Distributionally Robust Optimization for Design under Partially Observable Uncertainty

by

Michael George Kapteyn

BE. (Hons) Engineering Science, The University of Auckland (2016)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Signature redacted

Author .....
Department of Aeronautics and Astronautics
May 24, 2018

Signature redacted

Certified by .....
Karen E. Willcox
Professor of Aeronautics and Astronautics
Thesis Supervisor

Signature redacted

Accepted by .....
Hamsa Balakrishnan
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Distributionally Robust Optimization for Design under Partially Observable Uncertainty

by

Michael George Kapteyn

Submitted to the Department of Aeronautics and Astronautics
on May 24, 2018, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

Deciding how to represent and manage uncertainty is a vital part of designing complex systems. Widely used is a probabilistic approach: assigning a probability distribution to each uncertain parameter. However, this presents the designer with the task of assuming these probability distributions or estimating them from data, tasks which are inevitably prone to error. This thesis addresses this challenge by formulating a distributionally robust design optimization problem, and presents computationally efficient algorithms for solving the problem. In distributionally robust optimization (DRO) methods, the designer acknowledges that they are unable to exactly specify a probability distribution for the uncertain parameters, and instead specifies a so-called ambiguity set of possible distributions. This work uses an acoustic horn design problem to explore how the error incurred in estimating a probability distribution from limited data affects the realized performance of designs found using traditional approaches to optimization under uncertainty, such as multi-objective optimization. It is found that placing some importance on a risk reduction objective results in designs that are more robust to these errors, and thus have a better mean performance realized under the true distribution than if the designer were to focus all efforts on optimizing for mean performance alone. In contrast, the DRO approach is able to uncover designs that are not attainable using the multi-objective approach when given the same data. These DRO designs in some cases significantly outperform those designs found using the multi-objective approach.

Thesis Supervisor: Karen E. Willcox
Title: Professor of Aeronautics and Astronautics

Acknowledgments

Firstly I would like to thank Professor Karen Willcox, who provided me with an incredible opportunity when she welcomed me into her research group, and who continues to offer me unrivaled guidance, teaching, mentorship, and support. I would also like to thank Professor Andy Philpott for the initial idea for this research, and for his guidance and insightful, thought-provoking discussions throughout it's completion.

I would like to thank the SUTD-MIT International Design Center for providing me with the opportunity to visit Singapore, and Karen, Max and Alex for making the experience an enjoyable and memorable one.

I would also like to thank all of my fellow students in the Aerospace Computational Design Lab (ACDL) for their support, friendship, and welcome distractions from productive work. Additional thanks go out to my housemates Conor, Tim, Andrew and Gary, as well as Richard, Ben, Jeff, Nicole and the rest of my expatriated Kiwi friends for providing me with a home away from home.

Last but certainly not least I would like to thank my family. Mum, Dad, Carmen, Andy, and Megan—Thank you for your everlasting love and support. I hope I can make all of you proud.

None of this would have been possible without financial support. This work was supported in part by AFOSR grant FA9550-16-1-0108 under the Dynamic Data Driven Application Systems Program, by the Defense Advanced Research Projects Agency [EQUiPS program, award W911NF-15-2-0121, Program Manager F. Fahroo], by a New Zealand Marsden Fund grant under contract UOA1520, and by the SUTD-MIT International Design Center.

TABLE OF CONTENTS

NOMENCLATURE

Greek Symbols

Symbol	Description
Γ_{in}	Computational boundary at the acoustic horn inlet
Γ_N	Computational Boundary at the acoustic horn walls
Γ_R	Truncated absorbing boundary of the computational domain in the acoustic horn problem
κ	Kolmogorov-Smirnov (K-S) distance between a random sample and the true underlying distribution
λ	Mean versus standard deviation trade-off parameter in the Multi-Objective Optimization $\mathcal{M}$
$\mu(\cdot)$	Average of the argument over T samples
Ω^m	Space of all vectors of representing discrete probability measures over m support values
$\phi(\cdot)$	CVaR_{10} of the argument over T samples

