Electrical Engineering and Computer Science Department
Case Western Reserve University

Distributionally Robust Monte Carlo Simulation: A Tutorial Survey

Constantino M. Lagoa and B. Ross Barmish

Technical Report EECS 01-01, September 2001

Distributionally Robust Monte Carlo Simulation: A Tutorial Survey*

Constantino M. Lagoa† and B. R. Barmish‡

Abstract

Whereas the use of traditional Monte Carlo simulation requires probability distributions for the uncertain parameters entering the system, *distributionally robust* Monte Carlo simulation does not. The description of this new approach to Monte Carlo simulation is the focal point of this tutorial survey. According to the new theory, instead of carrying out simulations using some rather arbitrary probability distribution such as Gaussian for the uncertain parameters, we provide a rather different prescription based on *distributional robustness* considerations. The new approach which we describe, does not require a probability distribution f for the uncertain parameters. Instead, motivated by manufacturing considerations, a class of distributions $\mathcal{F}$ is specified and the results of the simulation hold for all $f \in \mathcal{F}$. In a sense, this new method of Monte Carlo simulation was developed with the *robustician* in mind. That is, the motivation for this new approach is derived from the fact that robusticians often object to classical Monte Carlo simulation on the grounds that the probability distribution for the uncertain parameters is unavailable. They typically begin only with bounds on the uncertain parameters and are unwilling to assume an *a priori* probability distribution. This is the same starting point for the methods provided here.

†EE Department, The Pennsylvania State University, University Park, PA 16802.

‡EECS Department, Case Western Reserve University, Cleveland, OH 44106.

*Funding for this research was provided by the National Science Foundation under Grants ECS-9811051 and ECS-9984260.

1. Introduction

When the model a system depends nonlinearly on uncertain parameters, a Monte Carlo analysis is often insightful when mathematical manipulation of the equations would otherwise be prohibitive; e.g., see [1]. The focal point of this tutorial paper are questions of the following sort: For the case when there is little or no statistical description of the random variables entering a system, what Monte Carlo simulation procedure, if any, is appropriate for analysis?

This tutorial survey describes the new approach to Monte Carlo simulation which originates in [2] and [3]. Whereas the use of traditional Monte Carlo simulation software requires probability distributions for the uncertain parameters as input, *distributionally robust* Monte Carlo simulation method of this paper does not. Instead, similar to classical robustness theory, the uncertain parameters are described solely in terms of their bounds with no *a priori* statistics assumed. In this setting, instead of carrying out simulations using some rather arbitrary probability distribution such as Gaussian, we provide a rather different prescription for simulation based *distributional robustness* considerations. More specifically, motivated by manufacturing considerations, we define a class of probability distributions $\mathcal{F}$ and prescribe a method of simulation which leads to conclusions which hold robustly for all $f \in \mathcal{F}$. To this end, the theory characterizes some distinguished distribution $f^* \in \mathcal{F}$ with which the simulation should be carried out. In this sense, our approach is *a posteriori* in nature. That is, instead of assuming a probability distribution *a priori* as in a the classical Monte Carlo setting, the theory determines what distribution to use.

To illustrate the situation above by way of example, we consider a typical circuit analysis problem with uncertain resistors, capacitors and inductors described by the manufacturer only in terms of percentage tolerances about some nominal manufacturing values. In other words, no statistical description for the circuit parameters is assumed. In such a case, if one

wishes to carry out a circuit simulation and simply imposes ad hoc probability distributions on the parameters in order to proceed, it is arguable that the results obtained may be unduly optimistic; e.g., see [47] and [48]. Instead, the distributional robustness approach presented here leads to a probability distribution for Monte Carlo simulation which is an outcome of the analysis rather than assumed *a priori*.

In a sense, this new method of Monte Carlo simulation was developed with the *robustician* in mind. That is, the motivation for this new approach is derived from the fact that robusticians often object to classical Monte Carlo simulation on the grounds that the probability distribution for the uncertain parameters is unavailable. In classical robustness analysis with parametric uncertainty, for example, see [50], one starts only with bounds on the uncertain parameters and no *a priori* probability distribution is assumed. This is the same starting point for the probabilistic method provided here.

This distinction between *a priori* and a posteriori probability distributions is what makes the distributional robustness approach different from many which appear in the systems literature. Be it the Monte Carlo analysis and design methods in papers such as [9], [12]–[21], [23], [24] and [37]– [39], the learning theory approach as in [42] and [43], the simulations based on sample size considerations as in [39] and [40], in each case an *a priori* probability distribution is assumed for simulation purposes. For the plethora of cases for which such information is available, there is no need to consider the methods described in this paper. Finally, is should also be noted that the literature is abound with other approaches to uncertain parameters with even more significant differences in starting assumptions; e.g., see [22], [25] and [38].

1.1 Example: To illustrate the issue addressed in this paper at the most basic of levels, consider the mass-spring-damper system of Figure 1 with applied force $u(t)$, unit mass $M = 1$,

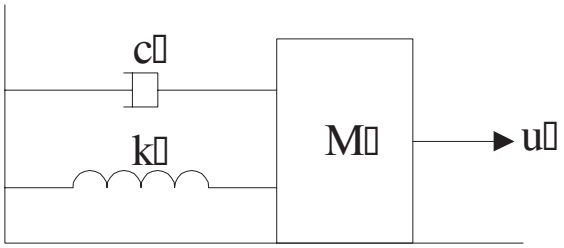

Figure 1: Mass-Spring-Damper System

uncertain spring constant

$$0.2 \leq k \leq 0.8$$

and uncertain damping constant

$$0.3 \leq c \leq 0.9.$$

In view of the parameter uncertainty above, at frequency $\omega \geq 0$, the gain of the system relating displacement for equilibrium to the applied force

$$g(\omega, k, c) = \frac{1}{\sqrt{(\omega^2 + k)^2 + c^2\omega^2}}$$

may vary. In studying such variations, a classical Monte Carlo simulation dictates assignment of probability distributions to the uncertain parameters k and c. Subsequently, one generates samples $k^1, k^2, \ldots, k^N$,$c^1, c^2, \ldots, c^N$ and computes an estimate

$$\hat{g}(\omega) \doteq \frac{1}{N} \sum_{i=1}^{N} g(\omega, k^i, c^i).$$

With regard to the issue under consideration in this paper, the main point to note is that the value of $\hat{g}(\omega)$ obtained via Monte Carlo simulation can change dramatically based on the probability distributions assigned to k and c. To illustrate, at frequency $\omega = 0.01$, if one models highly imprecise manufacturing values for k and c with a uniform distribution, the expected value of the gain is $\hat{g}(0.01) \approx 2.31$. On the other hand, if one postulates a highly precise manufacturing process with normal distribution centered on the intervals for k and c and having standard deviation $\sigma = 0.01$, the result becomes $\hat{g}(0.01) \approx 2.00$.

This significant difference between the two computed gains poses a dilemma for the systems engineer when no *a priori* probability distributions for k and c are given. For example, if one rates the performance of the system using the uniform distribution whereas the "true" distribution is the normal distribution, one obtains an erroneous assessment of performance which is unduly optimistic. To address this problem, the remainder of this paper is devoted to a tutorial exposition of the distributional robustness approach to Monte Carlo simulation. The reader interested in more mathematical detail than that provided here may consult some of the underlying references such as [2], [3], [26], [27], [29] and [30].

2. Preliminaries for Distributional Robustness

In this section, we introduce some of the basic concepts and motivation leading to the distributional robustness formulation to follow. With the mass-spring-damper example above in mind, we entertain one objection to Monte Carlo simulation which the robustician may raise: Namely, in the absence of *a priori* probability distributions for the uncertain parameters q_i, the results of a classical Monte Carlo simulation may be highly suspect.

It turns out that, when working in a distributional robustness framework rather than a classical robustness framework, it is often the case that a larger radius of uncertainty can be tolerated while keeping the risk of performance violation acceptably small. Moreover, when uncertain parameters enter nonlinearly into the system equations, it is often the case that a Monte Carlo approach based on distributional robustness considerations is computationally tractable, whereas a robustness approach is not.

2.1 Uncertainty Notation: We consider a system with *uncertain parameters*

$$q \doteq (q_1, q_2, \ldots, q_\ell) \in \mathbf{R}^\ell$$

and given bounds

$$|q_i| \leq r_i$$

for $i = 1, 2, \ldots, \ell$. Since the variations on q_i are centered at $q_i = 0$, these parameters are viewed as deviations from the so-called nominal. To illustrate, for the mass-spring-damper system of Section 1.1, the spring constant is expressed as $k = 0.5 + q_1$, $|q_1| \leq r_1$, $r_1 = 0.1$ and the damping constant as $c = 0.625 + q_2$, $|q_2| \leq r_2$, $r_2 = 0.125$. With this notation, the *set of admissible uncertainties*

$$Q \doteq \{q : |q_i| \leq r_i \text{ for } i = 1, 2, \ldots, \ell\}$$

is a hypercube in the ℓ-dimensional parameter space.

2.2 Robustician's Point of View: Given a performance specification, call it Property $\mathcal{P}$, for the system under consideration, a typical robustness problem is as follows: Determine if property $\mathcal{P}$ is satisfied for all $q \in Q$. Since this is essentially a worst-case criterion, the robustician recognizes the fact that the assessment of a system from this point of view can be rather conservative. This conservatism provides motivation for the Monte Carlo approach described here and can be linked to the fact that a classical robustness analysis only partially accounts for the shapes of the *good set*

$$Q_{good} \doteq \{q \in Q : \mathcal{P} \text{ is satisfied}\}$$

and the *bad set*

$$Q_{bad} \doteq \{q \in Q : \mathcal{P} \text{ is violated}\}$$

in parameter space.

A metaphor to describe the conservatism associated with classical robustness analysis is provided by Figure 2. In many cases, especially when the dimension of the uncertain parameter vector q is high, the bad set Q_{bad} behaves as if it is a union of "icicles." More specifically,

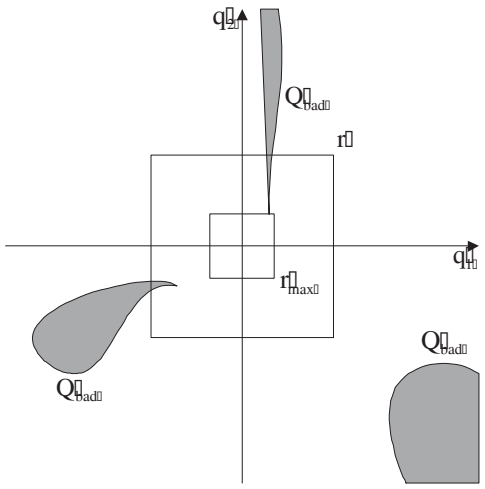

Figure 2: A Two-dimensional Representation of the Geometry of Q_{bad}

over a box of radius r as shown in the figure, the volume of the bad set Q_{bad} is quite small compared to the total volume of Q. For the situation which is depicted, it is noted that a classical robustness analysis leads to a tolerable radius of uncertainty $r = r_{max}$. However, since Q_{bad} has area much less than that of Q, it can be argued that one can work with larger uncertainty radii than r_{max} while keeping the risk of performance violation acceptably small. Hence, one can often justify system operation with uncertainty radius $r > r_{max}$.

This discussion above leads to the following question: Is the so-called icicle geometry of Q_{bad} in Figure 2 just a theoretical possibility or do most physical systems behave in this manner? Simulations based on the approach in this paper indicate that the icicle phenomenon described above is common and that classical approaches tend, in general, to be very conservative — especially when the number of uncertain parameters is high. These statements are substantiated both in the sequel and in the cited references such as [2]–[4] and [26]–[36].

