frequency ω. Due to the randomness of P, the chordal distance becomes a random scalar quantity. We define this random variable for a fixed ω as

$$K := \kappa(P,\bar{P}) = \frac{|\bar{P}-P|}{\sqrt{1+|P|^2}\sqrt{1+|\bar{P}|^2}}. \tag{4.4}$$

We are interested in studying its cumulative distribution function

$$F_K(d) := \mathbb{P}(K < d). \tag{4.5}$$

Before stating the main result we will first introduce several notations for the sake of brevity.

$$c_1(z,w) := \frac{1}{2} - \frac{z-w+1}{2r^2} \tag{4.6a}$$

$$c_2(z,w) := \frac{1}{2}\sqrt{\frac{2(z+w-1)}{r^2} - \frac{(z-w+1)^2}{r^4} - 1} \tag{4.6b}$$

$$\begin{bmatrix} x_i(z,w) \\ y_i(z,w) \end{bmatrix} = c_1(z,w)\begin{bmatrix} a \\ b \end{bmatrix} + (-1)^{i-1} c_2(z,w)\begin{bmatrix} -b \\ a \end{bmatrix},\ i = 1,2 \tag{4.6c}$$

$$u(d,t) := \min\{td^2(1+r^2), (r+\sqrt{t-1})^2\} \tag{4.6d}$$

$$\ell(d,t) := \min\{td^2(1+r^2), (r-\sqrt{t-1})^2\} \tag{4.6e}$$

We shall also require the following lemma before continuing onwards.

LEMMA 1
Given real random variables x and y, with joint distribution $f_{\mathsf{xy}}(x,y)$, and constants $a,b \in \mathbb{R}$, define $r := \sqrt{a^2+b^2}$, and the real random variables

$$\mathsf{z} := g_1(\mathsf{x},\mathsf{y}) = (\mathsf{x}-a)^2 + (\mathsf{y}-b)^2 \geq 0, \tag{4.7a}$$

$$\mathsf{w} := g_2(\mathsf{x},\mathsf{y}) = \mathsf{x}^2 + \mathsf{y}^2 + 1 \geq 1. \tag{4.7b}$$

Then, the joint distribution of z and w is given by

$$f_{\mathsf{zw}}(z,w) = \begin{cases} f(z,w), & |r-\sqrt{w-1}| < \sqrt{z} \\ \text{undefined}, & |r-\sqrt{w-1}| = \sqrt{z} \\ 0, & \text{otherwise} \end{cases} \tag{4.8}$$

where

$$f(z,w) := \frac{1}{4r^2 c_2(z,w)} \sum_{i=1}^{2} f_{\mathsf{xy}}(x_i(z,w), y_i(z,w)) \tag{4.9}$$

Proof. We will use the known result (2.5) where x_i, y_i are all solutions to the mapping (4.7) for the corresponding z, w. Furthermore $J(x_i,\ y_i)$ is the Jacobian determinant of the mapping evaluated at the solutions $x_i,\ y_i$ and is given by

$$|J(x,y)| = 4\left|\begin{bmatrix} b & -a \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right|. \tag{4.10}$$

We can interpret (4.7) as circles with radii $\sqrt{\mathsf{w}-1}$ and $\sqrt{\mathsf{z}}$ and centers in the origin and $\begin{bmatrix} a \\ b \end{bmatrix}$ respectively, illustrated in Figure 4.1.

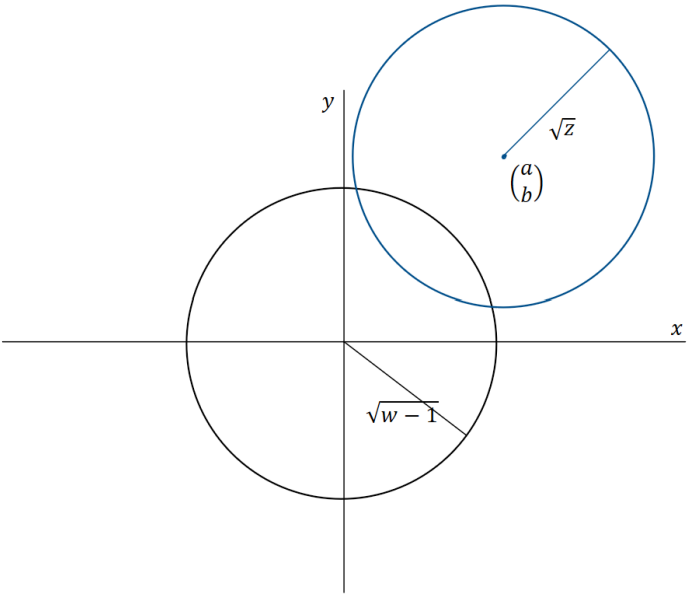


Figure 4.1 An interpretation of (4.7) in terms of circles.

Then x_i, y_i are the points of intersection of the circles given by the following cases namely:

1. When $|r-\sqrt{w-1}| < \sqrt{z}$ and $z > 0,\ w > 1$, we have two intersection points which are given by (4.6c).
2. When $|r-\sqrt{w-1}| = \sqrt{z}$ and $z > 0,\ w > 1$, we get $c_2(z,w) = 0$ and leads to a unique intersection at $\begin{bmatrix} x_1 & y_1 \end{bmatrix}^\top = c_1(z,w) \begin{bmatrix} a & b \end{bmatrix}^\top$.
3. For all other cases, there is no intersection.

Inserting these results into (4.10), we obtain the following results for when two intersections exist.

$$\begin{aligned} &|J(x_i(z,w),\, y_i(z,w))| \\ &= 4 \left| c_1(z,w) \begin{bmatrix} b & -a \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \pm c_2(z,w) \begin{bmatrix} b & -a \end{bmatrix} \begin{bmatrix} -b \\ a \end{bmatrix} \right| \\ &= 4r^2 c_2(z,w). \end{aligned}$$

Inserting this into (2.5) for each case yields (4.8).

We also need to show that the PDF f_{zw} is valid. To do so, we first note that it is easy to show it is non-negative. We also need to show that the integral of f_{zw} over the region in the z-w plane defined by the cases in (4.8) converges. Formally, the integral should also equal 1 when computed on its support. In this proof, this calculation is omitted since the equality (2.5) guarantees the correct normalization provided the integral is convergent.

We want to find

$$F_{\mathsf{K}}(1) = \int_1^\infty \int_{(r-\sqrt{t-1})^2}^{(r+\sqrt{t-1})^2} \frac{1}{4r^2 c_2(l,t)} \sum_{i=1}^{2} f_{\mathsf{xy}}(x_i, y_i)(l,t)\, dl\, dt. \tag{4.11}$$

Now, utilize the following change of variables

$$u = c_1(l,t) = \left(\frac{1}{2} - \frac{l-t+1}{2r^2} \right), \tag{4.12a}$$

$$v = c_2(l,t) = \frac{1}{2} \sqrt{\frac{2(l+t-1)}{r^2} - \frac{(l-t+1)^2}{r^4} - 1}. \tag{4.12b}$$