Roman Symbols

Symbol	Description
a	Half-width of the horn inlet
b	Half-width of the horn outlet
$\mathcal{D}$	Distributionally robust design optimization problem
$D_{\text{L2}}(\cdot, \cdot)$	Function measuring the K-L divergence between two discrete probability vectors
$D_{\text{KL}}(\cdot, \cdot)$	Function measuring the L_2-norm of the difference between two discrete probability vectors
$\mathbf{e}_j$	Unit vector with a 1 in position j
h_1, h_2	Half-widths at uniformly spaced points in the horn flare

i	Imaginary unit, $\sqrt{-1}$
k	Wave-number in the acoustic horn problem
L	Length of the horn from inlet to outlet
$\mathcal{M}$	Multi-objective design optimization problem
m	Sample size
$\mathcal{P}$	Full observability design optimization problem
$\mathcal{P}$	Ambiguity Set
$\hat{\mathbf{p}}$	Empirical distribution from a sample of the uncertain parameters
$\mathbb{P}_Q$	Probability density function of the quantity of interest
$\mathbf{p}^*$	Worst-case distribution within the ambiguity set
$\mathbb{P}_{\mathbf{u}}$	Probability density function of the uncertain parameters
Q	Quantity of interest (QoI)
R	Radius of the truncated absorbing boundary in the horn model
r	Radius of ambiguity
$\bar{r}$	Normalized radius of ambiguity (0–1)
$\mathcal{S}$	Sample average approximation-based design optimization problem
s	Reflection coefficient of the acoustic horn
s_m	Standard deviation in performance of design $\mathbf{x_m}$ under the true distribution $\mathbb{P}_{\mathbf{u}}$
T	Number of random samples used in computational experiments
$\mathbf{u}$	Vector of uncertain parameters
$\mathcal{U}$	Space of possible uncertain parameter vectors
v	Non-dimensionalized pressure
$\mathbf{x}$	Vector of design variables
$\mathcal{X}$	Space of possible design vectors
$\mathbf{x}_\infty$	True optimal design—an optimizer of $\mathcal{P}$
$\mathbf{x}_m$	Optimal design computed using a sample size m

z	Dummy variable used in the optimization over a K-L divergence ambiguity set
Z_∞	True optimal mean performance—the optimal objective value in $\mathcal{P}$
Z_m	Mean performance of design $\mathbf{x_m}$ under the true distribution $\mathbb{P}_\mathbf{u}$
$\hat{Z}_m$	Mean performance of design $\mathbf{x_m}$ under the sample distribution $\hat{\mathbf{p}}$

List of Abbreviations

CDF cumulative distribution function. 61

CVaR conditional value-at-risk. 32, 38

DRO distributionally robust optimization. 3, 16, 21, 43–48, 52, 57–60, 65–71, 74–80

K-L Kullback-Leibler. 7, 11, 16, 21, 52–54, 58–60, 62, 65–68, 70, 71, 73–76

K-S Kolmogorov-Smirnov. 9, 16, 61, 62, 65–67, 69, 71, 75, 77–79

MOO multi-objective optimization. 15, 16, 21–23, 35, 37–39, 42, 45, 48, 58–60, 73

QoI quantity of interest. 15, 19, 24, 28–30, 33, 58, 75

SAA sample average approximation. 17, 21–26, 32, 35, 40, 44, 46, 59–61, 63, 69, 73, 76

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

In designing a system to achieve optimal performance, the designer typically evaluates system performance using some quantity of interest (QoI) (e.g., cost, efficiency, weight). The QoI depends on design variables, which the designer wishes to select optimally. Typical design variables include geometric quantities (e.g., span, sweep angle for an aircraft wing), material properties (e.g., material stiffness, density for an aircraft wing spar), or operational decisions (e.g., nominal altitude for an aircraft). The QoI will also depend on other parameters which are out of the designers control. Furthermore, in any real-world system the designer may have imperfect knowledge of their true values, rendering these parameters uncertain. Typical examples of such uncertain parameters in the design of an aircraft are windspeed (and thus operational Mach number), or cargo weight.