2.3 Motivation for Distributional Robustness: The astute robustician might object to the analysis of r versus r_{max} above on the grounds that a uniform distribution was implicitly assumed for the vector of uncertain parameters q. That is, the comparison of the volumes

of Q_{bad} versus Q does not provide an indication of the risk when the probability distribution of q is unknown. The theory of distributional robustness to follow addresses this concern. Once an appropriate class $\mathcal{F}$ of probability distributions is defined, we only conclude that r can be taken much larger than r_{max} with small risk only if the volume of Q_{bad} is much smaller than the volume of Q under all possible measures obtained with $f \in \mathcal{F}$. In other words, we study robustness with respect to $f \in \mathcal{F}$.

2.4 Problem Formulation: Let $\mathcal{F}$ denote the class of admissible probability distributions for q. Then, for $f \in \mathcal{F}$, we take q^f to be the associated random vector and consider a performance measure $\phi(q)$ of the system in question. For example, $\phi(q)$ might represent the gain of the system at some frequency, rise time to a step input, overshoot to a step input, etc. Equally well, $\phi(q)$ can be of a discrete nature. For example, for a feedback system, we can set $\phi(q) = 1$ if stability is guaranteed with uncertainty q and $\phi(q) = 0$ otherwise. In this setting, we concentrate on two specific probabilistic measures, taking the distribution $f \in \mathcal{F}$ to be a probability density function. The first measure of interest is the probability of satisfying the performance specifications; i.e., for desired performance level $\gamma > 0$, let

$$\begin{aligned}\Phi(f) &= \text{Prob}\{\phi(q^f) \leq \gamma\} \\ &= \int_{\{q \in Q : \phi(q) \leq \gamma\}} f(q)dq.\end{aligned}$$

The second measure is the expected value of $\phi(q^f)$. In this case,

$$\Phi(f) = \mathcal{E}[\phi(q^f)] = \int_Q \phi(q) f(q) dq.$$

With the setup above, the *distributional robustness problem* is to find $f^* \in \mathcal{F}$ minimizing $\Phi(f)$; i.e.,

$$\Phi(f^*) = \min_{f \in \mathcal{F}} \Phi(f),$$

or, equivalently,

$$\Phi(f^*) \leq \Phi(f)$$

for all $f \in \mathcal{F}$.

2.5 Remarks: Upon solving the problem above for $f^* \in \mathcal{F}$ and using this distribution in a Monte Carlo simulation, we obtain more reliable estimates of probability and expected value than would be the case using some ad hoc distribution for q. To illustrate, if stability is of concern, then for any $f \in \mathcal{F}$, it follows that

$$\text{Prob}\{\text{stability under } q^f\} \geq \text{Prob}\{\text{stability under } q^{f^*}\}.$$

Hence, a Monte Carlo simulation performed with some ad hoc distribution $f \in \mathcal{F}$ instead of f^* leads to an unduly optimistic estimate of performance. From a robustician's point of view, it is also of interest to determine the extent to which the *worst-case* performance

$$\phi^* \doteq \min_{q \in Q} \phi(q)$$

differs from the expected performance. To this end, the basic inequality

$$\min_{q \in Q} \phi(q) \leq \mathcal{E}[\phi(q^{f^*})]$$

can be used to understand the icicle metaphor described in Section 2.2.

The desirability of distributional robustness is seen via a simple illustration: Suppose one is assessing the probability that a performance specification is met and a distributionally robust probability estimate $\hat{p} = 0.99$ is obtained. Then, this probability is guaranteed no matter which probability density function $f \in \mathcal{F}$ is realized. Hence, without the knowledge of the "true" probability distribution, one can nevertheless be confident about the assessment of performance. This provides a rationale for a new approach to Monte Carlo simulation for cases when little or no *a priori* statistical information about the uncertain parameters is

available. Namely, in contrast to classical Monte Carlo methods, which require specification of a probability distribution f for q *a priori*, we solve a distributional robustness problem and select a distinguished distribution $f^* \in \mathcal{F}$. This a posteriori distribution is used in the random number generation for the associated Monte Carlo simulation. By proceeding in this manner, one avoids the need to specify some ad hoc distribution when no statistical information about the uncertainty is available.

3. The Class of Distributions $\mathcal{F}$

In this section, attention is turned to the class of probability distributions $\mathcal{F}$; to this point in the paper, this class has not been specified. The paradigm of [2] is now described and it is argued that the definition of $\mathcal{F}$ is physically meaningful for a large class of problems. In later sections, it is seen that this definition of $\mathcal{F}$ leads to a rich theory characterizing the distinguished distribution $f^* \in \mathcal{F}$ which is used for Monte Carlo simulation. That is, in order to carry out computer simulations, the computer program generating random numbers must be "told" what probability distribution $f \in \mathcal{F}$ to use.

Based on robustness considerations in the systems sciences, an interval bound description of the uncertainty is the takeoff point for the new paradigm. Motivated in large measure by manufacturing considerations, the fundamental assumptions in the exposition to follow are that the uncertain parameters are *independent, large deviations in the parameters q_i away from their nominal values is less probable than small deviations* and *positive and negative deviations in the q_i are equally likely.* In other words, no assumptions made about the probability distribution other than its salient characteristics above. In Section 3.2, after making these notions precise, the class $\mathcal{F}$ emerges. This setup is reminiscent of formulations such as Huber's [10] in the field of robust statistics. In contrast to his formulation and

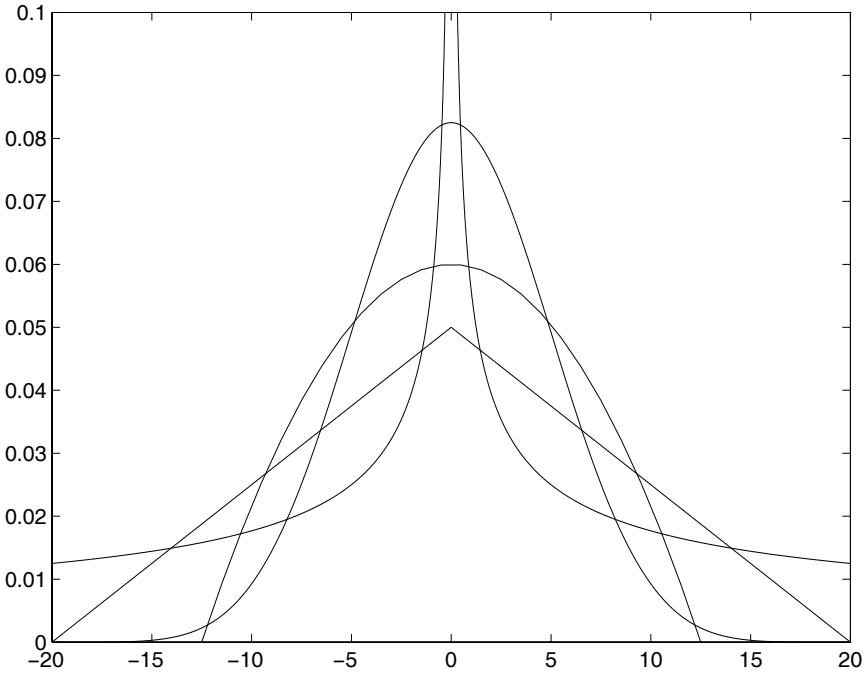

Figure 3: Admissible Distributions for Capacitor Uncertainty

others, however, no *a priori* parameterization of the underlying probability density functions is assumed. This is explained in more detail below.

3.1 Motivating Example: To motivate the definition of $\mathcal{F}$, consider a circuit with an uncertain capacitor 30 μfd $\leq C \leq$ 70 μfd which is nominally manufactured with *nominal value* $C_0 = 50\mu$fd. For this capacitor, the manufacturing process is modelled by assuming that positive and negative deviations about C_0 are equally likely and that large deviations from C_0 are less likely than small deviations. In other words, if $|\Delta C_1| < |\Delta C_2|$, then the capacitor with value $C = 50 + \Delta C_1$ is more likely to be manufactured than the resistor with value $C = 50 + \Delta C_2$. This situation is illustrated in Figure 3 where the possible probability density functions for capacitor uncertainty ΔC are depicted with zero mean. These ideas are now precisely formulated in the more general setting of this paper.

3.2 Class of Admissible Distributions $\mathcal{F}$: It is assumed that the uncertainty vector q is a zero mean random vector with independent components q_i. Furthermore, for $i = 1, 2, \ldots, \ell$, it is assumed that each component q_i is supported in the interval

$$Q_i \doteq [-r_i, r_i].$$

Therefore, the support for the random vector q is the hypercube

$$Q = Q_1 \times Q_2 \times \cdots \times Q_\ell.$$

Now, a density function $f_i(x_i)$ is said to be *admissible* for q_i if it is symmetric and non-increasing with respect to $|x_i|$. More precisely, f_i is an *admissible probability density function* for q_i if

$$f_i(x_i) \geq f_i(y_i),$$

for $|x_i| \leq |y_i|$ and

$$f_i(x_i) = f_i(-x_i)$$

for all x_i.To make the definition of $\mathcal{F}$ complete, the behavior of $f_i(x_i)$ at $x_i = 0$ needs to be specified. In this paper, $f_i(x_i)$ is allowed to be a probability density function which contains a Dirac delta function at $x_i = 0$. Finally, by writing $f \in \mathcal{F}$ for the joint density function

$$f(x) \doteq f(x_1, x_2, \ldots, x_\ell) = f_1(x_1) f_2(x_2) \cdots f_\ell(x_\ell)$$

of the random vector q^f, the understanding is that each f_i is an admissible probability density function for q_i.

3.3 Distributionally Robust Performance: As indicated in Section 2.4, each admissible density function $f \in \mathcal{F}$ results in a value $\Phi(f)$ for system performance. Now, we define the *distributionally robust cost*

$$\Phi^* \doteq \inf_{f \in \mathcal{F}} \Phi(f).$$

Since it has not yet been guaranteed that the infimum above is attained for some $f^* \in \mathcal{F}$, we use the "inf" notation for the definition above. This "inf" will later be sharpened to "min" in the results to follow.

3.4 Summary: The fundamental difference between "ordinary" Monte Carlo theory and distributionally robust Monte Carlo theory is as follows: In the ordinary Monte Carlo problem, the probability density function f used in the simulation is specified *a priori*. In the distributionally robust Monte Carlo problem, f is determined a posteriori; that is, one solves a variational problem to obtain $f^* \in \mathcal{F}$ minimizing $\Phi(f)$.

3.5 Preview Example: To consolidate the development to date and preview the exposition to follow, an example from the theory of robust stability, for example, see [50], is provided to demonstrate some of the basic ideas. To this end, we consider the theory of interval polynomials based on Kharitonov's Theorem [51] within the probabilistic setting of this paper. For the uncertain polynomial

$$p(s,q) = p_0(s) + \sum_{i=1}^{14} s^{i-1} q_i$$

with interval bounds $q_i \in [-r, r]$ for $i = 1, 2, \ldots, 12$ and stable nominal

$$p_0(s) = (s+1)^{12}(s^2 + 0.002s + 1),$$

with lightly damped roots $s = -0.001 \pm j$ and good set

$$Q_{good} = \{q \in Q : p(s,q) \text{ is stable}\},$$

we compare a classical Monte Carlo solution of the stability problem with the robust solution. Whereas ordinary Monte Carlo is used here, in Section 6.6, this same problem is revisited from the distributional robustness point of view.

First, using Kharitonov's Theorem [51], robust stability for $p(s,q)$ is guaranteed if

$$r < 0.021.$$

Now, to illustrate an ordinary Monte Carlo solution, let $r = 0.03$. Noting that this bound is approximately 40% above the stability limit provided by Kharitonov's Theorem, the objective is to estimate the risk of instability and the number of samples required to achieve a prescribed degree of confidence in the result.