The mapping has the absolute Jacobian determinant

$$|J(l,t)| = \frac{1}{8r^4 c_2(l,t)}. \tag{4.13}$$

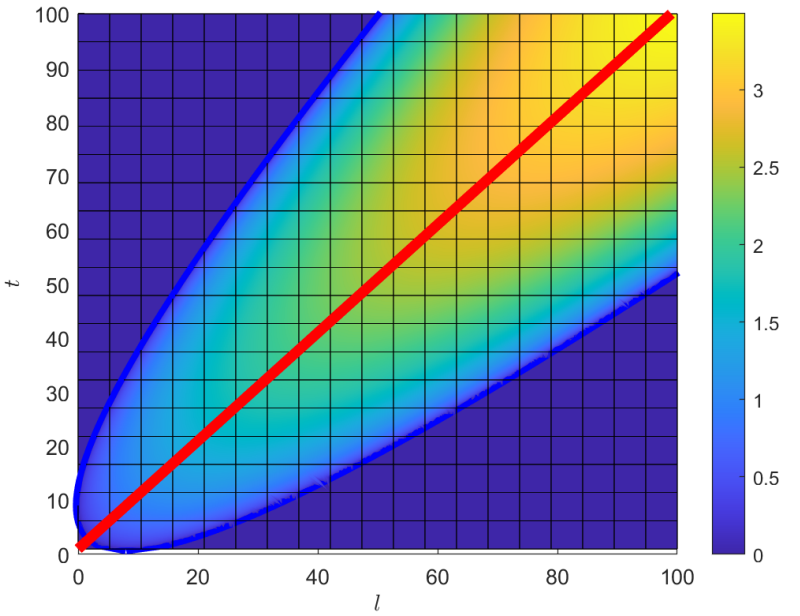


Figure 4.2 A plot of $c_2(l,t)$ for $r = 2\sqrt{2}$ with the region limits given by $|r - \sqrt{l}| < \sqrt{t-1}$ marked by blue solid lines and the red solid line constitutes the symmetry line $t = l + 1$. The disjoint regions on either side of the symmetry line and limited by the blue solid lines correspond to the regions which after rotation and translation into the l_r, t_r plane coincide with E_1 and E_2.

We need the Jacobian of the inverse transformation of (4.12) to perform the variable change which requires the transformation to be invertible. Since c_1 is linear in l and t it is invertible but c_2 is not invertible on the entire integration region. However, the function is invertible on either side of the symmetry line $t = l + 1$ visualized by Figure 4.2. This fact is verified by the following calculation. Applying the bijective transformation $(l_r, t_r) = \frac{1}{\sqrt{2}}(l - (t-1), l + (t-1))$ (corresponding to a translation and rotation) yields

$$c_2(l_r, t_r) = \frac{1}{2}\sqrt{\frac{2\sqrt{2}}{r^2}t_r - \frac{2}{r^4}l_r^2 - 1}$$

which is symmetric around the t_r axis. Furthermore, it is monotone on either side of the t_r axis and increasing for $t_r > 0$ and fixed l_r and thus invertible. Transforming back maintains this property but transforms the symmetry line $l_r = 0$ into $t = l + 1$ as promised. Dividing the integral (4.11) into the two disjoint regions E_1, E_2 separated by this line guarantees (4.12b) is invertible in each region. As such, we can use the result that the Jacobian of the inverse is the inverse of the Jacobian to transform

(4.11). Now let D_1 and D_2 be the mapping of E_1 and E_2 by (4.12) respectively in the $u-v$ plane. We also write

$$h(l,t) := \sum_{i=1}^{2} f_{\mathsf{xy}}(x_i(l,t), y_i(l,t)). \tag{4.14}$$

Note that, after the transformation in the following we have $l = l(u,v)$ and $t = t(u,v)$ to get

$$F_{\mathsf{K}}(1) = \int\int_{E_1 \cup E_2} \frac{h(l,t)}{4r^2 c_2(l,t)}\, dl\, dt, \tag{4.15}$$

$$= \int\int_{D_1} 2r^2 h(l,t)\, du\, dv + \int\int_{D_2} 2r^2 h(l,t)\, du\, dv. \tag{4.16}$$

Since we know h from (4.14) is a sum of two PDFs, we know it is integrable and therefore both integrals in the last step are convergent which proves the result. □

Now we are ready to state the main result.

THEOREM 1
Given $d \in [0,1]$, let $P(j\omega) = \mathsf{x} + \mathsf{y}j \in \mathbb{C}$ for a fixed ω follow the distribution f_{P} parameterized by $f_{\mathsf{xy}}(x,y)$. Further, let $\bar{P}(j\omega) \in \mathbb{C}$ denote a nominal model at the same frequency ω. Then, the CDF of K is given by:

$$F_{\mathsf{K}}(d) = \int_1^{\infty} \int_{\ell(d,t)}^{u(d,t)} \frac{1}{4r^2 c_2(l,t)} \sum_{i=1}^{2} f_{\mathsf{xy}}(x_i, y_i)(l,t)\, dl\, dt, \tag{4.17}$$

where $r = |\bar{P}(j\omega)|$ and notations follow (4.6).

Proof. Let us parameterize the known nominal model as $\bar{P} = a + bj$ and the unknown random model as $P = \mathsf{x} + \mathsf{y}j$ at the given frequency. As a consequence x and y are real random variables with a known density $f_{\mathsf{xy}}(x,y)$ derived from f_{P}. Then we have

$$K = \frac{1}{\sqrt{1+r^2}} \sqrt{\frac{g_1(\mathsf{x},\mathsf{y})}{g_2(\mathsf{x},\mathsf{y})}} =: \frac{1}{c}\sqrt{Q}$$

where g_1, g_2 are given by (4.7). Then, for a given $d \in [0,1]$, the CDF of K using Lemma 1 is given by

$$F_{\mathsf{K}}(d) = \mathbf{P}\left(\frac{\sqrt{Q}}{c} < d\right) = \mathbf{P}\left(Q < c^2 d^2\right) = F_{\mathsf{Q}}(c^2 d^2),$$

where $F_{\mathsf{Q}}(\cdot)$ denotes the CDF of $Q = \frac{\mathsf{z}}{\mathsf{w}}$ and $\mathsf{z} := g_1(\mathsf{x},\mathsf{y})$, $\mathsf{w} := g_2(\mathsf{x},\mathsf{y})$. In (2.6), F_{Q} is stated as

$$F_{\mathsf{Q}}(q) := \int_0^\infty \int_0^{tq} f_{\mathsf{zw}}(l,t)\, dl\, dt$$

since g_1 and g_2 guarantees non-negative z and w. Here, $f_{\mathsf{zw}}(l,t)$ denotes the joint PDF of z and w obtained using Lemma 1. By realizing that $f_{\mathsf{zw}}(z,w)$ is zero outside the region specified by $\left|r - \sqrt{w-1}\right| < \sqrt{z}$, we get

$$F_{\mathsf{Q}}(q) = \int_1^\infty \int_{\min\{tq,\left(r-\sqrt{t-1}\right)^2\}}^{\min\{tq,\left(r+\sqrt{t-1}\right)^2\}} f(l,t)\, dl\, dt, \quad \text{where}$$

$$f(l,t) = \frac{1}{4r^2 c_2(l,t)} \sum_{i=1}^{2} f_{\mathsf{xy}}(x_i(l,t), y_i(l,t)).$$

Here, $x_i(l,t), y_i(l,t)$ are given by (4.6c). Setting $q = c^2 d^2 = (1+r^2)d^2$ yields the final result. □

A General Perspective

In this section we shall provide another, more general perspective on how to find the cumulative distribution function (4.5). Investigating the function using sets instead of directly manipulating (2.12), we arrive at an alternative expression which can be more easily generalized to MIMO systems since the definition (2.12) only holds for SISO systems as stated. While the following still will regard the SISO case, we will treat κ purely as a metric, independent of how we defined it in (2.12). As such, a corresponding MIMO definition could take its place with only minor changes to the argument that follows.