Computational modeling of the system of interest provides the designer a means to cheaply and rapidly explore how changes in the underlying parameters affect the output quantity of interest, and thus supports more informed design decisions. We consider the setting in which the designer has access to a high-fidelity computational model of the system of interest. The model takes, as input, values for the design variables and the uncertain parameters, and returns an accurate evaluation of the QoI. In practice, such computational models are often complex, proprietary, legacy, and/or written in poorly documented code that the designer may not have time to decipher. With these settings in mind, we treat the model as a black-box. This means we assume no knowledge of the model structure or the behavior of the output variable

and its functional relationship to input variables. This makes the design methodology discussed herein general in the sense that it can be applied to any system, provided the designer has a computational model with input and output as described above.

The field of design optimization under uncertainty concerns the specification, representation, and treatment of uncertain parameters during a design process. The importance of considering the nature of the uncertain parameters in engineering design has long been established,[2–4] the primary reason being that methods that simply consider all parameters to be deterministic tend to over-fit designs to the chosen parameter values. This often results in degraded performance of the system when it is exposed to the true uncertain operating environment.[5,6]

One decision a designer must make is in how to characterize and represent the uncertainty. The most prevalent approach is to treat the uncertain parameters as random variables, with each parameter being assigned a probability distribution describing how it varies.[2,5] Other treatments and representations of uncertainty have also been established,[7] such as interval or set-based uncertainty,[8–10] and possibility theory.[11] In this work we focus on the case of probabilistic uncertainties, where the designer must specify a probability distribution for the uncertain parameters.

In practice the designer will never have perfect knowledge of how the underlying system parameters vary, and thus they will not have access to the exact probability distribution governing an uncertain parameter. Consequently, the designer will be forced to assume a probability distribution, or to estimate it from known data, processes that will always be prone to error.

Fortunately, the designer usually has some incomplete knowledge about the uncertainty, which can help mitigate this error. In this case we say the designer has *partial observability* over the uncertainty. One situation in which partially observable uncertainty commonly arises is the case when the designer has no knowledge about the distribution of uncertain parameters itself, but instead has access to a sample of parameter values drawn from this distribution. This is the case, e.g., when the designer takes discrete measurements of the uncertain parameters. In this work we will consider this scenario, denoting by m the size of the sample (number of measurements).

Of particular interest is the case where the sample size, m, is low. This represents the situation where only limited data are available to characterize the probability distribution.

Recently, a body of work has emerged from within the optimization community studying an approach that explicitly accounts for the fact that one is never able to exactly specify a probability distribution in practice.[12–18] This approach, called distributionally robust optimization (DRO), weakens the assumption of specifying a single probability distribution for the uncertain parameters. Instead, it only requires the designer to select a *set* of possible probability distributions—the optimization considers all distributions within this set.

In this work we explore how imperfect knowledge of uncertainties affects the performance of traditional optimization under uncertainty methods. In particular, we show that these methods often produce designs that perform poorly when realized under the true distribution of uncertainty. We then present a formulation of the design optimization as a DRO problem. This formulation explicitly seeks designs that are robust to deviations in the distribution of the uncertainty. We present computationally efficient algorithms for solving the DRO problem, when the set of distributions is constructed using either the L_2-norm or the Kullback-Leibler (K-L) divergence. We apply this formulation to a model design problem and show that it outperforms traditional design optimization under uncertainty methods when the uncertainty is only partially observable.

The outline of this thesis is as follows. Chapter 2 formulates the problem of design optimization using only m realizations of the uncertain parameters and presents two traditional methods for solving this problem using sample average approximation (SAA), and multi-objective optimization (MOO). A model design problem is also presented, which serves as a motivating example to illustrate the challenges associated with design under partially observable uncertainty. Chapter 3 presents a formulation of the design problem as a DRO problem, and presents algorithms for solving the resulting optimization problems. Chapter 4 presents the results of the DRO approach applied to the model design problem. These results are used to an-

alyze the performance of the approach, and compare it to both the SAA and MOO approaches. Finally, Chapter 5 concludes the thesis and suggests possible areas to explore in future work.