In accordance with the notation of the preceding section, for a given probability density function f for q and

$$\Phi(f) = \text{Prob}\{p(s, q^f) \text{ is stable}\},$$

the basics of ordinary Monte Carlo simulation are illustrated with $f = u$ being the uniform distribution. Taking $\phi(q) = 1$ if $p(s, q)$ is stable and $\phi(q) = 0$ otherwise, an ordinary Monte Carlo simulation involves randomly generating N samples $q^1, q^2, \ldots, q^N$ for q and creating the relative frequency estimate for stability

$$\widehat{\Phi}(u) = \frac{1}{N} \sum_{i=1}^{N} \phi(q^i).$$

For the moment, a sample size $N = 10^5$ is arbitrarily specified while noting that the choice of N is explained in the next section. In Figure 4, a convergence plot of the partial estimates

$$\widehat{\Phi}_k(u) \doteq \frac{1}{k} \sum_{i=1}^{k} \phi(q^i)$$

is given. This leads to the estimate

$$\text{Prob}\{p(s, q^u) \text{ is stable}\} \approx 0.99951$$

In other words, with uncertainty bound approximately 40% above Kharitonov's limit, only a small risk of instability is obtained.

To conclude this section, it is important to remind the reader that the probability density function f for q was assumed *a priori*. Therefore, the computed probability is simply an ordinary Monte Carlo estimate rather than a distributionally robust estimate. In Section 6.6, this example is revisited from the distributional robustness point of view.

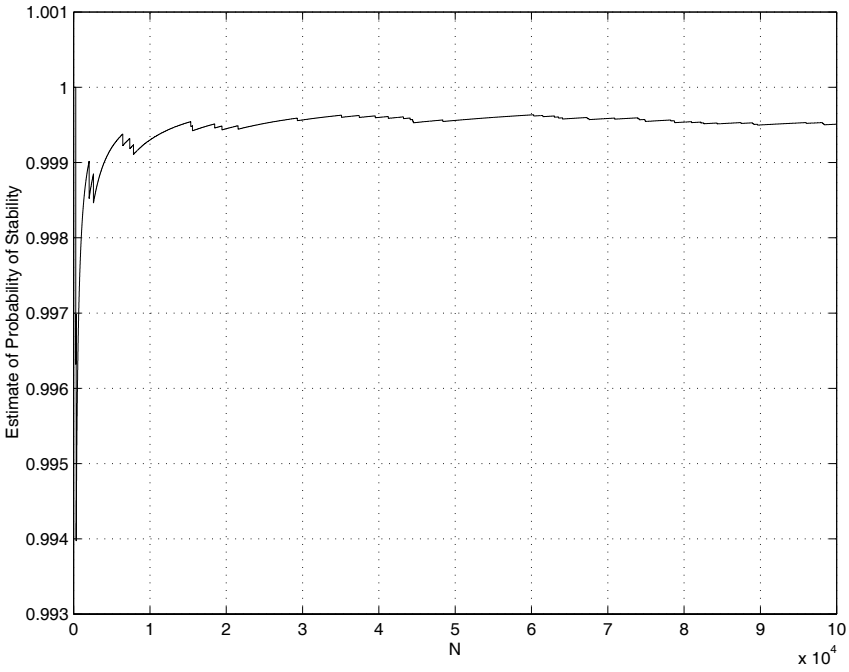

Figure 4: Estimated Probability of Stability Versus Sample Size N

3.6 Sample Generation: Although not the focal point of this paper, it is important to provide an indication of how one generates the random samples and how one chooses the sample size. Since these problems are important in both the classical and the distributionally robust Monte Carlo settings, some highlights in the sample generation and sample size literature are now described.

Currently, many software packages contain routines for generating random samples with commonly used distributions. For example, the basic distribution of Matlab contains routines for generating samples with either uniform or Gaussian distributions. However, when dealing with robustness problems, one is confronted with the necessity of generating samples with distributions other than the ones mentioned above. For example, when using Monte Carlo methods for analyzing systems with unstructured uncertainty, one might have uncertainty that is uniformly distributed on a sphere. For cases like this, one needs more powerful tools for sample generation; see [5] for a comprehensive treatment.

Of special interest to robusticians, is the problem of generating samples uniformly distributed over a given compact convex set. This problem has been addressed by several authors; we mention two approaches which have been proposed. The first approach involves the design of a Markov chain whose stationary distribution is the required one. Examples of this approach can be found in [5] and [7], where algorithms have been proposed for very general compact convex sets. For cases when one has a lot of information about the convex set in which we want the samples to be generated, one can develop more direct sample generation algorithms which do not rely on the asymptotic behavior of a Markov chain. For example, this is the case addressed in [8] and [9]. where algorithms are provided for uniform sampling over ℓ^p ball described by

$$q_1^p + q_2^p + \cdots + q_\ell^p \leq 1$$

and over the "sphere" of matrices

$$\mathcal{M} \doteq \{A \in \mathbf{R}^{n\times n} : \overline{\sigma}(A) \leq 1\}$$

where $\overline{\sigma}(A)$ denotes the maximum singular value of the matrix A.

3.7 Sample Size: In the literature on sample generation, the following question arises: For a given uncertainty dimension ℓ and a given probability density function f for q, how many samples N are required to obtain a "reliable" estimate $\widehat{\Phi}_N(f)$? Surprisingly, with reliability defined in terms of *probable approximate correctness* (PAC) as indicated below, it can be shown that there are upper bounds for the required number of samples which are independent of both ℓ and f. To illustrate the use of such results, following [42], the PAC reliability criterion is defined and illustrated using sample size bounds provided in [39] and [40].

3.8 Reliability Based on Probable Approximate Correctness: In this framework, the estimate $\widehat{\Phi}_N(f)$ is viewed as random variable and one seeks to find probability of this

quantity being in error by no more than a prescribed tolerance $\varepsilon > 0$. With this setup, an estimate $\widehat{\Phi}_N(f)$ is said to have *reliability* of $0 < \delta < 1$ if

$$\text{Prob}\{|\widehat{\Phi}_N(f) - \Phi(f)| > \varepsilon\} \leq \delta.$$

In other words, the probability of an estimation error exceeding ε is less than or equal to δ.

With the definition above, there are many papers with upper bounds on the number of samples $N = N(\varepsilon, \delta)$ which are needed; e.g., see [39] and [40]. To illustrate, a simple upper bound based on the Law of Large Numbers is

$$N(\varepsilon, \delta) = \frac{1}{4\varepsilon^2\delta}.$$

A second upper bound, obtained using the Bienaymé inequality is

$$N(\varepsilon, \delta) = \frac{\sqrt{3}}{4\epsilon^2\sqrt{\delta}}.$$

A third upper bound, obtained using the Chernoff inequality is

$$N(\varepsilon, \delta) = \frac{\ln(2/\delta)}{2\epsilon^2}.$$

None of the bounds above is "best" in the sense of requiring less samples than the others for all (ε, δ) pairs. Therefore, for a given ε and δ, one can look at all available bounds and take the smallest of the $N(\varepsilon, \delta)$ values obtained. To illustrate, using the three bounds above with $\varepsilon = 0.01$ and $\delta = 0.06$, the Bienaymé bound is the tightest and leads to $N \approx 1.77 \times 10^4$. For the tighter specification $\varepsilon = \delta = 0.001$ corresponding to a 0.1% error, one obtains $N \approx 3.8 \times 10^6$ using the Chernoff bound.

For the case of the interval polynomial in Section 3.5 above, $\varepsilon = \delta = 0.005$ was taken as the reliability specification and the number of samples dictated is $N = 1.2 \times 10^5$. However, as a practical matter, it is seen in Figure 4 that with $N = 3 \times 10^4$, convergence is obtained with far fewer samples than that prescribed by the theory. This is consistent with the authors' experience involving Many Monte Carlo case studies of this sort.

4. The Truncation Principle

The Truncation Principle of [2] is a fundamental result in the theory of distributional robustness and serves as the takeoff point for the Monte Carlo simulation techniques in the sequel. This principle indicates that when minimizing the performance functional $\Phi(f)$ over $f \in \mathcal{F}$, one need only consider truncated uniform distributions. In other words, it turns out to be the case that distributional robustness is achieved by restricting attention to probability distributions which are constant over some subinterval $[-t_i, t_i]$ of $[-r_i, r_i]$. For example, even though a truncated normal distribution is admissible, it can be ignored when minimizing $\Phi(f)$ over $f \in \mathcal{F}$. This means that the prescription for distributionally robust simulation involves sampling only a subinterval of the uncertainty rather than the entire interval.

It is interesting that the subinterval sampling scheme above may seem counterintuitive in certain applications. For example, for Monte Carlo simulation of a system with spring constant parameter $0.4 \leq k \leq 0.6$, a distributionally robust simulation may require sampling restricted to $0.45 \leq k \leq 0.55$. These ideas are now formalized.

4.1 Truncated Uniform Distributions: A probability density function $u^t(x)$ is called a *truncated uniform distribution* if each of its components $u_i^t(x_i)$ is either distributed uniformly over a symmetric interval $[-t_i, t_i] \subseteq [-r_i, r_i]$ for $t_i > 0$ or zero with probability one for $t_i = 0$; that is, a Dirac delta function. The interval $[-t_i, t_i]$ might be different for each uncertainty component. Using the notation

$$T \doteq \{t = (t_1, t_2, \ldots, t_\ell) \in \mathbf{R}^\ell : 0 \leq t_i \leq r_i \text{ for } i = 1, 2, \ldots, \ell\},$$

for $t \in T$, we take

$$u^t(x) \doteq u_1^t(x_1) u_2^t(x_2) \cdots u_\ell^t(x_\ell)$$

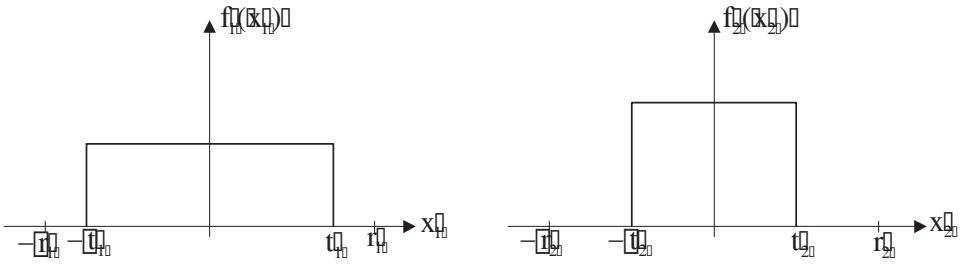

Figure 5: Truncated Uniform Distribution

to be the associated truncated uniform distribution. For the special case obtained with $t_i = r_i$ for $i = 1, 2, \ldots, \ell$, one obtains the uniform distribution $u(x)$ over Q. It is also observed that all truncated uniform distributions u^t belong to the class $\mathcal{F}$; i.e., the inclusion

$$\{u^t : t \in T\} \subseteq \mathcal{F}.$$

holds. An example of a truncated uniform distribution for $\ell = 2$ is presented in Figure 5.

The Truncation Principle of [2] indicates that, in the search for the minimum of $\Phi(f)$, one need not to consider all possible distributions $f \in \mathcal{F}$; that is attention can be restricted to the class truncated uniform distributions u^t obtained with $t \in T$.

4.2 The Truncation Principle: *With the notation above,*

$$\inf_{f \in \mathcal{F}} \Phi(f) = \inf_{t \in T} \Phi(u^t).$$

4.3 Distributional Robustness: The theorem above provides a prescription for distributionally robust Monte Carlo simulation for many cases when some *optimal truncation* $t^* \in T$ minimizing $\Phi(u^t)$ can be found. Namely, one simply performs the simulation using uniform sampling over the interval $[-t_i^*, t_i^*]$ in lieu of $[-r_i, r_i]$.