Firstly, define the open ball

$$\mathscr{B}_d(\bar{P}) := \{P \in \mathbb{C} : \kappa(P,\bar{P}) < d\}. \tag{4.18}$$

Using this we can find the following expression for (4.5)

$$F_K(d) = \mathbf{P}(\kappa(\bar{P},P) < d) = \mathbf{P}(P \in \mathscr{B}_d(\bar{P})) = \int_{\mathscr{B}_d(\bar{P}))} f_P\, dA$$

where dA is an area element on $\mathbb{C}$. Since the distribution f_P is only non-zero on its support set $\mathbf{S}$ we finally find

$$F_K(d) = \int_{\mathscr{B}_d(\bar{P}))\cap\mathbf{S}} f_P\, dA \tag{4.19}$$

With a given $\mathbf{S}$, f_P and d, this integral is well defined, albeit perhaps difficult to compute for certain regions and distributions.

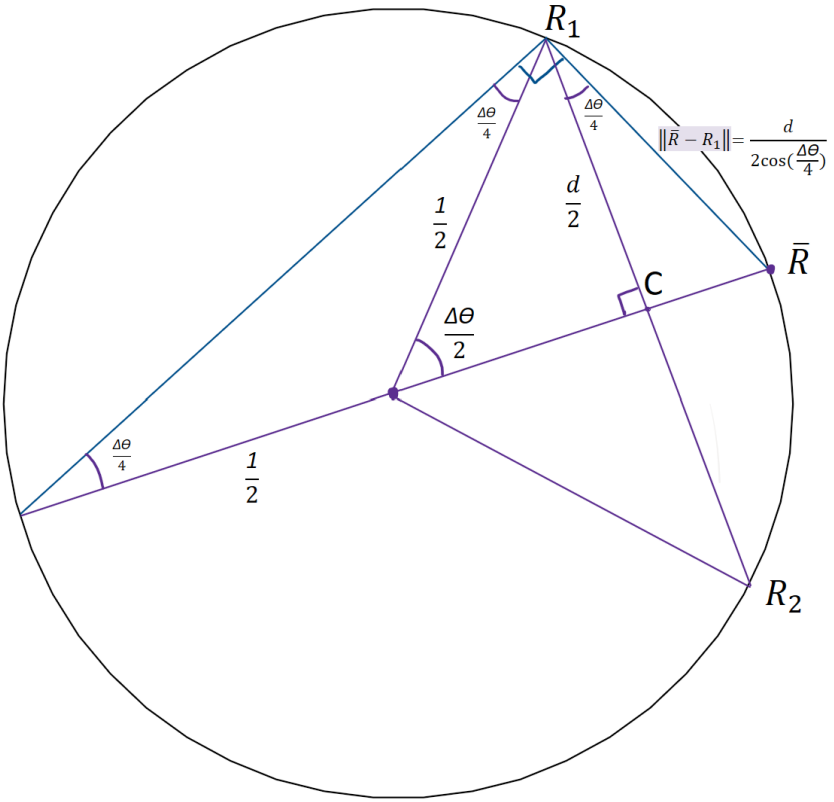


Figure 4.3 Illustration of how chordal diameter relates to chordal radius.

Note that the open ball $\mathscr{B}_d(\bar{P})$, when projected onto $\partial\mathfrak{R}$ constitutes the intersection between $\partial\mathfrak{R}$ and an open ball in $\mathbb{R}^3$ centered on $\bar{R} := \phi^{-1}(\bar{P})$ with the $\mathbb{R}^3$-Euclidean radius d. Its boundary is a circle on $\partial\mathfrak{R}$ with a center point C that lies inside $\mathfrak{R}$ on the intersection between the radial line to $\bar{R}$ and a chordal diameter line. Formally we have

$$\phi^{-1}\left(\mathscr{B}_d(\bar{P})\right) = \{R \in \partial\mathfrak{R} : \left\|R - \phi^{-1}(\bar{P})\right\|_2 < d\}.$$

A cross-section of this for a vertical plane parallel with the azimuthal angle of P is shown in the illustration in Figure 4.3.

We also note that given a circle on $\partial\mathfrak{R}$ with chordal *diameter* d centered on the point C inside $\mathfrak{R}$ on the radial line to $\bar{R} \in \partial\mathfrak{R}$. The chordal distance from a point R_1 on the the circle to $\bar{R}$, which we will call the chordal radius, will be $\frac{d}{2}\left(\cos\frac{\Delta\theta}{4}\right)^{-1}$ as shown in Figure 4.3, where $\sin\frac{\Delta\theta}{2} = d$. This fact will be used in the following section in the formulation of an important lemma.

Uniform Distribution with Elliptic Support

The previous result (4.19), while numerically solvable for known nominal models $\bar{P}$, sets $\mathbf{S}$ and distributions f_P, has no guarantee for an analytic solution and could present with numerical issues for complicated cases. We will now consider a special case where the integral can be alternatively expressed in a more easily computable way.

Consider $\mathbf{S}$ to be an ellipse with major and minor semi-axes $a, b > 0$ and f_P to be a uniform distribution on $\mathbf{S}$. This setting might be suitable when the engineer can place bounds on the uncertainty in gain and phase and thus bound a point on the Nyquist plot in the radial and tangential directions, but has no knowledge of the distribution inside the bounds. A good model to capture this could be to assume no prior distribution and thus assume every model in the set to be equally likely leading to a uniform distribution. In this case the integral (4.19) turns into a simple proportion of area. With

$$f_P = \begin{cases} \frac{1}{\int_{\mathbf{S}} dA}, & \text{on } \mathbf{S} \\ 0, & \text{otherwise} \end{cases}$$

we get

$$F_K(d) = \frac{\int_{\mathscr{B}_d(\bar{P}))\cap \mathbf{S}} dA}{\int_{\mathbf{S}} dA}. \tag{4.20}$$

Before stating the result, we will once again require a lemma. The open ball (4.18) is defined by a center point and a radius described in terms of the chordal distance, and thus the distance of points on $\partial\mathfrak{R}$ to the inverse stereographic projection $\bar{R} = \phi^{-1}(\bar{P})$. The following lemma will describe how the ball can be alternatively expressed using the Euclidean norm on $\mathbb{C}$. We shall also see that the center point of the resulting disc in $\mathbb{C}$ is not necessarily $\bar{P}$ and in fact depends on $d \in [0,1]$.

LEMMA 2
Let $\bar{R} = (x,y,z) \in \partial\mathfrak{R}$ be given. Then, given a standard open ball $\mathscr{B}_{\frac{d}{2}\left(\cos\frac{\Delta\theta}{4}\right)^{-1}}\left(\overline{R}\right) \in \mathbb{R}^3$ of diameter $d \in [0,1]$ in the Euclidean metric that does not include the north pole of the Riemann sphere $(0,\ 0,\ 1)$, the open disc $D := \phi\left(\mathscr{B}_{\frac{d}{2}\left(\cos\frac{\Delta\theta}{4}\right)^{-1}}\left(\overline{R}\right) \cap \mathfrak{R}\right) \subset \mathbb{C}$ has a diameter $\hat{d}$ in the Euclidean metric given by

$$\hat{d} = \left| \frac{\sin(\theta+\Delta\theta)}{1-\cos(\theta+\Delta\theta)} - \frac{\sin\theta}{1-\cos\theta} \right|, \quad \text{with} \tag{4.21a}$$

$$\sin\frac{\Delta\theta}{2} = d, \text{ and } \theta = \theta(\bar{R}) \text{ such that } \cos\left(\theta + \frac{\Delta\theta}{2}\right) = 2z - 1. \tag{4.21b}$$

The stereographic projection ϕ is given by

$$\phi\colon \mathfrak{R} \to \mathbb{C},$$
$$\phi(x,y,z) = \frac{x}{1-z} + \frac{y}{1-z}j.$$