CHAPTER 2

DESIGN OPTIMIZATION UNDER PARTIALLY OBSERVABLE UNCERTAINTY

In this chapter we formulate a design optimization under uncertainty problem adapted to the setting in which the designer only has access to a sample of realizations of the uncertain parameters. In Section 2.1 we introduce a methodology for optimizing the mean performance of designs using sample average approximation (SAA). In Section 2.2 we introduce a model design problem, to which we apply the SAA approach and illustrate the consequences of partial observability. Finally, Section 2.3 discusses multi-objective optimization (MOO), and explores how this widely used approach to optimization under uncertainty performs in the partial observability setting.

2.1 Direct optimization using sample average approximation

We denote the vector of design variables by $\mathbf{x} \in \mathcal{X}$, where $\mathcal{X}$ is the feasible design space encoding constraints on the design. To represent the uncertain parameters, we define a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\mathbf{u}})$ with sample space Ω, σ-algebra $\mathcal{F}$, and probability measure $\mathbb{P}_{\mathbf{u}}$. We define the random variable $\mathbf{u} : \Omega \to \mathcal{U}$, which has dimension equal to the number of uncertain parameters in the design problem. The uncertainty space, $\mathcal{U}$, represents the space of all possible realizations of the random variable, $\mathbf{u}$. We use subscripts to denote particular realizations of the uncertain parameters, i.e., $\mathbf{u}_i =$

$\mathbf{u}(\omega_i)$ for some $\omega_i \in \Omega$. We represent the QoI of our system as a function, $Q(\mathbf{x}, \mathbf{u})$, of the design variables $\mathbf{x}$, and the uncertain variables, given by the random variable $\mathbf{u}$. As the QoI, $Q(\mathbf{x}, \mathbf{u})$, is a function of the random variable $\mathbf{u}$, it too is a random variable, with probability measure denoted $\mathbb{P}_{\mathbf{Q}}$. Thus, the performance of a design $\mathbf{x}$ is completely characterized by the distribution $\mathbb{P}_{\mathbf{Q}}$. Note that the randomness in the QoI is induced only by the randomness in $\mathbf{u}$. In a slight abuse of notation, we will use $Q(\mathbf{x}, \mathbf{u}_i)$ to denote the deterministic value of the QoI evaluated at a realization of the uncertain parameters, $\mathbf{u}_i$. The problem of design optimization under uncertainty is thus to find a set of optimal design variables that produce a favorable distribution of performance, $\mathbb{P}_{\mathbf{Q}}$.

In this work we will be focus on optimizing the mean of the distribution in design performance. In this case, assuming that a lower QoI is favorable, the optimization problem we wish to solve is

$$\mathcal{P}: \qquad \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}_{\mathbf{u}}} \left[Q(\mathbf{x}, \mathbf{u}) \right]. \tag{2.1}$$

We denote the optimal objective value of $\mathcal{P}$ by Z_∞, and a corresponding minimizing design by $\mathbf{x}_\infty$. A key limitation in solving $\mathcal{P}$ is that it requires the designer to have exact knowledge of the probability distribution, $\mathbb{P}_{\mathbf{u}}$, governing the uncertain parameters. In practice the designer will not have complete knowledge of the exact distribution of the uncertainty. In the setting we consider, the designer has access to a sample of m realizations of the uncertain parameters, drawn from the distribution $\mathbb{P}_{\mathbf{u}}$. Consequently, the designer will be forced to use this sample to make an assumption about the true probability distribution, with unknown impacts on the resulting optimal design.