4.4 Example Illustrating Truncation Principle: The Truncation Principle raises the possibility that distributionally robust Monte Carlo simulation may lead to results which differ significantly from what one might obtain using a more traditional Monte Carlo approach.

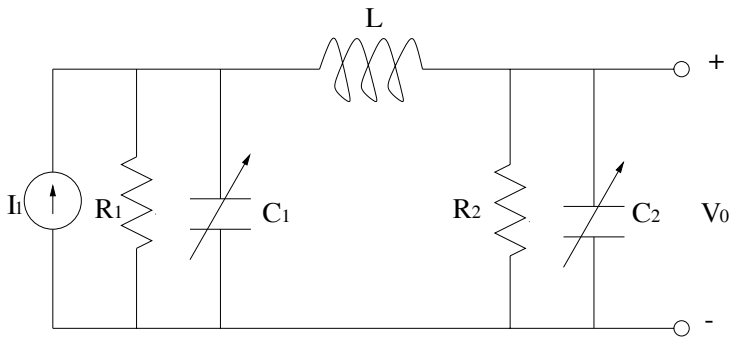

Figure 6: RLC Circuit

That is, in the example below, taken from [30], the Truncation Principle leads to sampling over a subinterval of the range of q_i whereas a classical Monte Carlo analysis typically dictates sampling over the entire range of parameter variation. Subsequently, the two methods may lead to dramatically different assessments of performance. In this regard, the point of view in this paper is that traditional Monte Carlo simulation provides an unduly optimistic estimate of the performance whereas the distributionally robust approach does not.

The RLC circuit of [30] is now studied with random parameters corresponding to independent uncertainties in the interstage capacitors C_1 and C_2; see Figure 6. The amplifier has fixed parameters $R_1 = 1000, R_2 = 100, L = 0.01$ and uncertain parameters

$$0.755 \times 10^{-6} \leq C_1 \leq 1.695 \times 10^{-6}; \qquad 0.75 \times 10^{-6} \leq C_2 \leq 4.55 \times 10^{-6}.$$

For this example, performance is defined in terms of the overshoot to a step input. The specification is that $|V_0(t)|$ not exceed 96.3 volts. This leads to an interest in computing the probability that this performance specification is satisfied.

To study this circuit using the Truncation Principle, q_1 and q_2 are identified with deviations from the center points of the intervals of capacitance. Next, letting $V_0(q_1, q_2, t)$ denote the dependence of the output voltage on the q_i and taking

$$Q \doteq \{(q_1, q_2) : |q_1| \leq r_1 = 0.940 \times 10^{-6}; \;\; |q_2| \leq r_2 = 4.8 \times 10^{-6}\}$$

and

$$Q_{good} \doteq \{(q_1, q_2) : |V_0(q_1, q_2, t)| \leq 96.3 \text{ for all } t \geq 0\},$$

we seek to compute the distributionally robust performance

$$\Phi^* \doteq \min_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good}\}.$$

In accordance with the Truncation Principle, a solution to this problem is obtained with marginals u_i^t which are truncated uniform distributions described by t_i . For the given uncertainty bounds for the two capacitors variations, a two variable optimization in the truncation variable $t \doteq (t_1, t_2) \in T$ was carried out. Using the Matlab rand function to estimate

$$p_t \doteq \text{Prob}\{q^{u^t} \in Q_{good}\}$$

and generating 100,000 samples for (t_1, t_2) pairs, the estimate

$$\Phi^* = \Phi(u^{t^*}) = p_{t^*} \approx 0.486$$

was reached with truncations given by $t_1 = t_1^* \approx 0.17 \times 10^{-6}$ and $t_2 = t_2^* \approx 0.275 \times 10^{-6}$. This result is shown in the contour plot in Figure 7. It is noted that the truncation t^* maximizing p^t is obtained as an interior point within the rectangle of capacitor variation.

In order to compare the result above with a traditional Monte Carlo simulation, we take $f = u$ to be the uniform distribution and obtain the estimate

$$\Phi(u) = \text{Prob}\{q^u \in Q_{good}\} \approx 0.6912,$$

which is more than 50% larger than Φ^*.

4.5 Remarks: The successful use of the Truncation Principle to solve the circuit problem above was facilitated by the fact that the uncertain parameter vector q was only two-dimensional. For problems with higher dimensional uncertainty, finding an "optimal truncation" t^* is generally a nonlinear programming problem. Whereas a gridding method sufficed

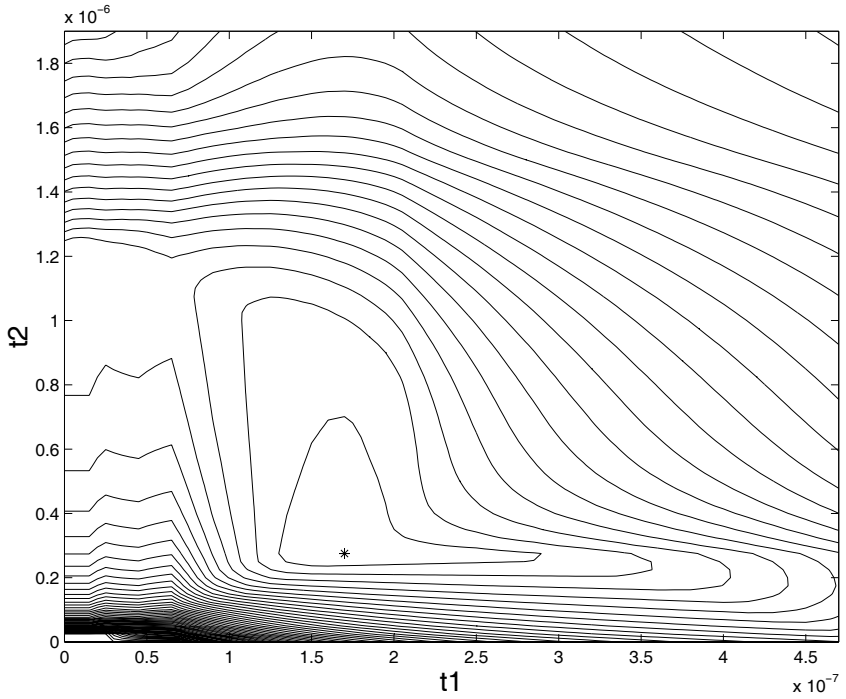

Figure 7: Contours of Equal Probability in the (t_1, t_2) Plane

for the circuit illustration above, for a high number of uncertainties, such an approach is no longer computationally tractable. This motivates an ongoing line of research aimed at exploiting the structure of the probabilistic robustness problem at hand in order to obtain an optimal truncation $t^* \in T$. For example, as seen in Sections 5.1–5.7, for large classes of circuit configurations, the equations associated the Kirchoff's laws for circuits make it possible to characterize the optimal truncation t^* without recourse to nonlinear programming. A second case where one can actually determine t^* occurs when the good set Q_{good} is convex and symmetric. As seen in Section 5.9, in this case, the uniform distribution u turns out to be the optimal truncation; i.e., we set each t_i to its maximum value r_i. In other words, in this special case it can be argued that the traditional Monte Carlo approach leads to a distributionally robust result. Finally, it is also worth mentioning the case when Q_{good} is convex but not necessarily symmetric, the concept of symmetrization proves useful to obtain a bound on the performance function $\Phi(f)$ using the uniform distribution; see Section 6.1.

5. The Convex and Componentwise Convex Cases

In the absence of an *a priori* statistical description of the uncertain parameters, it has been argued that one may obtain unduly optimistic results using traditional Monte Carlo simulation. In this section, we provide conditions under which the traditional approach agrees with the distributionally robust approach. In other words, under the conditions described below, there is indeed a solid rationale for use of the folk theorem which says: *When in doubt, use the uniform distribution.* To this end, of the results of [2]–[4] are now described.

The first result below applies to the case when $\Phi(f)$ is the expected value of some *componentwise convex* performance function; i.e., for each $i = 1, 2, \ldots, \ell$, the function $\phi_i(q_i)$ obtained with q_k held fix for $k \neq i$, is convex in q_i. To illustrate, for large classes of robustness problems with a so-called *multilinear* uncertainty structure, this componentwise convexity condition is satisfied; e.g., one can obtain a performance function of the form

$$\phi(q) = 3q_1q_2q_3 + 10q_1q_2 - 9q_1q_3 - q_2 + 15.$$

Two examples are provided, which illustrate the application of the Componentwise Convexity Principle to resistive networks and to H_∞ performance. The second result, the Uniformity Principle in Section 5.9, applies to the case when $\Phi(f)$ is the probability of performance satisfaction and the set Q_{good} is convex and symmetric (if $q \in Q_{good}$, then $-q \in Q_{good}$). Associated with each of these results, examples are given which illustrate the satisfaction of the required conditions on $\Phi(f)$ and Q_{good}

5.1 The Componentwise Convexity Principle: *If $\phi(q)$ is convex with respect to component q_k, then the minimization of $\mathcal{E}(\phi(q^f))$ is attained with $f^* \in \mathcal{F}$ having k-th component $f_k^* = \delta$, the Dirac delta function. Similarly, if $\phi(q)$ is concave with respect to q_k then, the minimization of $\mathcal{E}(\phi(q^f))$ is attained with $f^* \in \mathcal{F}$ having k-th component $f_k^* = u$, the*

uniform distribution.

5.2 Remark: The result above provides conditions under which a minimizing truncation t_k^* in Theorem 4.2 is attained at an extreme $t_k^* = 0$ or $t_k^* = r_k$.

5.3 Resistive Networks: For a large class of resistive networks described below, it is seen that the extreme point solutions obtained above solve the distributional robustness problem for the expected gain. Such a result is considerably at odds with what one obtains using Monte Carlo sampling scheme or common sense traditional considerations; i.e, in such a case, the prescription of the theorem is that some uncertain parameters, those with the Dirac Delta function distribution, should be held fixed at their nominal values whereas other uncertain parameters should be sampled uniformly over their range of variation. In other words, one should resist the temptation to sample those uncertain parameters q_k corresponding to $f_k^* = \delta$ in the Componentwise Convexity Principle. To illustrate, if $t_1 = 0$ for q_1 and $t_2 = r_2$ for q_2, then to obtain a distributionally robust performance estimate, one should ignore the temptation to sample q_2 uniformly over its range $[-r_2, r_2]$. Instead, one should generate samples (q_1, q_2) with q_1 fixed at its nominal value and q_2 uniformly distributed over $[-r_2, r_2]$.

The situation above is more fully described in [34] where the authors consider a planar network $\mathcal{N}$ consisting of an input voltage source V_{in}, an output voltage V_{out} across a designated resistor $R_{out} = R_n$ and uncertain positive resistor n-tuple $R \doteq (R_1, R_2, \ldots, R_n)$; see Figure 8. With q_i identified with resistor uncertainty ΔR_i representing deviations from the nominal manufacturing value $R_{i,0} > 0$ and gain

$$g(q) \doteq \frac{V_{out}(q)}{V_{in}},$$

the Truncation Principle applies to the problem of finding the maximum and minimum values

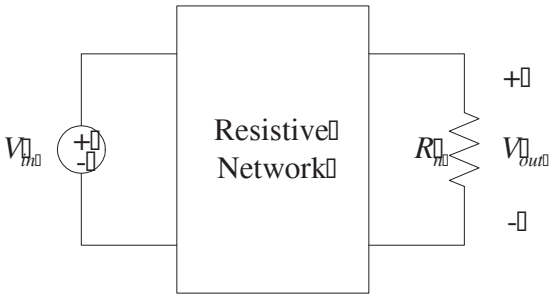

Figure 8: Network Configuration

of the expected gain

$$\mathcal{E}(g(q^f)) = \int_Q f(q)g(q)dq.$$

The definition below, given in [34], leads to the delineation of a class of circuits for which the optimal truncations for Monte Carlo admit extreme point solution per Section 5.1.