Furthermore, the center point $\hat{P} \in \mathbb{C}$ of the disc D is given by

$$\hat{P} = \frac{\left|\frac{\sin(\theta+\Delta\theta)}{1-\cos(\theta+\Delta\theta)}\right| + \left|\frac{\sin\theta}{1-\cos\theta}\right|}{2}\left(\cos\varphi(\bar{R}) + \sin\varphi(\bar{R})j\right), \quad \text{where} \tag{4.22a}$$
$$\varphi(\bar{R}) \colon = \operatorname{sign}(y)\arccos\frac{x}{\sqrt{x^2+y^2}} \tag{4.22b}$$

Proof. We know that a circle which does not intersect the north pole on the Riemann sphere maps to another circle on the complex plane through the stereographic projection [Vinnicombe, 2000]. Furthermore, a meridian on the Riemann sphere maps to a straight radial line. An illustration in Figure 4.4 showcases the main argument of the proof. Take some circle on $\mathfrak{R}$ with chordal diameter $d \in [0,1]$ and some center point $\bar{R} = (x,y,z)$. Now consider the meridian through $\bar{R}$ and its intersection with the circle which yields two points R_1 and R_2 between which the chordal distance is d. Figure 4.3 illustrates a cross section of a plane in which such a meridian lies. Furthermore, $P_1 \colon = \phi(R_1)$ and $P_2 \colon = \phi(R_2)$ will lie on the opposite sides of the projected circle and thus at a distance $\hat{d}$ equal to the projected diameter. Consider $R_1 = (\frac{1}{2}, \theta, \varphi)$ and $R_2 = (\frac{1}{2}, \hat{\theta}, \varphi)$, with $\hat{\theta} = \theta + \Delta\theta$ to be the points described above given in spherical coordinates with its origin in $(x,y,z) = (0,0,\frac{1}{2})$. The radial coordinate is fixed at $\frac{1}{2}$ due to the points being on $\mathfrak{R}$. They lie on the meridian at the fixed azimuth angle φ and at a chordal distance d from each other such that $\sin\frac{\Delta\theta}{2} = d$. We know φ only determines the direction of the projected diameter line but not its length. As such, $\hat{d}$ is invariant of φ and we set it equal to zero. Now, equivalently consider the points in Cartesian coordinates to get

$$R_1 = \frac{1}{2}(\sin\theta, 0, 1+\cos\theta), \quad R_2 = \frac{1}{2}\left(\sin\hat{\theta}, 0, 1+\cos\hat{\theta}\right).$$

Projecting these onto $\mathbb{C}$ using (2.9), we get real points

$$P_1 = \frac{\sin\theta}{1-\cos\theta}, \quad P_2 = \frac{\sin\hat{\theta}}{1-\cos\hat{\theta}}.$$

We then get

$$\hat{d} = |P_2 - P_1| = \left|\frac{\sin(\theta+\Delta\theta)}{1-\cos(\theta+\Delta\theta)} - \frac{\sin\theta}{1-\cos\theta}\right|$$

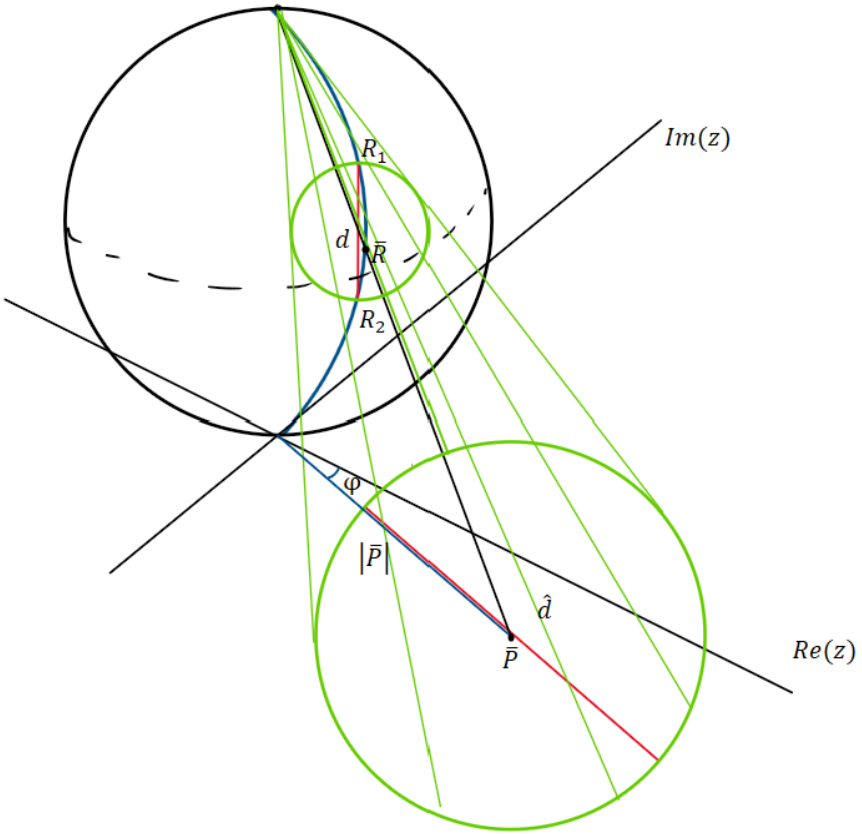


Figure 4.4 An illustration of the argument used in the proof of Lemma 2.

where

$$\sin\frac{\Delta\theta}{2} = d, \quad \text{and} \quad \theta = \theta(\bar{R}) \text{ such that } \cos\left(\theta + \frac{\Delta\theta}{2}\right) = 2z - 1.$$

Furthermore we know the center point is given by $\frac{P_1+P_2}{2}$. Since direction now plays a part we need to consider the azimuth angle φ again, however we note that the distance from the origin to P_1 and P_2 respectively is still unaffected. We therefore find for any $\varphi \in [0,\, 2\pi)$ that

$$P_1 = \frac{\sin\theta}{1-\cos\theta}\,(\cos\varphi + \sin\varphi\, j)\,, \quad P_2 = \frac{\sin\hat{\theta}}{1-\cos\hat{\theta}}\,(\cos\varphi + \sin\varphi\, j)$$

where $\varphi = \varphi(\bar{R})$ is described by (4.22b) and the expression for the center point $\hat{P}$ follows from $\hat{P} = \frac{p_1+p_2}{2}$ yielding the result. □

While not strictly used in the derivation of the following theorem and only referred to, this lemma provides a useful bridge between the setting we introduced in the previous section to arrive at (4.19) and the starting point of the following theorem. We state the theorem here and comment on the connection to the lemma afterwards.

THEOREM 2
Suppose $\mathbf{S} \in \mathbb{C}$ is an ellipse centered in $\bar{P} \in \mathbb{C}$ with minor-axis $B > 0$, major-axis $A \geq B$ and angle between major axis and the real axis φ_e (in positive direction). Let $D \in \mathbb{C}$ be the open disc with diameter $\hat{d} > 0$ and center in $\hat{P} \in \mathbb{C}$. Then the following holds

$$\frac{\int_{D\cap\mathbf{S}} dA}{\int_{\mathbf{S}} dA} = \frac{1}{AB\pi} \int_{\max\left(-A, c_x - \frac{\hat{d}}{2}\right)}^{\min\left(A, c_x + \frac{\hat{d}}{2}\right)} (U(x) - L(x))\, dx \tag{4.23}$$

where

$$\begin{pmatrix} c_x \\ c_y \end{pmatrix} := M(-\varphi_e) \left(\begin{pmatrix} Re(\hat{P}) \\ Im(\hat{P}) \end{pmatrix} - \begin{pmatrix} Re(\bar{P}) \\ Im(\bar{P}) \end{pmatrix} \right) + \begin{pmatrix} 0 \\ B \end{pmatrix} \tag{4.24a}$$

$$D_i(x) := (-1)^{i-1} \sqrt{\frac{\hat{d}^2}{4} - (x - c_x)^2} + c_y, \quad i = 1,2 \tag{4.24b}$$

$$E_i(x) := (-1)^{i-1} B \sqrt{1 - \left(\frac{x}{A}\right)^2} + B, \quad i = 1,2 \tag{4.24c}$$

$$U(x) := \max(E_2(x), \min(D_1(x), E_1(x))) \tag{4.24d}$$

$$L(x) := \min(E_1(x), \max(D_2(x), E_2(x))) \tag{4.24e}$$

and $M(-\varphi_e)$ denotes the standard 2×2 rotation matrix of angle $-\varphi_e$.