In this setting, a simple and commonly used approach is to use SAA. This amounts to approximating the true distribution $\mathbb{P}_{\mathbf{u}}$ using the empirical distribution $\hat{\mathbf{p}}$ associated with the sample. In this case $\hat{p}_1, \ldots, \hat{p}_m$ denote the likelihood that the uncertain parameters $\mathbf{u}$ are realized as $\mathbf{u}_1, \ldots, \mathbf{u}_m$ respectively. This approach is justified by the fact that as $m \to \infty$ the empirical distribution $\hat{\mathbf{p}}$ is will converge to the true dis-

tribution $\mathbb{P}_{\mathbf{u}}$ for most well-behaved distributions encountered in practice. For a given random sample $\mathbf{u}_1, ..., \mathbf{u}_m$, with associated empirical distribution $\hat{\mathbf{p}}$, we use SAA to write a finite sample analogue to $\mathcal{P}$:

$$\mathcal{S}: \qquad \min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}_{\hat{\mathbf{p}}}\left[Q(\mathbf{x},\mathbf{u})\right] = \min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{m} \hat{p}_i Q(\mathbf{x},\mathbf{u}_i). \tag{2.2}$$

We denote the optimal objective value of $\mathcal{S}$ by $\hat{Z}_m$, and the corresponding optimal design by $\mathbf{x}_m$. Note that $\hat{Z}_m$ is the expected performance of $\mathbf{x}_m$ under the *sample* distribution $\hat{\mathbf{p}}$. In practice, what we really care about is how the design $\mathbf{x}_m$ performs under the *true* distribution $\mathbb{P}_{\mathbf{u}}$. To this end, we define the expected performance of the design $\mathbf{x}_m$ under the true distribution of uncertainty, $\mathbb{P}_{\mathbf{u}}$, and denote this by

$$Z_m = \mathbb{E}_{\mathbb{P}_{\mathbf{u}}}\left[Q\left(\mathbf{x}_m,\mathbf{u}\right)\right]. \tag{2.3}$$

Note that $\hat{Z}_m$ is able to be calculated by the designer, as it requires only the sampled values of the uncertainty, whereas Z_m is not observable by the designer, since it requires knowledge of the true distribution $\mathbb{P}_{\mathbf{u}}$.

As $m \to \infty$ we expect $\mathcal{P}$ and $\mathcal{S}$ to be equivalent, so that $Z_m \to Z_\infty$ and $\mathbf{x}_m \to \mathbf{x}_\infty$. However, for finite m this approximation will be subject to sampling error. Consequently, the solution $\mathbf{x}_m$ obtained by the designer in the finite sample setting will differ from the true solution $\mathbf{x}_\infty$. Moreover, $\mathbf{x}_m$ and Z_m will vary between different draws of the random sample. If the empirical distribution $\hat{\mathbf{p}}$ generated by a particular sample draw is a good approximation of the true distribution $\mathbb{P}_{\mathbf{u}}$, then we can reasonably expect $\mathbf{x}_m$ to be close to the true optimal design $\mathbf{x}_\infty$. On the other hand, if the sample is not representative of the true distribution $\mathbb{P}_{\mathbf{u}}$, then $\mathbf{x}_m$ may differ greatly from the true solution $\mathbf{x}_\infty$, and the corresponding design performance, Z_m, may be poor.

The designer has no control or a priori knowledge of whether a particular sample draw is representative of the true distribution. We thus seek design methodologies that are robust to this sampling error, in the sense that over all possible sample

draws the methodology consistently results in designs that perform well under the true distribution. To this end, we will be primarily interested in the performance of designs averaged over all possible sample draws of size m. We will also be interested in the risk level of the methodology, as indicated by how designs perform in the worst-case sample draws.

2.2 Motivating example: the challenges of optimization under uncertainty

In Section 2.2.1 we introduce a motivating example: the problem of designing an acoustic horn for maximum efficiency subject to an uncertain operating condition. In Section 2.2.2 we apply the SAA approach outlined in Section 2.1 to the horn design problem, and investigate how partial observability affects the performance of the resulting designs.

2.2.1 Model problem formulation

In this section, we introduce an example design problem that will be used throughout this thesis to explore different design methodologies, and illustrate various aspects of their performance. We suppose that we are tasked with optimizing the design of an acoustic horn in order to minimize the amount of internal reflection present, and thus maximize the efficiency of the horn. In order to analyze how the design of the horn affects the amount of internal reflection, we utilize a computational model. A brief summary of the model is presented in Figure 2-1 below and the description that follows. For more details on the theory behind the acoustic horn and the computational model used, we refer the reader to References 19,20 and 21.