5.4 Essential Resistors: For the class of resistive networks under consideration, physical interpretations of componentwise convexity and concavity are available. Namely, a resistor R_k is said to be *essential* if the following condition holds: There does not exist admissible values of the $n-1$ remaining resistors R_i, $i \neq k$ making the gain g independent of R_k. If R_k is essential, it can readily be shown that, with $q_k = \Delta R_k$ as identified above, the gain is either componentwise convex or concave with respect to q_k. To make the convexity/concavity assignment more precise, it is noted that essentiality guarantees that the partial derivative $\partial g/\partial q_k$ has one sign over Q. Letting

$$s_k \doteq \text{sign}\left(\frac{\partial g}{\partial q_k}\right)$$

denote this invariant sign, exploitation of the Componentwise Convexity Principle leads to the result of [34] given below.

5.5 Theorem: *Assume that all resistors in $\mathcal{N}$ are essential. For the case of maximizing $\mathcal{E}(g(q^f))$, define probability density function f^* with marginals f_i^* as follows: Set $f_i^* = u$*

if $s_i = -1$ *and* $f_i^* = \delta$ *if* $s_i = 1$. *Then,*

$$\mathcal{E}(g(q^{f^*})) = \max_{f \in \mathcal{F}} \mathcal{E}(g(q^f)).$$

For the case of minimizing $\mathcal{E}(g(q^f))$, *define probability density function* f^* *with marginals* f_i^* *as follows: Set* $f_i^* = \delta$ *if* $s_i = -1$ *and* $f_i^* = u$ *if* $s_i = 1$. *Then,*

$$\mathcal{E}(g(q^{f^*})) = \min_{f \in \mathcal{F}} \mathcal{E}(g(q^f)).$$

5.6 Example: To illustrate the use of the theorem above, consider the ladder network studied in [34] and shown in Figure 9. Applying the theorem above, it can be shown that all resistors are essential with maximum expected gain being attained by using $t_i = 0$ for the inter-stage resistors R_{3k} and $t_i = r_i$ for the remaining resistors. To illustrate how this result is applied, for a three stage network with nominal values $R_{1,0} = R_{4,0} = R_{5,0} = R_{7,0} = R_{8,0} = 1$, $R_{2,0} = 2$, $R_{3,0} = 3$, $R_{6,0} = 5$ and $R_{9,0} = 7$, and uncertainty bounds $r_i = 0.8R_{i,0}$ for the inter-stage resistors and $r_i = 0.1R_{i,0}$ for the remaining resistors, the results above indicate that a distributionally robust Monte Carlo simulation should be performed as follows: Hold the interstage resistors R_3, R_6 and R_9 fixed corresponding to the Dirac Delta function; i.e., do not sample these parameters despite the fact that sampling ranges are given. For the remaining resistors, sample uniformly over prescribed ranges $[R_{i,0} - r_i, R_{i,0} + r_i]$. This sampling scheme leads to the estimate

$$\mathcal{E}(g(q^{f^*})) \approx 0.1864$$

with $n = 100,000$ samples. In contrast, a more traditional Monte Carlo simulation using the uniform distribution for all resistors leads to the estimate

$$\mathcal{E}(g(q^u)) \approx 0.1554.$$

was obtained. In other words, the classical analysis leads to a result which we view as over optimistic by about 20%.

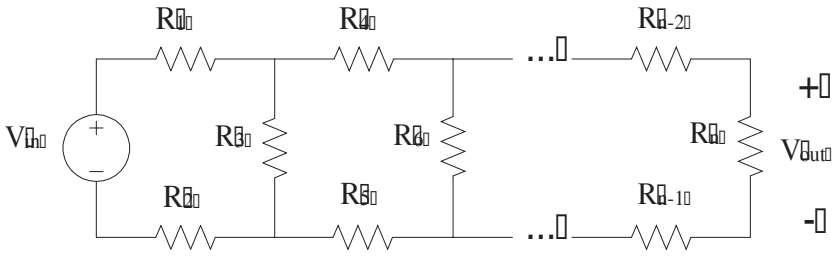

Figure 9: Ladder Network

5.7 Multilinearly Parameterized H_∞ Norm: To illustrate a second application of the Componentwise Convexity Principle, we consider the problem formulation of [36]. Namely, the starting point is a transfer function matrix $H(s,q)$ which depends multilinearly on the uncertain vector q. Furthermore, it is assumed that $H(s,q)$ is decomposable as a multilinear combination fixed stable transfer functions with the uncertain parameters q_i being the matrix multipliers. For example, if $H_0(s), H_1(s), H_2(s)$ and $H_{12}(s)$ are proper stable $n \times m$ transfer function matrices, then

$$H(s,q) = H_0(s) + q_1 H_1(s) + q_2 H_2(s) + q_1 q_2 H_{12}(s)$$

is such a multilinear combination. Another example is obtained from a feedback system which is set up in the so-called $M - \Delta$ configuration with $M(s)$ being a square $\ell \times \ell$ proper stable transfer function matrix and $\Delta(q) = \text{diag}\{q_1, q_2, \ldots, q_\ell\}$. Now,

$$H(s,q) \doteq \det(I + M(s)\Delta(q))$$

satisfies the multilinearity requirement of this section.

Using the fact that the norm function is convex and each q_i enters affine linearly into $H(s,q)$ with the remaining parameters fixed, it can readily be shown that that with performance measure

$$\phi(f) = ||H(s,q^f))||_\infty,$$

the uncertain parameter vector q enters in a componentwise convex manner. Hence,

$$\max_{f\in\mathcal{F}} \mathcal{E}\left(||H(s,q^f)||_\infty\right) = \mathcal{E}\left(||H(s,q^u)||_\infty\right).$$

In fact, it can also be shown (see [36]) that the same result hold for all moments of $||H(s,q)||_\infty$ as well.

5.8 Convex Symmetric Sets: Attention now is turned to the case where $\Phi(f)$ is the probability of performance satisfaction. As seen below, if the set of parameters that satisfy the performance specifications is convex and symmetric then the the uniform distribution is the one that should be used in the Distributional Robustness setting. We now formally present the result of [2] which initiated the distributional robustness line of research.

5.9 The Uniformity Principle: *If Q_{good} is convex and symmetric, then it follows that*

$$\min_{f\in\mathcal{F}} \text{Prob}\{q^f \in Q_{good}\} = \text{Prob}\{q^u \in Q_{good}\}.$$

Equivalently, the minimizing truncation t^* in Theorem 4.2 has components $t_i^* = r_i$ corresponding to the uniform distribution.

5.10 Example (Interval Polynomial): The interval polynomial, analyzed in Section 3.5 from a traditional Monte Carlo point of view is now studied using the Uniformity Principle It is shown that a distributionally robust under-estimate of the probability of stability can be obtained using the Uniformity Principle. Indeed, recalling $p(s,q) = p_0(s) + \sum_{i=1}^{12} s^{i-1}q_i$ with interval bounds $q_i \in [-r,r]$ for $i = 1,2,\ldots,12$, uncertainty radius $r = 0.03$ and stable nominal $p_0(s) = (s+1)^{12}(s^2+0.002s+1)$, in lieu of defining Q_{good} in terms of stability, we generate this set based on frequency domain considerations. Namely, with target set $\mathcal{P}(\omega)$ given in Figure 10, for robust stability, classical robustness theory (for example, see [50]) can be used to show that with a fixed $q \in Q$, stability of $p(s,q)$ is assured if

$$p(j\omega,q) \in \mathcal{P}(\omega)$$

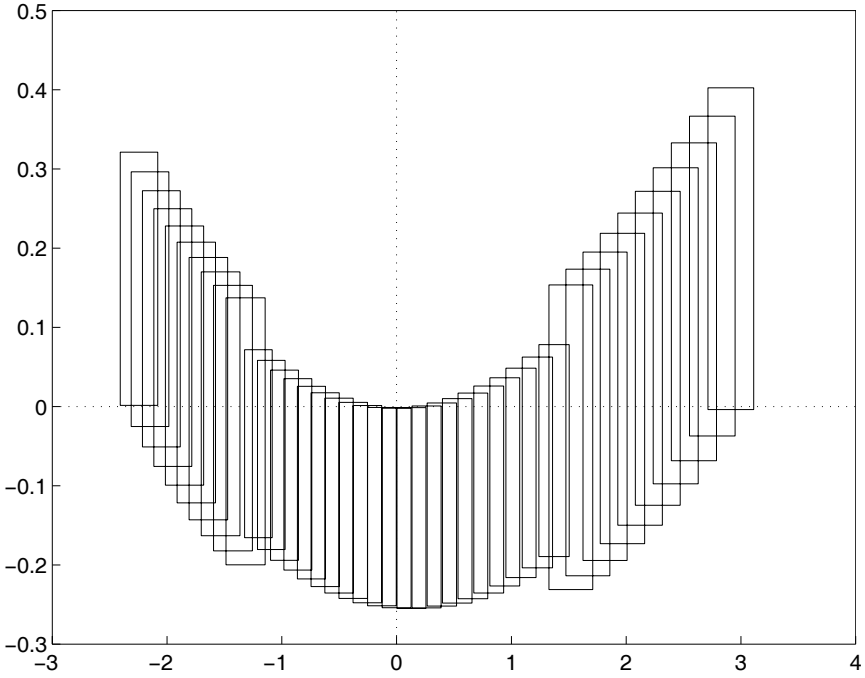

Figure 10: The Target Set $\mathcal{P}(\omega)$ for $0.98 \leq \omega \leq 1.02$.

for the critical range of frequencies $0.98 \leq \omega \leq 1.02$.

Now, to obtain the desired distributionally robust Monte Carlo estimate, we take

$$Q_{good} \doteq \{q \in Q : p(j\omega, q) \in \mathcal{P}(\omega) \text{ for } 0 \leq \omega < \infty\}$$

and note that $p_0(j\omega)$ is the center of the frequency dependent rectangles in Figure 10. Hence, Q_{good} is convex and symmetric and the Uniformity Principle applies; that is

$$p_\Omega \doteq \min_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good}\} = \text{Prob}\{q^u \in Q_{good}\}.$$

Now, an estimate $\widehat{p}_\Omega$ of p_Ω is obtained using a uniform sampling distribution over Q. For this example, using $N = 10^6$ samples, it turns out that

$$\widehat{p}_\Omega \approx 0.9969.$$

In conclusion, the inequality

$$\widehat{p}_\Omega \leq \min_{f \in \mathcal{F}} \text{Prob}\{p(s, q^f) \text{ is stable}\},$$

guarantees a distributionally robust probability estimate of stability of at least 0.9969.

5.11 Example (Robust Least Squares): Consider an over-determined set of linear equations

$$Ax = b; \quad x \in \mathbf{R}^n; \quad b \in \mathbf{R}^m; \quad m \leq n,$$

with matrix $A \in \mathbf{R}^{m \times n}$ having rank m. Given data $A = A_0$ and $b = b_0$, the classical least squares problem is to as find $x = x_{LS}$ which minimizes the residual cost function

$$J_{LS}(x) \doteq \|A_0 x - b_0\|.$$

It is well known that the solution for this problem is

$$x_{LS} = (A_0^T A_0)^{-1} A_0^T b_0.$$

Now, in the presence of data uncertainty, a distributional robustness problem arises: Indeed, with both A and b having uncertain parameters q_i entering affine linearly, we write

$$A(q) \doteq A_0 + \sum_{i=1}^{\ell} q_i A_i$$

and

$$b(q) \doteq b_0 + \sum_{i=1}^{\ell} q_i b_i$$

where the $n \times m$ matrices and the $m \times 1$ vectors b_i above are fixed.