Proof. Since the left hand side of (4.23) is a proportion of areas the expression is invariant of translation and rotation. As such, we introduce the transformation from (4.24a) where we interpret $\begin{pmatrix} c_x \\ c_y \end{pmatrix}$ as real and imaginary parts of $c_x + c_y j$ respectively. As such consider it as a mapping $T\colon\ \mathbb{C} \to \mathbb{C}$ where (4.24a) showcases $T(\hat{P}) = c_x + c_y j$. It consists of a translation by $-\bar{P}$ and rotation by $-\varphi_e$ which centers the ellipsis and aligns its major axis with the x-axis (real axis). Then a shift of B along the y-axis (imaginary axis) lifts the ellipse to where it tangents the x-axis at its south pole. The procedure is shown in Figure 4.5. Applying the transformation yields the right figure in Figure 4.5. We split the circle D and ellipse E into its upper and lower halves in order to be able to express each curve segment as functions $D_i(x)$, $E_i(x)$, $i = 1,2$ as defined in (4.24b) and (4.24c) and illustrated in Figure 4.6. The area of the ellipsis is $AB\pi$ and the area of the intersection can now be calculated as the integral of the

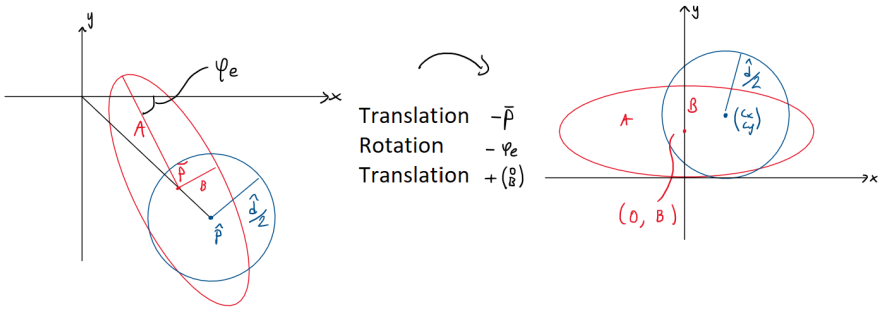


Figure 4.5 Visual representation of transformation utilized in the proof for Theorem 2.

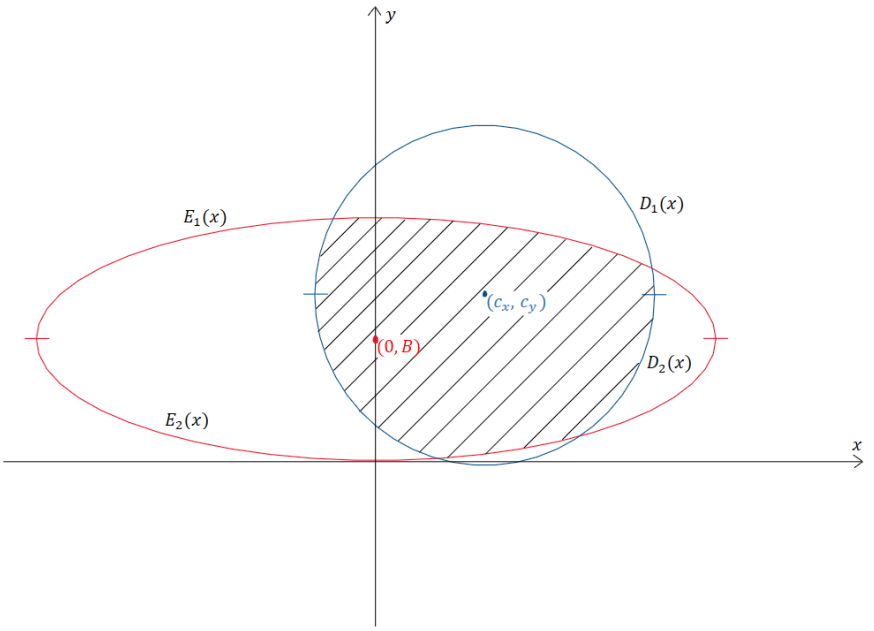


Figure 4.6 An example illustration of the D_i and E_i notations and the shaded area to be represented by the integral factor in (4.23).

difference between the upper and lower segments of the intersection boundary as shown in Figure 4.6. The integral limits are determined by either the ends of the ellipsis or the circle in the x-direction, whichever is more limiting. The integrand consists of the difference between the upper and lower sections for all different cases of c_x, c_y and $\hat{d}$ resulting in (4.24d) and (4.24e). The final result then follows.

Remark: As promised we will discuss the connection between Theorem 2 and Lemma 2 briefly. We note that if the point $\hat{P}$ and diameter $\hat{d}$ in Theorem 2 were determined by the situation described in Lemma 2 then D would correspond to

$$\mathscr{B}_{\frac{d}{2}\left(\cos\frac{\Delta\theta}{4}\right)^{-1}}(\bar{P}) \in \mathbb{C}$$

as defined by (4.18). The integral result (4.23) would then be equal to $F_K\left(\frac{d}{2}\left(\cos\frac{\Delta\theta}{4}\right)^{-1}\right)$ providing a way to use this to calculate the CDF. This would hold for all $d \in [0,1]$ such that $\min(\theta(\bar{R}), \theta(\bar{R}) + \Delta\theta) > 0$ which guarantees that the inversely projected ball does not intersect or encircle the north pole.

CDF Result Summary

In this chapter we have now provided two different perspectives on the CDF of the chordal distance along with a numerical example showcasing the first of the two. In the following section we will provide a result which makes use of the knowledge of the CDF to place probabilistic guarantees on the point-wise generalized stability margin ρ for a stabilizing controller with respect to the nominal system $\bar{P}$. This provides a point-wise version of the answer to Question 1 and provides a suitable conclusion to the series of results provided in this thesis before finally showcasing some numerical examples.