Now, to study this problem using the results in this paper, the uncertain parameter vector q is assumed to have probability distribution $f \in \mathcal{F}$ and $\gamma > 0$ is taken to be an acceptable residual cost level. Now, with

$$\Phi(f) \doteq \text{Prob}\{\|A(q^f)x_{LS} - b(q^f)\| \leq \gamma\}$$

and nominal solution x_{LS} satisfying $A_0 x_{LS} = b_0$, the associated good set

$$Q_{good} \doteq \{q \in Q : \|A(q)x_{LS} - b(q)\| \leq \gamma\}$$

is readily shown to be both convex and symmetric. Hence, the Uniformity Principle applies and it is concluded that

$$\min_{f\in\mathcal{F}} \Phi(f) = \Phi(u).$$

5.12 Numerical Example: We now present an example which illustrates application of the least squares result above. Indeed, with

$$A(q) = \begin{bmatrix} -1+q_1 & -2+q_2 & -4+q_3 \\ -5+q_4 & 5+q_5 & -9+q_6 \\ -3+q_7 & -3+q_8 & -7+q_9 \\ -1+q_{10} & -2+q_{11} & -4+q_{12} \\ -1+q_{13} & 4+q_{14} & -1+q_{15} \end{bmatrix}$$

and

$$b(q) = b_0 = [-7 \ -28 \ -14 \ -7 \ -7]^T,$$

we first compute the classical least squares solution

$$x_{LS} = [1 \ -1 \quad 2]^T$$

corresponding to $q = 0$. Now, with uncertainty dimension $\ell = 15$, we assume that the distribution of the uncertain vector q belongs to the class $\mathcal{F}$ and we analyze the performance of x_{LS} for different radii for the uncertainty. More precisely, we assume that $|q_i| \leq r$ and study the effects of varying the radius r with x_{LS} held fixed. First, it is noted that classical robustness theory indicates that the maximum allowed radius is $r_{max} \approx 0.0112$ with performance specification

$$\|A(q)x_{LS} - b(q)\| \leq 0.1$$

is satisfied by all allowed values of q if and only if $r < r_{max} \approx 0.0112$. Now, we take the distributionally robust point of view and seek to compute

$$\Phi^* = \min_{f\in\mathcal{F}} \text{Prob}\{\|A(q^f)x_{LS} - b(q^f)\| \leq \gamma\}$$

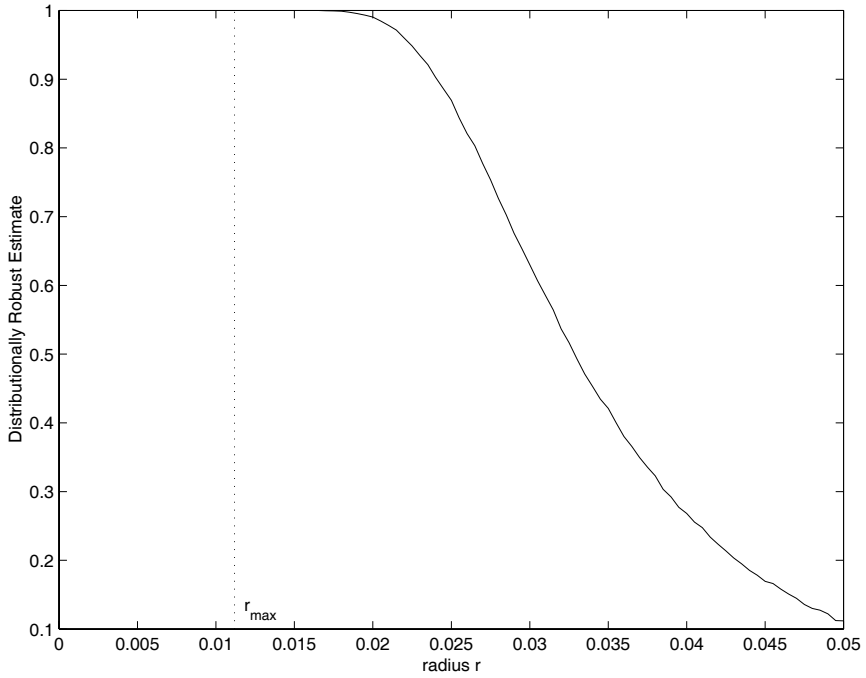

Figure 11: Distributionally Robust Probability of Performance Satisfaction

as a function of the uncertainty radius r. The results obtained are depicted in Figure 11. To illustrate how conservative a classical robustness measure can be, we take radius of uncertainty of $r = 0.018$, which is approximately 60% larger then the r_{max}. For this radius, the distributionally robust risk of performance violation is only $\varepsilon \approx 0.0001$.

For the case of expected performance, it is also interesting to note that least squares analysis can be carried out for so-called multilinear uncertainty structures. More specifically, considering the setup above with the entries of $A(q)$ and $b(q)$ depending multilinearly on q, it follows from the Componentwise Convexity Theorem (see [4] for details) that

$$\max_{f \in \mathcal{F}} \mathcal{E}(\|A(q^f)x_{LS} - b(q^f)\|) = \mathcal{E}(\|A(q^u)x_{LS} - b(q^u)\|).$$

6. Non-Symmetric and Non-Convex Cases

When Q_{good} in not convex or symmetric, we seek an optimal truncation $t^* \in T$, as prescribed in Theorem 4.2. Since finding t^* may be computationally complex, it is of interest to circumvent this problem. To this end, as illustrated by the polynomial example in Section 5.10, it is often possible to obtain a lower bound on the probability of performance satisfaction. In Section 6.1 below, results are given when the convexity property is satisfied but the symmetry property is violated. In Section 6.3, a method is described which is applicable to cases for which there exists a deterministic algorithm for testing the satisfaction of performance specifications on "rectangles" of uncertainty.

6.1 Symmetrization: To motivate the so-called symmetrization approach, consider the problem of Lyapunov stability with a $n \times n$ state space matrix $A(q)$ having entries depending affinely on the uncertainty vector q and fixed $n \times n$ positive-definite Lyapunov matrix P. Now, consistent with standard Lyapunov theory, for example, see [52], we take

$$Q_{good} \doteq \{q \in Q : A^T(q)P + PA(q) < 0\}.$$

To motivate the construction below, it is noted that the set Q_{good} above is readily verified to be convex but is not necessarily symmetric.

For cases such as the one above, it proves useful to consider the *symmetrization* of Q_{good} given by

$$Q_{good,s} \doteq \{q : q \in Q_{good} \text{ and } -q \in Q_{good}\}.$$

Now, since $Q_{good,s}$ is both convex and symmetric, the Uniformity Principle guarantees

$$\min_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good,s}\} = \text{Prob}\{q^u \in Q_{good,s}\}.$$

Furthermore, since the containment

$$Q_{good,s} \subseteq Q_{good}$$

holds, the performance estimate obtained using $Q_{good,s}$ is a lower bound for the true performance. Now, combining these considerations with the applicability of the Uniformity Principle for $Q_{good,s}$, we obtain

$$\text{Prob}\{q^u \in Q_{good,s}\} = \min_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good,s}\} \leq \inf_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good}\}.$$

In practice, it often turns out that this bound is quite useful and the truncation problem is avoided.

Given that a lower bound on performance is being computed above, the issue of the conservatism of the estimate arises. In this regard, it can be easily seen, using standard reasoning on probability of sets, that for any $f \in \mathcal{F}$,

$$\text{Prob}\{q^f \in Q_{good}\} \geq \text{Prob}\{q^f \in Q_{good,s}\} \geq 2\ \text{Prob}\{q^f \in Q_{good}\} - 1.$$

In other words, for high performance problems, the bound obtained using symmetrization becomes tight.

6.2 Example: Figure 12 depicts a mechanical system consisting of four blocks with uncertain dampers and springs. With eight uncertain parameters with ranges $0.8 \leq b_i \leq 2.2$ and $0.8 \leq k_i \leq 2.2$ for $i = 1, \ldots, 4$ and all unit masses $m_i = 1$, the performance objective is to keep the gain of the system below a level $\overline{g}$ for all frequencies. In other words, one wants to keep the transfer function magnitude $|H(j\omega)|$ from F to y below $\overline{g}$ at all frequencies $\omega \geq 0$. Identifying q_1, q_2, q_3 and q_4 with deviations from the center points of the intervals of spring constants and q_5, q_6, q_7 and q_8 with deviations from the center points of the intervals of damper constants, the theory of Linear Matrix Inequalities, for example, see [52], provides a sufficient condition for this specification to be satisfied. Namely, with

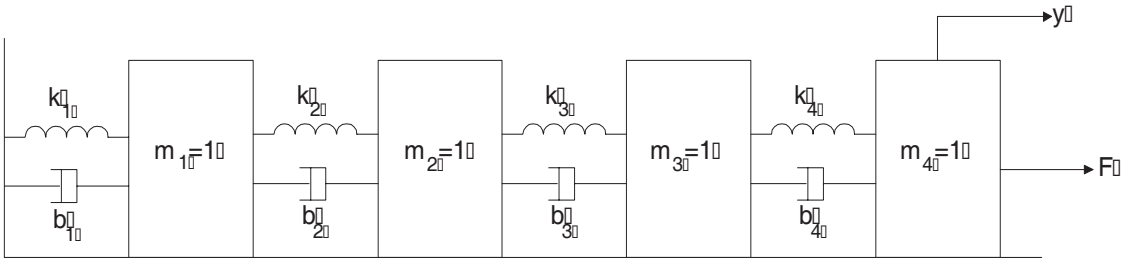

Figure 12: Mechanical System

$$A_0 \doteq A(0) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1.5 & -1.5 & 1.5 & 1.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1.5 & 1.5 & -3 & -3 & 1.5 & 1.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1.5 & 1.5 & -3 & -3 & 1.5 & 1.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1.5 & 1.5 & -3 & -3 \end{bmatrix},$$

$$B_0 = B(0) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T, \quad C_0 = C(0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M_0 \doteq \begin{bmatrix} A_0 S + S A_0^T & S C_0^T & B_0 \\ C_0 S & -\overline{g} I & 0 \\ B_0^T & 0 & -\overline{g} I \end{bmatrix},$$

$\Delta A(q)$ having the appropriate uncertainties corresponding to the non-zero entries of A_0 and $\Delta B(q) = \Delta C^T(q) = 0$, performance is guaranteed if

$$M_0 + \Delta M(q) < 0$$

where

$$\Delta M(q) \doteq \begin{bmatrix} \Delta A(q) S + S \Delta A^T(q) & S \Delta C^T(q) & \Delta B(q) \\ \Delta C(q) S & 0 & 0 \\ \Delta B^T(q) & 0 & 0 \end{bmatrix}.$$

Taking

$$Q_{good} \doteq \{q : M_0 + \Delta M(q) < 0\}$$

and noting that this set is convex but not symmetric, symmetrization was used to assess the probability of performance satisfaction. A performance level of $\overline{g} = 6.2076$ was considered and 20,000 samples were used to estimate performance. The value obtained via a Monte Carlo simulation was

$$\min_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good,s}\} = \text{Prob}\{q^u \in Q_{good,s}\} \approx 0.99.$$

Hence, in this case, we obtain an estimate of probability of performance satisfaction which satisfies

$$0.99 \leq \inf_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good}\} \leq 0.995.$$

6.3 Unirectangularity: In this section, we describe the method in [29] which applies to many cases when the set Q_{good} is non-convex. Central to this method is the concept of a *unirectangular* set which is described below.

As a first step, we define the concept of *rectangular projection.* That is, given a point $q \in \mathbf{R}^\ell$, its rectangular projection $\mathcal{R}(q)$ is taken to be the box whose extremes are the point q and the origin. Namely,

$$\mathcal{R}(q) \doteq \{(\alpha_1 q_1, \alpha_2 q_2, \ldots, \alpha_\ell q_\ell) : \ \alpha_i \in [0,1] \text{ for } i = 1, 2, \ldots, \ell\}.$$

Now, a set Q_{good} is said to be *unirectangular* if the rectangular projection of any point q belonging to Q_{good} is contained in Q_{good}; i.e., if $q \in Q_{good}$ then $\mathcal{R}(q) \subseteq Q_{good}$. An example of a unirectangular set is shown in Figure 14. The result below, established in [29], motivates some of the analysis to follow.

6.4 Unirectangularity Principle: *If Q_{good} is unirectangular then,*

$$\min_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good}\} = \text{Prob}\{q^u \in Q_{good}\}.$$

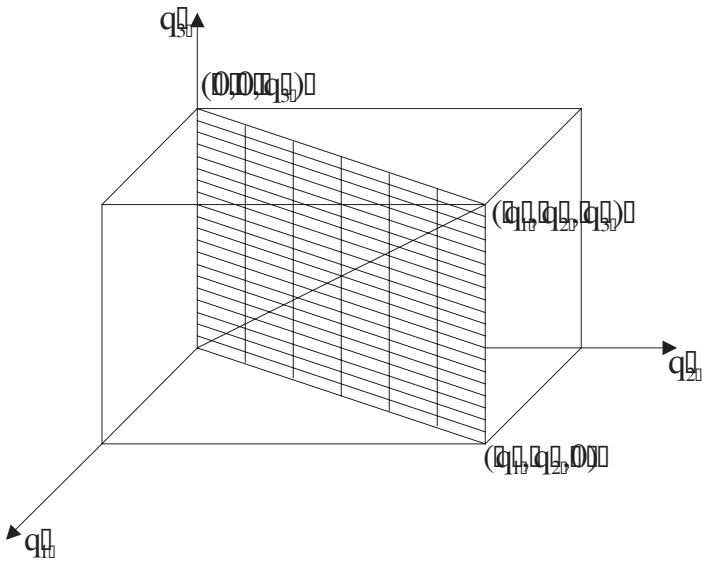

Figure 13: Rectangular Projection Set

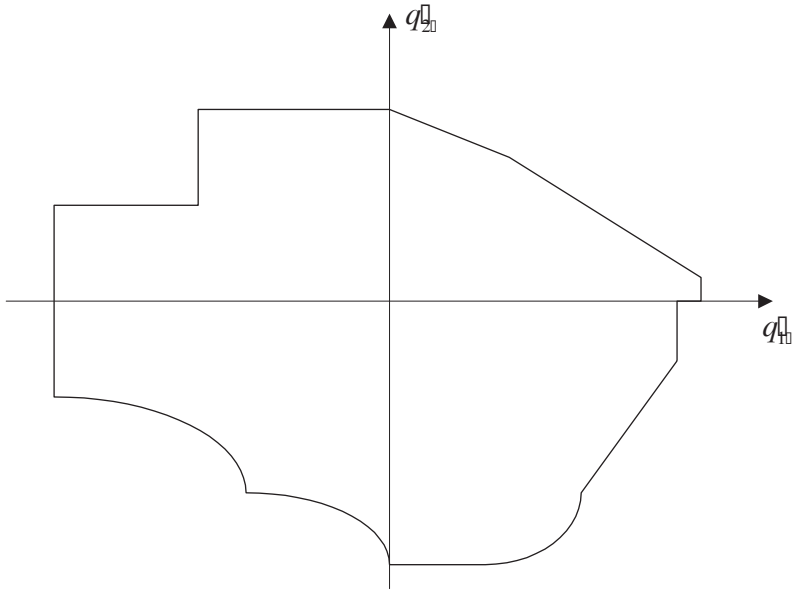

Figure 14: A Unirectangular Set

6.5 Continuation of Unirectangularity: The fact that a Uniformity Principle is also valid for unirectangular sets is the basis for the method described in [29]. This method is applicable to all problems for which there exists a deterministic algorithm $\mathcal{A}$ which can test if a given rectangle is contained in Q_{good}. More specifically, to obtain a lower bound on the probability of performance satisfaction, for a given uncertainty box Q, let

$$\mathcal{A}(Q) \doteq \begin{cases} 1 & \text{if } q \in Q_{good} \text{ for all } q \in Q; \\ 0 & \text{otherwise.} \end{cases}$$

For example, if $\mathcal{A}$ corresponds to an algorithm for testing some inequality guaranteeing satisfaction of the desired performance specifications, then $\mathcal{A}(Q) = 1$ indicates that this inequality is satisfied for all $q \in Q$. Another possibility is that the algorithm $\mathcal{A}$ corresponds to the implementation of some robustness result such as Kharitonov's Theorem or a structured singular value criterion.

Next, we describe the method for estimating the probability of performance. In accordance with [29], if one draws N samples $q^1, q^2, \ldots, q^N$ uniformly distributed over Q, it can be shown that

$$\inf_{f \in \mathcal{F}} \text{Prob}\{q^f \in Q_{good}\} \geq \frac{\sum_{i=1}^{N} \mathcal{A}(\mathcal{R}(q^k))}{N} \doteq \widehat{p}.$$

Hence, the estimate $\widehat{p}$ above is a lower bound on the probability of performance satisfaction.

6.6 Example (Interval Polynomial): For the second time, the interval polynomial of Section 3.5 is revisited with the same uncertainty bound $r_i = 0.03$ for $i = 1, 2, \ldots, 12$. In this case, the algorithm $\mathcal{A}$ corresponds to the application of Kharitonov's Theorem. That is, $\mathcal{A}[\mathcal{R}(q)] = 1$ if the four Kharitonov polynomials associated with $\mathcal{R}(q)$ are stable and zero otherwise. The algorithm above was applied with $N = 100,000$ resulting in the estimate

$$\widehat{p} \approx 0.99936.$$

7. Spherical Setting: A Brief Introduction

Thus far, this paper has concentrated on cases with the so-called *structured uncertainty* entering the model. In this section, we consider cases where the uncertainty is unstructured. In this regard, the method for analysis of *unstructured* uncertainty of [31] is briefly introduced. In this new setting, the first point to note is that the description of $\mathcal{F}$ given in Section 3.2 is unsuitable. That is, for the case of unstructured uncertainty, it is unreasonable to assume that the uncertain parameters vary independently. This observation motivates a new definition for the set of probability distributions $\mathcal{F}$ so as to accommodate parameter dependency.

7.1 New Definition of the Class $\mathcal{F}$: Using the Euclidean norm for q and taking

$$Q \doteq \{q : \|q\| \leq r\},$$

a probability density function f is said to belong to the class $\mathcal{F}$ if there exists a nondecreasing function $g(\cdot)$ with scalar argument such that

$$f(x) = g(\|x\|)$$

for all x. Intuitively, this says that larger uncertainty values are less likely than smaller values and that all uncertainty "directions" are equally probable.

7.2 Truncations: Analogous to the development in Sections 1–6, in this spherical setting, a class of radially truncated uniform distributions is defined. Indeed, letting $0 \leq t \leq r$ denote a truncation radius, the truncated uniform distribution u^t is the uniform distribution over the truncated sphere

$$Q_t \doteq \{q : \|q\| \leq t\}.$$

For example, if Q is the unit sphere, then the uniform distribution over the sphere of radius $t = 1/3$ would be a radial truncation.

In this radial distribution framework, perhaps a most important observation is that there is only one truncation parameter, no matter what the dimension of q. Therefore, the problem of finding a optimal truncation $t^* \in T$ is greatly simplified. That is, one need only conduct a single variable line search in the variable t.

7.3 Truncation Principle: Analogous to the case of independent uncertainty, it is shown in [31] that the Truncation Principle

$$\inf_{f \in \mathcal{F}} \Phi(f) = \inf_{t \in T} \Phi(u^t)$$

also holds in the spherical uncertainty case. As seen in the example below, this result readily lends itself to numerical computation.

7.4 Example: To illustrate a typical problem which is addressed in the spherical setting, consider the n-dimensional state space system

$$\dot{x} = A(q)x$$

having uncertain matrix $A(q)$ which is partially structured and given by

$$A(q) \doteq A_0 + B_0 \Delta(q) C_0$$

where A_0 is a fixed *nominal* state matrix, B_0 is a fixed $n \times m$ matrix, C_0 is a fixed $r \times n$ matrix and $\Delta(q)$ is an $m \times r$ random matrix representing the uncertainty. Now, with Q being a sphere representing unstructured variations and

$$Q_{good} \doteq \{q \in Q : A_0 + B_0 \Delta(q) C_0 \text{ is stable}\},$$

we consider the problem of finding $t^* \in T$ leading to

$$\min_{t \in T} \text{Prob}\{q^{u^t} \in Q_{good}\} = \text{Prob}\{q^{u^{t^*}} \in Q_{good}\}.$$

Now, with the specific problem data

$$A_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -6812 & -3090.6 & -913.6 & -235.10 & -28.1 \end{bmatrix};$$

$$B_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 \end{bmatrix}^T;$$

$$C_0 = \begin{bmatrix} 520 & 226 & 56 & 11 & 1 \\ 15.6 & 6.78 & 1.68 & 0.33 & 0.03 \end{bmatrix},$$

it follows that

$$\Delta = [q_1 \quad q_2]$$

is a 2-dimensional row vector and Property $\mathcal{P}$ is deemed to be satisfied if and only if the uncertain matrix $A = A_0 + B_0 \Delta C_0$ is stable. Now, using the Truncation Principle in this spherical setting with uncertainty radius $r = 13.188$ and $N = 300,000$ samples, the function

$$\Phi(t) = \text{Prob}\{u^t \in Q_{good}\}$$

is studied and found to have minimizer $t^* \approx 11.78$; this which corresponds to a probability of stability $\Phi(t^*) \approx 0.8149$. In contrast, with uniform distribution, one obtains $\Phi(r) \approx 0.8193$. That is, common sense use of the uniform distribution in lieu of u^{t^*} leads to a probability estimate which is slightly more optimistic than the distributionally robust result.

7.5 Uniformity Principle: For the case of spherical uncertainty, it is shown in [31] that a Uniformity Principle holds under weaker hypothesis than in the independent parameter case. That is, instead of requiring Q_{good} to be convex and symmetric, we only require Q_{good} to be *star-shaped*; i.e., if $q \in Q_{good}$ then $\lambda q \in Q_{good}$ for all $\lambda \in [0, 1]$.

An example illustrating satisfaction of the star-shaped requirement is obtained from the theory of quadratic stability. Indeed, suppose that A_0 is an $n \times n$ stable matrix and $P = P^T > 0$ is an $n \times n$ candidate Lyapunov matrix satisfying

$$A_0^T P + P A_0 < 0.$$

Now, suppose A_0 is replaced by $A = A_0 + \Delta A(q)$ and we want to determine how large $||\Delta A(q)||$ can be while preserving the stability inequality above. Then, if $\Delta A(q)$ is a linear function of q and

$$Q_{good} \doteq \{q \in Q : A^T P + P A < 0\},$$

it is easy to verify that the resulting set Q_{good} is star-shaped. Hence, in view of the Uniformity Principle, a uniform sampling scheme can be used in a distributionally robust Monte Carlo simulation.

8. Conclusion

Distributionally robust Monte Carlo simulation is a research area which is still in its infancy. As seen in this paper, many of the problems in the area reduce to finding a so-called optimal truncation vector $t^* \in T$ which defines the required interval for uniform sampling. It was also seen that there are many special cases for which this truncation-finding problem is readily solved. For example, when Q_{good} is convex and symmetric, the Uniformity Principle was seen to apply; i.e., one simply takes all $t_i^* = r_i$ corresponding to a uniform distribution. A second special case was seen to involve classes of circuits for which distributional robustness was obtained with an extreme distribution, uniform or impulsive. Finally, a number of special cases were described for which one obtains a distributionally robust lower bound for the probability of performance satisfaction.