4.3 Potential Performance Degradation and Violation Probability Duality

With knowledge of $F_K(d)$ we are interested in how a random perturbation from $\bar{P}$ (a realization of P) impacts performance on a controller $\bar{C}$ which stabilizes $\bar{P}$. Recall (2.18) holds for any P if $\bar{C}$ stabilizes $\bar{P}$. We state the following result

THEOREM 3
Let $\bar{P}$ be a known system for which $\bar{C}$ is a stabilizing controller. Further let P be a random system following the distribution f_P on some set $\mathbf{S}$, both known for every frequency ω. If, for some frequency ω the random variable $K := \kappa(P(j\omega), \bar{P}(j\omega))$ is determined by the known cumulative distribution $F_K(d)$, $d \in [0,1]$, derived from f_P. Then the following holds:

$$P\left(\rho_{P,\bar{C}} \geq \rho_{\bar{P},\bar{C}} - d\right) \geq F_K(d) \tag{4.25}$$

Proof. The proof follows from direct calculation where the last step uses that the inequality (2.18) given a stable pair $[\bar{P}, \bar{C}]$ always holds. We have

$$\begin{aligned} F_K(d) &= P(K \leq d) = P\left(K - \rho_{\bar{P},\bar{C}} \leq d - \rho_{\bar{P},\bar{C}}\right) = \\ &= P\left(\rho_{\bar{P},\bar{C}} - d \leq \rho_{\bar{P},\bar{C}} - K\right) \leq P\left(\rho_{\bar{P},\bar{C}} - d \leq \rho_{P,\bar{C}}\right) \end{aligned}$$

which directly provides the result. □

The result of Theorem 3 does what we set out to do and provides an answer to the point-wise version of Question 1. With known $[\bar{P}, \bar{C}]$ we can calculate $\rho_{\bar{P},\bar{C}}$ and pick a $d \in [0,1]$ such that $\rho_{\bar{P},\bar{C}} - d$ constitutes the desired performance requirement, bounded from above by $\rho_{\bar{P},\bar{C}}$. Then Theorem 3 gives a lower bound on the probability that the designed controller $\bar{C}$ satisfies the performance requirement on a random plant P in the uncertainty set $\mathbf{S}$.

We can use the following corollary to provide an answer to the question where if we pick a violation probability, what performance requirement will only be violated with a probability at most equal to that violation probability?

COROLLARY 1
Given a violation probability $\varepsilon \in [0,1]$ the following choice of $d = F_K^{-1}(1-\varepsilon)$, whenever it is properly defined will ensure

$$P\left(\rho_{P,\bar{C}} \geq \rho_{\bar{P},\bar{C}} - d\right) \geq 1 - \varepsilon.$$

Proof. The proof of the corollary follows directly from Theorem 3 by setting $F_K(d) = 1-\varepsilon$ and assuming the inverse to be properly defined at that point. □

Remark : Since F_K is a CDF it is monotone (not strictly), the only case when it is not invertible in a point $1-\varepsilon$ is when there exists an interval $[d_1,\ d_2]$ where $F_K(d) = 1-\varepsilon$ for all $d \in [d_1,\ d_2]$. In such a case we can define the inverse to mean the smallest of such d, namely $F_K^{-1}(1-\varepsilon) := d_1$ and maintain the result.

4.4 Numerical Examples

To provide validation and concretize the results produced in this chapter, this section will provide numerical examples where the results above are put to use. The code used to produce these results along with the simulation in the motivating example can be found on the public repository [Nyström, 2024].

Gaussian Distribution with Unbounded Support

With an expression of the CDF from Theorem 1, we will now study a practical example where it is computed numerically. One such example is the example problem from Section 1.1 where an estimation of the distribution has been made and assumed to be Gaussian. This section will deal with the example where f_P, parameterized as $f_{xy}(x,y)$, is a Gaussian distribution with an unbounded support set $\mathbf{S} = \mathbb{C}$.

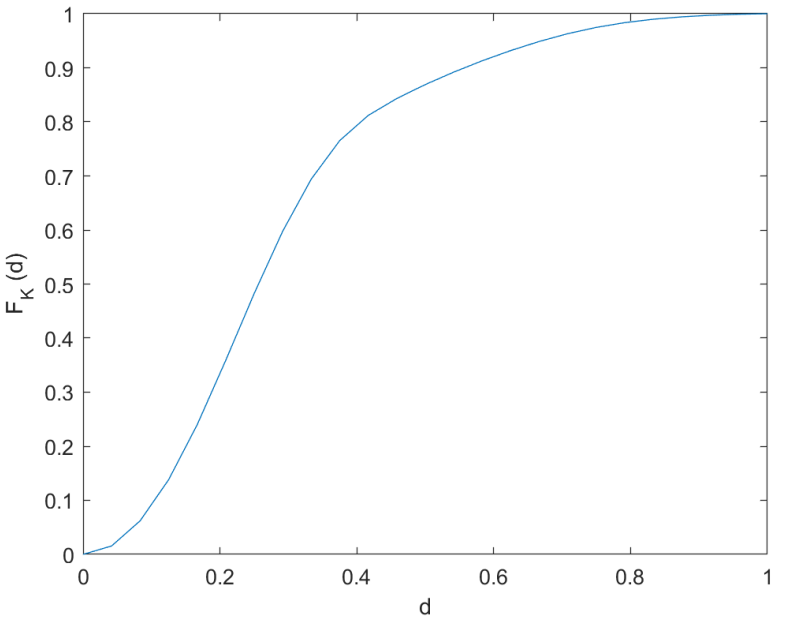


Figure 4.7 Numerical results in calculating $F_K(d)$ as expressed in (4.17) with a Gaussian underlying distribution on unbounded support.

Such a case can be considered to be a generalization of the elliptical support set from Section 1.1 but with a truncation at infinity rather than at 5 standard deviations. This would be a general case of the following. Let f_{xy} have the covariance matrix Σ and the support set be an ellipse given by $k\mathbf{x}^T\Sigma^{-1}\mathbf{x} \leq 1$ for some scalar k and where $\mathbf{x}^T = (x, y)$. This corresponds to the boundary of $\mathbf{S}$ being a level set to the distribution. This is natural since the set in Section 1.1 was determined as the level set at 5 standard deviations. By setting the level at 0 we include all of $\mathbb{C}$. The numerically computed CDF $F_K(d)$, $d \in [0,1]$ given by Theorem 1 of such a case can be seen in Figure 4.7. As seen the resulting interpolated function is a valid CDF with $F_K(1) = 1$ as expected. Here $\bar{P} = 1 + j$ along with $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$ were chosen. For an elliptic restriction on $\mathbf{S}$, the same distribution f_{xy} would apply but be truncated at some cutoff and re-normalized resulting in a new numerical result.

Uniform Distribution with Elliptic support

To illustrate and validate the result given in Theorem 2 as well as provide a connection between the result and the calculation of a CDF, we provide another numerical example where the result is used. Consider a case commensurate with Theorem 2

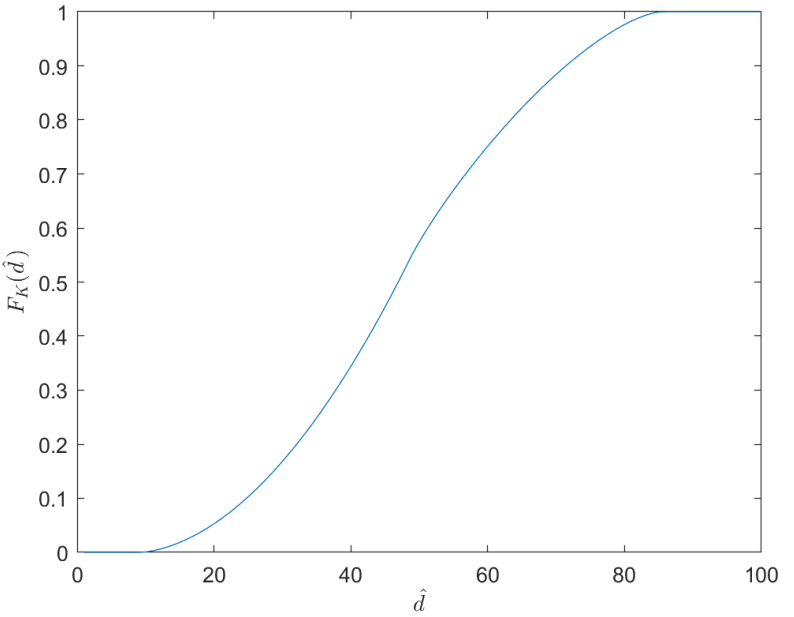


Figure 4.8 An example plot of the numerically calculated CDF of the special case when f_P is uniform with elliptic support. Note that it is expressed in terms of $\hat{d}$ from Lemma 2 here.

such that $P \sim f_P$ where f_P is a uniform distribution with support on the elliptical $\mathbf{S}$ with major-axis $A = 4$ and minor-axis $B = 1$. Let the ellipse be centered in the origin and be rotated to where the angle between major axis and real axis, $\varphi_e = \frac{\pi}{4}$. Furthermore, let $\bar{P} = 1 + j$ and consider how(4.23) varies when we vary $\hat{d}$. This is equivalent to considering its variation with respect to d if one uses the relation between d and $\hat{d}$ from Lemma 2. For the sake of simplicity this example does not consider the relation and instead sweeps across a large enough span of $\hat{d} \in \mathbb{R}_{\geq 0}$ to cover all $d \in [0, 1]$ for this specific case. The result of this procedure is the plot displayed in Figure 4.8 where we once again can see the CDF validity.