By way of future research, there are many open problems involving some aspect of truncation-finding. Most notably, when the performance specification function $\phi(q)$ comes from a specific physical generating mechanism, analogous to the case of resistive networks in Section 5.3, it is of interest to investigate the extent to which exploitation of the structure of ϕ may lead to a solution of the truncation problem. In this regard, there are many control theoretic problems of interest. To illustrate, if $H(s,q)$ is a transfer function obtained from a signal flow graph with uncertain branch gains q_i, the manner in which these gains enter H might be exploited to find the desired truncations t_i^*. This is simply one of many examples of problems with a system theoretic flavor which would be worthy of investigation in the distributional robustness framework. Finally, it is felt that further research involving the spherical setting of Section 7 would be worthwhile. For problems lending themselves to this setting, truncation-finding is not a serious problem because only one truncation parameter is involved.

A second important line of future research involves what might appropriately be termed *distributionally robust design*. To this end, it should be noted the results described in this paper were entirely of an analysis nature; i.e., there were no design variables entering the performance specification $\phi(q)$. It would be of interest to extend the results reported here classes of problems for which a design vector x enters ϕ. For example, one considers a performance specification $\phi(x,q)$ and the goal is to select x so as to provide the best possible level of performance which is distributionally robust with respect to $f \in \mathcal{F}$. Some initial results in this area are given in [3] and [44].

References

[1] Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Methods*, Wiley, New York.

[2] Barmish, B. R. and C. M. Lagoa (1997). "The Uniform Distribution: A Rigorous Justification for its Use in Robustness Analysis," *Mathematics of Control Signals and Systems*, vol. 10, pp. 203-222.

[3] Lagoa, C. M. (1998). *Contributions to the Theory of Probabilistic Robustness*, Ph.D. Dissertation, ECE Department, University of Wisconsin, Madison.

[4] Barmish, B. R. and P. Shcherbakov (1999). "Distributionally Robust Least Squares," Proceedings of the *International Symposium on Adaptive Systems*, St. Petersburg, Russia.

[5] Devroye, L. (1986). *Non-Uniform Random Variable Generation*, Springer-Verlag, New York.

[6] Chernoff, H. (1952). A Measure of Asymptotic Efficiency for Test of Hypothesis Based on the Sum of Observations, *Annals of Mathematical Statistics*, vol. 23, pp. 493–507.

[7] Fok, D. S. K. and D. Crevier (1989). Volume Estimation by Monte Carlo Methods, *Journal of Statistical Computation and Simulation*, vol. 31, pp. 223–235.

[8] Calafiore, G., F. Dabbene and R. Tempo (1998). "Uniform Sample Generation in l_p Balls for Probabilistic Robustness Analysis," Proceedings of the *IEEE Conference on Decision and Control*, Tampa.

[9] Calafiore, G., F. Dabbene and R. Tempo (2000). "Randomized Algorithms for Probabilistic Robustness With Real and Complex Structured Uncertainty," *IEEE Transactions on Automatic Control*, AC–45, pp. 2218–2235.

[10] Huber, P. J. (1981). *Robust Statistics*, John Wiley & Sons, New York.

[11] Barmish, B. R. and B. T. Polyak (1996). "A New Approach to Open Robustness Problems Based on Probabilistic Prediction Formulae," Proceedings of the *IFAC World Congress*, San Francisco.

[12] Chen X. and K. Zhou (1997). "A Probabilistic Approach to Robust Control," Proceedings of the *IEEE Conference on Decision and Control*, San Diego.

[13] Chen, X. and K. Zhou (1998). "Constrained Optimal Synthesis and Robustness Analysis by Randomized Algorithms," Proceedings of the *American Control Conference*, Philadelphia.

[14] Gazi, E., W. D. Seider and L. H. Ungar (1997). "A Non-parametric Monte Carlo Technique for Controller Verification," *Automatica*, vol. 33, pp. 901–906.

[15] Marrison, C. and R. Stengel (1994). "The Use of Random Search and Genetic Algorithms to Optimize Stochastic Robustness Functions," Proceedings of the *American Control Conference*, Baltimore.

[16] Petersen, I. R. , S. O. R. Moheimani and H. R. Pota (1999). "Minimax LQG Optimal Control of Acoustic Noise in a Duct," Proceedings of the *World Congress of IFAC,* Beijing, China.

[17] Ray, L. R. and R. F. Stengel (1993). "A Monte Carlo Approach to the Analysis of Control System Robustness," *Automatica*, vol. 29, pp. 229–236.

[18] Stengel, R. F. and L. R. Ray (1991). "Stochastic Robustness of Linear Time-Invariant Control Systems," *IEEE Transactions on Automatic Control*, AC–36, pp. 82–87.

[19] Stengel, R. F., L. R. Ray and C. I. Marrison (1995). "Probabilistic Evaluation of Control Systems Robustness," *International Journal of Systems Science*, vol. 26, pp. 1363–1382.

[20] Yoon, A., P. Khargonekar and K. Hebbale (1997). "Design of Computational Experiments for Open-Loop Control and Robustness Analysis of Clutch–to–Clutch Shifts in Automatic Transmissions," Proceedings of the *American Control Conference*, Albuquerque.

[21] Zhu, X., Y. Huang and J. Doyle (1996). "Soft vs. Hard Bounds in Probabilistic Robustness Analysis," Proceedings of the *IEEE Conference on Decision and Control*, Kobe, Japan.

[22] Goodwin, G. C., L. Wang and D. Miller (1999). "Bias-Variance Trade-Off Issues in Robust Controller Design Using Statistical Confidence Bounds," Proceedings of the *IFAC World Congress*, Beijing.

[23] Zhu, X. (1999) "Probabilistic μ Upper Bound Using Linear Cuts," Proceedings of the *IFAC World Congress*, Beijing.

[24] Calafiore, G., F. Dabbene, R. Tempo. (1999). "The Probabilistic Real Stability Radius," Proceedings of the *IFAC World Congress*, Beijing.

[25] Hara, S. and T. Miyazato. (2000) "A Probabilistic Approach To Robust Control Design," Proceedings of the *IEEE Conference on Decision and Control,* Sydney.

[26] Barmish, B. R., C. M. Lagoa and P. S. Shcherbakov (1996). "Probabilistic Enhancement of Robustness Margins Provided by Linear Matrix Inequalities," Proceedings of the *Allerton Conference on Communication, Control and Computing*, Monticello.

[27] Barmish, B. R., C. M. Lagoa and R. Tempo (1997). "Radially Truncated Uniform Distribution for Probabilistic Robustness of Control Systems," Proceedings of the *American Control Conference*, Albuquerque.

[28] Barmish, B. R. and P. S. Shcherbakov (1999). "Distributionally Robust Least Squares," Proceedings of *SPAS'99*, St. Petersburg, Russia.

[29] Lagoa, C. M., P. S. Shcherbakov and B. R. Barmish (1998). "Probabilistic Enhancement of Classical Robustness Margins: The Unirectangularity Concept," *Systems and Control Letters*, vol. 35, pp. 31–43.

[30] Zhang, J., C. M. Lagoa and B. R. Barmish (1998). "Probabilistic Robustness: An RLC Realization of the Truncation Phenomenon," Proceedings of the *American Control Conference*, Philadelphia.

[31] Barmish, B. R., C. M. Lagoa and R. Tempo (1997). "Radially Truncated Uniform Distribution for Probabilistic Robustness of Control Systems," Proceedings of the *American Control Conference*, Albuquerque.

[32] Muhler, M. L. and B. R. Barmish, (1999). "A Uniformity Principle for Probabilistic Robustness of the Discrete-Time Linear Quadratic Regulator," Proceedings of the *American Control Conference*, San Diego,.

[33] Shcherbakov, P. S., C.M. Lagoa and B. R. Barmish (1999). "Characterization of Worst-Case Uncertainty Geometry in the Theory of Probabilistic Robustness," Proceedings of the *IFAC World Congress*, Beijing.

[34] Barmish, B. R. and H. Kettani (2000). "Monte Carlo Analysis of Resistive Networks Without Apriori Probability Distributions," Proceedings of the *International Symposium on Circuits and Systems*, Geneva, Switzerland.

[35] Kettani, H. and B. R. Barmish (2000). "A New Approach to Monte Carlo Analysis Illustrated for Resistive Networks," Proceedings of the *Conference on Information Sciences and Systems,*, Princeton University, Princeton.

[36] Barmish, B. R. (2000) "A Probabilistic Robustness Result For a Multilinearly Parameterized H_∞ Norm," Proceedings of the *American Control Conference*, Chicago.

[37] Bai, E.-W., R. Tempo and M. Fu (1997). Worst Case Properties of the Uniform Distribution and Randomized Algorithms for Robustness Analysis, Proceedings of the *American Control Conference*, Albuquerque, New Mexico.

[38] Djavdan, P., H. J. A. F. Tulleken, M. H. Voetter, H. B. Verbruggen and G. J. Olsder (1989). "Probabilistic Robust Controller Design," Proceedings of the *IEEE Conference on Decision and Control*, Tampa.

[39] Khargonekar, P. and A. Tikku (1996). "Randomized Algorithms for Robust Constrol Analysis and Synthesis Have Polynomial Complexity," Proceedings of the *IEEE Conference on Decision and Control*, Kobe, Japan.

[40] Tempo, R., E. W. Bai and F. Dabbene (1996). "Probabilistic Robustness Analysis: Explicit Bounds for the Minimum Number of Samples," Proceedings of the *IEEE Conference on Decision and Control*, Kobe, Japan.

[41] Tempo, R., E. W. Bai and F. Dabbene (1997). "Probabilistic Robustness Analysis: Explicit Bounds for the Minimum Number of Samples," *Systems and Control Letters*, vol. 30, pp. 237–242.

[42] Vidyasagar, M. (1997). *A Theory of Learning and Generalization*, Springer Verlag, London, England.

[43] Vidyasagar, M. (1997). "Statistical Learning Theory: An Introduction and Applications to Randomized Algorithms," Proceedings of the *European Control Conference*, Brussels, Belgium.

[44] Barmish, B. R. and C. M. Lagoa (1999). "On the Convexity of the Probabilistic Design Problem for Quadratic Stabilizability," Proceedings of the *American Control Conference*, San Diego.

[45] Director, S. W. and G. D. Hachtel (1977). "The Simplicial Approximation Approach to Design Centering," *IEEE Transactions on Circuits and Systems*, vol. CAS-24, pp. 363-372.

[46] Director, S. W. and L. M. Vidigal (1981), "Statistical Circuit Design: A Somewhat Biased Survey," Proceedings of the *European Conference on Circuit Theory Design*, The Hague, The Netherlands.

[47] Pinel, J. F. and K. Singhal (1977). "Efficient Monte Carlo Computation of Circuit Yield Using Importance Sampling," Proceedings of the *IEEE International Symposium on Circuits and Systems*.

[48] Spence, R. and R. S. Soin (1988). *Tolerance Design of Electronic Circuits,* Addison-Wesley, New York.

[49] Barmish, B. R. (1985). "Necessary and Sufficient Conditions for Quadratic Stabilizability of an Uncertain Linear System," *Journal of Optimization Theory and Applications*, vol. 46, pp. 399–409.

[50] Barmish, B. R. (1994). *New Tools for Robustness of Linear Systems,* McMillan, New York.

[51] Kharitonov, V. L. (1978). "Asymptotic Stability of an Equilibrium Position of a Familiy of Systems of Linear Differential Equations," *Differentsial'nye Uravnenyia*, pp. 2086–2088.

[52] Boyd, S. P., L. El Ghaoui, E. Feron and V. Balakrishnan (1994). *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia.

[53] Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, Princeton.