5

Conclusion, Discussion and Future Outlook

In the previous chapter, all the produced results of this thesis were described and developed. Out of all of them, the main result and contribution of this thesis is Theorem 1 which can be used to compute the probability that the point-wise chordal distance at some frequency ω between a random SISO LTI system $P(j\omega)$ and a known nominal model $\bar{P}(j\omega)$ is smaller than a given $d \in [0,1]$ when a distribution of $P(j\omega) \in \mathbb{C}$, f_P with support on a set $\mathbf{S} \in \mathbb{C}$ is given. It is complemented by the equivalent result (4.19) which is more easily generalized into the MIMO case, and a specific use case of (4.19) which leads to an explicit expression in Theorem 2. The two results Theorem 1 and (4.19) yield two alternative ways of expressing the cumulative distribution function of the point-wise chordal distance. What follows from a determined CDF is that one can then place probabilistic guarantees on the point-wise performance measure ρ for a controller designed to stabilize $\bar{P}$ when applied to the random system P. These guarantees are specified by Theorem 3 and Corollary 1.

5.1 Discussion

The guarantees are two-fold and makes it possible to firstly answer the question: Given a lower bound on the point-wise performance, what is the lower bound on the probability that it will be achieved? and secondly: Given a violation probability threshold, what is the largest performance requirement one can set which will be violated with at most that probability? This provides a clear bridge between the theory produced in this thesis and practical application as these two questions are pertinent in a control engineering setting where performance guarantees are not strict. In some cases, the control of a process must satisfy some requirements at all points in time, but there are certainly situations where some leeway is acceptable.

In such a situation a probabilistic requirement is sufficient and can be determined using the theory in this thesis.

Further still, it is clear that Corollary 1 implies an inverse relationship between the violation probability and the allowed distance d where a smaller violation threshold yields a larger d which is associated with a larger uncertainty region and a poorer performance bound. As such, this theory can be used to determine a one-to-one correspondence between different levels of performance and a violation probability, providing a high degree of clarity into the performance-violation probability trade-off.

This can also be used to determine if the uncertainty set $\mathbf{S}$ is small enough to be useful. We noted earlier that if the uncertainty set stems from data collection, it is generally the case that more data collection yields a smaller uncertainty set, assuming the data provides new information and is free from bias. After collecting data and utilizing the theory in this thesis to illuminate the performance degradation-violation probability duality, it could be the case that the limitation provided by the duality is too harsh for the application and that required performance with at most a certain violation probability is impossible to achieve. In that case more data would be needed in order to reduce the uncertainty enough to guarantee better performance with the same violation probability. The question of interest is then the following.

Given a violation probability, how much data do I need to collect?

Corollary 1 implies, if we have a violation probability we can investigate what requirements are possible given F_K. If better requirements are needed one can find out how small $d = F_K^{-1}(1-\varepsilon)$ has to be and from there find out how much data is needed to be collected in order to achieve that level of certainty. For a sufficiently bounded set $\mathbf{S}$, the intersection between $\mathbf{S}$ and $\mathscr{B}_d(\bar{P})$ in (4.19) becomes identically $\mathbf{S}$ for some large enough $d \in [0,1]$, say $\bar{d}$. Then $F_K(d) = 1$, $\forall d \in [\bar{d},1]$ which by the remark related to Corollary 1 guarantees that even for a violation probability of $\varepsilon = 0$, the lower bound $\rho_{\bar{P},\bar{C}} - \hat{d}$ would be guaranteed with probability 1. Further investigation into exactly what restrictions need to be in place on $\mathbf{S}$ would likely not be trivial and constitutes a potentially fruitful area for future research.

As was shortly discussed previously, compactness of $\mathbf{S}$ was not necessary for any of the above results. However, when taking this research onwards to handle the ν-gap, the winding number condition needs to be contended with. One initial way of doing so is to restrict $\mathbf{S}$ to never include -1 for any frequency and combine with sufficient restrictions on the nominal model make the entire set of possible models satisfy the constraint by imposing compactness on $\mathbf{S}$. This would simplify the continuation immensely since it would equate to removing the condition in the

ν-gap definition. The assumption of compactness is not entirely unreasonable either as has been discussed with respect to truncation of a general uncertainty region and by utilizing the fact that the Nyquist plot of a proper rational open loop stable transfer function never will approach infinity. However, the restrictions needed to exclude -1 are likely too sharp to be realistic. To bring back generality, another possibility is to continue to consider general $\mathbf{S}$ and study the question:

> Given a distribution on an arbitrary set of models $\mathbf{S}$ and a nominal model, what is the probability that a random model in the set satisfies the winding number constraint with respect to the nominal model?

Answering this question could, in conjunction with the study of the supremum of the chordal distance, provide complete insight into the distribution of the ν-gap from where results akin to Theorem 3 can easily be posed with respect to b_{PC}. This would yield even more useful results by dispensing with the frequency dependence and study the systems directly.

Another use of the compactness constraint which could be interesting to study in the future is the worst-case analysis it enables, as stated previously. If $\mathbf{S}$ is compact there will be a subset of models in $\mathbf{S}$ for which the distance to the nominal model with respect to the ν-gap or the point-wise chordal distance is maximized. Using this one can consider this subset to perform a worst-case analysis and thereby producing similar, but deterministic, results to Theorem 3 and Corollary 1 thus providing sharp bounds for such cases where performance must be guaranteed.

5.2 Future Outlook

A fresh perspective on probabilistic robust control theory by treating the point-wise chordal distance metric between LTI dynamical systems as a random quantity was presented. Knowing the CDF of the plant model at a frequency, the corresponding CDF of the point-wise chordal distance metric was obtained. The proposed framework presents a fertile research landscape for the future research. A systematic development of future research articles is planned in the following sequence to concretize the proposed theory properly involving both SISO and MIMO extensions.

1. Given a frequency $\omega_1 \in \mathbb{R}_{\geq 0}$ for which the random LTI system $P(j\omega_1)$ has the known distribution $f_P(j\omega_1)$ with compact support $\mathbf{S}(j\omega_1)$, what can you say about $f_P(j\omega_2)$ and $\mathbf{S}(j\omega_2)$, for some $\omega_2 \in \mathbb{R}_{\geq 0}$, by utilizing tools from optimal transport theory such as constraints on the Wasserstein distance $W(f_P(j\omega_1), f_P(j\omega_2))$ in terms of $|\omega_1 - \omega_2|$ or the Wasserstein gradient? Can

such bounds be used to gain knowledge about the distribution of the supremum of the chordal distance when the point-wise distribution is known?

2. Extend the analysis from point-wise chordal distance setting made in this thesis to ν-Gap setting. This involves both the study of the winding number constraint and the distribution of the supremum over frequency of the chordal distance as discussed above. Handling the dependence between frequencies constitutes the main challenge here since the point-wise result given here pays no regard to the covariance between the chordal distance at two different frequencies. The previous point might provide some way to estimate such dependence.

3. Develop probabilistic versions of the robust stability results (see Proposition 1) that is available in [Vinnicombe, 1993]. One such result would be the non-frequency specific counterpart of Theorem 3 and Corollary 1.

4. Investigate how the results of this thesis and the points above can be utilized to determine requirements on $\mathbf{S}$ given pre-determined violation probabilities and performance bounds as discussed above. By considering the motivating example in Section 1.1 this would provide apriori insight into how much data is needed to be collected in order to guarantee the required performance. The practical usefulness of this is promising as it has the potential to reduce both monetary and time costs in data sampling of industrial processes.

5. Generalize the results in this thesis as well as the points above to MIMO LTI systems.

6. Given two random LTI systems P_1, P_2 following respective distributions f_{P_1}, f_{P_2} with compact supports $\mathbf{S}_1, \mathbf{S}_2$ at a frequency $\omega \in \mathbb{R}_{\geq 0}$, find the distribution of the point-wise chordal distance between their projected points onto the Riemann sphere. This would constitute one of the first attempts at computing the probabilistic distance between two stochastic dynamical systems.

7. In MPC [Rawlings et al., 2017], the common way of capturing uncertainty is to add a term to the state equation of the state space model and characterize as a stochastic process or using bounds. These in turn provide a field of potential trajectories from a given initial condition creating a tube around the nominal trajectory which can be used to find an optimal control law taking the worst case scenario into account. It would be very interesting to replace this term with *nu*-gap constraints or distributions f_P on support sets $\mathbf{S}$ as in this thesis to restrict the set of trajectories through a restriction of eligible models. The restriction of models leads to a family of trajectories similar to the tube which can be leveraged similarly.

Bibliography

Åström, K. J. and R. Murray (2021). *Feedback systems: an introduction for scientists and engineers*. Princeton university press.

Başar, T. and P. Bernhard (2008). *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media.

Cannon, M., B. Kouvaritakis, S. V. Raković, and Q. Cheng (2011). "Stochastic tubes in model predictive control with probabilistic constraints". *IEEE Transactions on Automatic Control* **56**:1, pp. 194–200. DOI: 10.1109/TAC.2010.2086553.

Desoer, C. A. and M. Vidyasagar (2009). *Feedback systems: input-output properties*. SIAM.

Georgiou, T. and M. Smith (1990). "Optimal robustness in the gap metric". *IEEE Transactions on Automatic Control* **35**:6, pp. 673–686. DOI: 10.1109/9.53546.

Georgiou, T. T. (1988). "On the computation of the gap metric". *Systems & Control Letters* **11**:4, pp. 253–257.

Hägglund, T. (2021). *Automatic Control, Lecture Notes*. Lund University.

Halder, A. (2014). *Probabilistic methods for model validation*. Texas A&M University.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". *Journal of Basic Engineering* **82**:1, pp. 35–45.

Ljung, L., T. Glad, and A. Hansson (2021). *Modeling and identification of dynamic systems*. Studentlitteratur.

McFarlane, D. and K. Glover (1992). "A loop-shaping design procedure using h/sub infinity/synthesis". *IEEE transactions on automatic control* **37**:6, pp. 759–769.

Megretski, A. and A. Rantzer (1997). "System analysis via integral quadratic constraints". *IEEE transactions on automatic control* **42**:6, pp. 819–830.

Nyström, A. (2024). *Github repository*. URL: https://github.com/antonnystrm/Understanding-Probabilistic-Uncertainty-Using-Nu-Gap.git.

Nyström, A., V. Renganathan, and M. Cantoni (2024). "Stereographic projection of probabilistic frequency-domain uncertainty". In: *Submitted to Conference on Decision and Control, 2024*.

Papoulis, A. and S. U. Pillai (2002). *Probability, random variables, and stochastic processes*. Fourth Edition, McGraw-Hill Europe: New York, NY, USA.

Pates, R. (2021). *The Nyquist stability criterion*. Youtube. URL: https://www.youtube.com/watch?v=XRDqbEVO2w0.

Rawlings, J. B., D. Q. Mayne, M. Diehl, et al. (2017). *Model predictive control: theory, computation, and design*. Vol. 2. Nob Hill Publishing Madison, WI.

Rohr, A. von, M. Neumann-Brosig, and S. Trimpe (2021). "Probabilistic robust linear quadratic regulators with gaussian processes". In: *Learning for Dynamics and Control*. PMLR, pp. 324–335.

Vidyasagar, M. (1984). "The graph metric for unstable plants and robustness estimates for feedback stability". *IEEE Transactions on Automatic Control* **29**:5, pp. 403–418.

Vinnicombe, G. (1992). *Measuring robustness of feedback systems*. PhD thesis. University of Cambridge.

Vinnicombe, G. (1993). "Frequency domain uncertainty and the graph topology". *IEEE Transactions on Automatic Control* **38**:9, pp. 1371–1383.

Vinnicombe, G. (2000). *Uncertainty and Feedback, H Loop-shaping and the V-gap Metric*. World Scientific.

Zames, G. and A. K. El-Sakkari (1980). "Unstable systems and feedback: the gap metric". In: *Proc. of the Allerton Conference, 1980*, pp. 380–385.

Lund University
Department of Automatic Control
Box 118
SE-221 00 Lund Sweden

Document name
MASTER'S THESIS

Date of issue
June 2024

Document Number
TFRT-6230

Author(s)
Anton Nyström

Supervisor
Venkatraman Renganathan, Dept. of Automatic Control, Lund University, Sweden
Anders Rantzer, Dept. of Automatic Control, Lund University, Sweden (examiner)

Title and subtitle
Understanding Probabilistic Uncertainty Using ν-Gap

Abstract

System uncertainty constitutes a fundamental restriction on control performance. System models are never perfect and the differences between system and model can be difficult to contend with. Applying a model based controller to the true system with uncertain dynamics can yield unpredictable results which led researchers to produce methods of robust control design. Existing theory on the ν-gap metric pro-vides control performance guarantees given bounds on the metric. However, it does not utilize any further information than the bounds, essentially restricting the set of possible systems into an uncertainty set for which the guarantees apply.

This thesis aims to investigate how additional information about the uncertain sys-tem can be leveraged to provide sharper results; specifically, by additionally consid-ering a probability distribution function (PDF) on the uncertainty set. Considering the uncertain system as a random quantity with a known distribution models it as more than simply belonging to some uncertainty set. It also incorporates further knowledge as to where in the set it is more likely to be. As such, this thesis opens up an entirely new perspective on the field of probabilistic uncertainty in control systems using the lens of the ν-gap.

Using the additional information, this thesis provides insight into how the differ-ence between a known model and the uncertain system is characterized as well as the potential effect on control performance. Two expressions for the cumulative dis-tribution function (CDF) of one such difference metric called the chordal distance is derived. With knowledge of this distribution, probabilistic guarantee results of a performance measure called the point-wise generalized stability margin are also produced. Some intermediate results which further illuminate the concepts and their relation to each other are also found. Lastly, a thorough discussion is given on how this field of research could be explored to expand the work started in this thesis.

Keywords
Control system uncertainty, Probabilistic uncertainty, Probabilistic control design, Nu-gap, Stereographic Projection

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280-5316

ISBN

Language
English

Number of pages
1-54

Recipient's notes

Security classification

http://www.control.lth.se/publications/