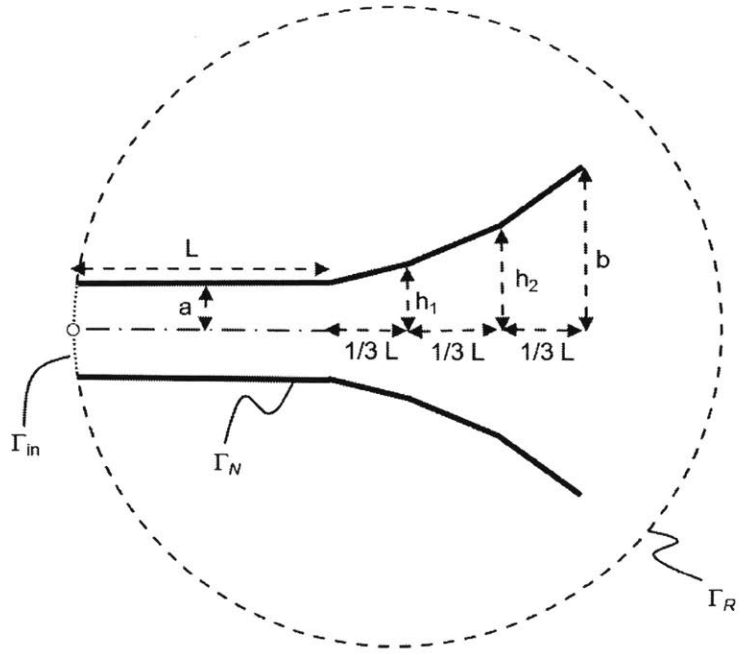


Figure 2-1: *The geometry of the acoustic horn model (Adapted from Ng et al.[1]). The design variables used in this thesis are h_1 and h_2, while the remaining parameters shown are considered fixed.*

The exterior domain is truncated by a circular absorbing boundary of radius $R = 25$. The non-dimensional horn geometry is axisymmetric and is parameterized by five variables. We consider three of these to be fixed parameters: the horn inlet length $L = 5$, and half-width $a = 0.5$, and the outlet half-width $b = 3$. The remaining two variables are design variables h_1 and h_2, corresponding to the half-widths at two uniformly spaced points in the horn flare (see Figure 2-1). Both design variables h_1 and h_2 are constrained to lie within the interval $[a, b] = [0.5, 3]$. The governing equation is the non-dimensional Helmholtz equation,

$$\nabla^2 v + k^2 v = 0, \tag{2.4}$$

where v is the non-dimensionalized pressure, and k is the wave-number, which we treat as the uncertain operating condition of the horn. The boundary conditions on

the horn inlet Γ_{in}, horn surface Γ_N, and far field boundary Γ_R are given by

$$\Gamma_{in}: \quad \mathrm{i}kv + \tfrac{\partial v}{\partial n} = 2\mathrm{i}k, \tag{2.5}$$

$$\Gamma_N: \quad \tfrac{\partial v}{\partial n} = 0, \tag{2.6}$$

$$\Gamma_R: \quad \tfrac{\partial v}{\partial n} - \left(\mathrm{i}k - \tfrac{1}{2R} + \tfrac{1}{8R(1-\mathrm{i}kR)}\right) v = 0, \tag{2.7}$$

where n is a unit vector normal to the corresponding boundary, and $\mathrm{i} = \sqrt{-1}$. The governing equation is solved to compute v using a reduced basis finite element model with $n = 116$ basis vectors. The QoI for the acoustic horn model is the reflection coefficient:

$$s = \left| \int_{\Gamma_{\text{in}}} v \, d\Gamma - 1 \right|, \tag{2.8}$$

which is a fractional measure describing how much of an incoming wave is internally reflected in the horn, as opposed to being transmitted out into the environment. It is thus considered a measure of the horn efficiency, with a lower reflection coefficient giving more favorable performance. Thus, framing the acoustic horn design problem using the notation introduced in Section 2.1, we have:

$$\text{Design variables:} \quad \mathbf{x} = [\, h_1,\ h_2 \,]^\top, \tag{2.9}$$

$$\text{Design space:} \quad \mathcal{X} = [\, 0.5\ ,\ 3 \,] \times [\, 0.5\ ,\ 3 \,], \tag{2.10}$$

$$\text{Uncertain parameters:} \quad \mathbf{u} = k, \tag{2.11}$$

$$\text{Uncertainty space:} \quad \mathcal{U} = [1.3, 1.5], \tag{2.12}$$

$$\text{Output QoI} \quad Q(\mathbf{x}, \mathbf{u}) = s. \tag{2.13}$$

We perform an exploratory analysis of the acoustic horn design space to show how the uncertain wave number affects the reflection coefficient for different horn designs. For this study, we suppose that the wave number follows a fixed truth probability distribution

$$\mathbf{u} \sim \mathbb{P}_u = \text{Uniform}(1.3, 1.5). \tag{2.14}$$

The design points explored comprise a 50×50 point, uniformly spaced grid in the design space $\mathcal{X}$. For each design, the mean and variance of the QoI under the uniformly distributed uncertainty is computed using 10-point Gaussian quadrature. Figures 2-2 and 2-3 show contour plots of the mean and standard deviation of the reflection coefficient over the design space $\mathcal{X}$.

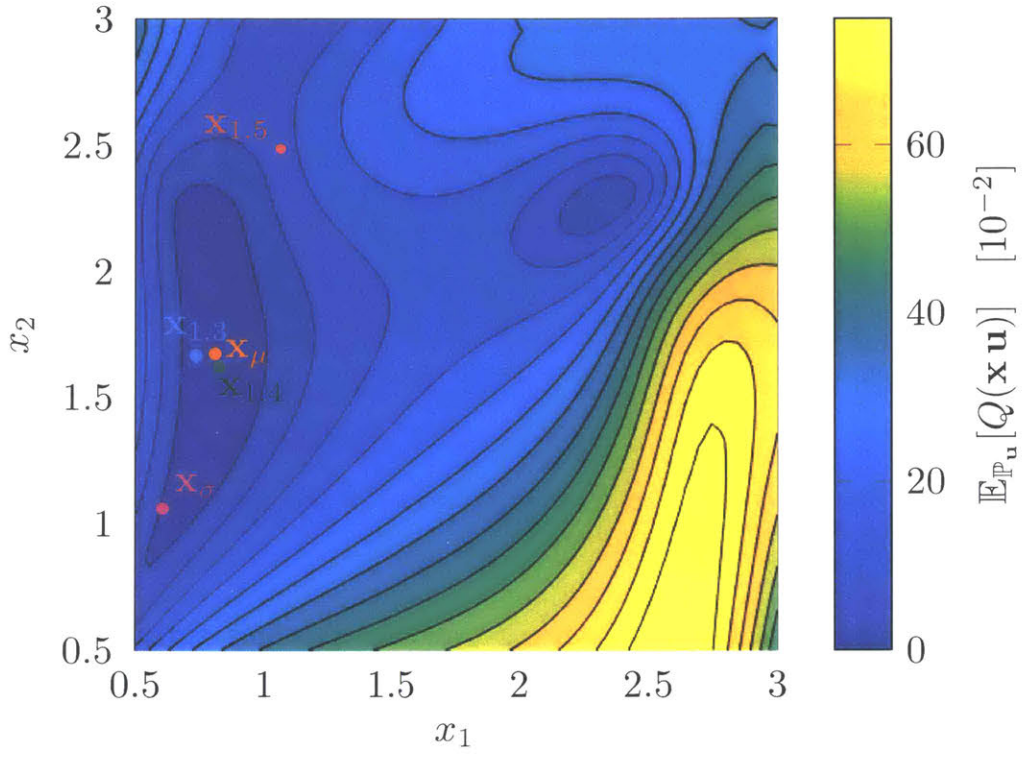


Figure 2-2: *Mean value of the QoI over the acoustic horn design space, for a uniformly distributed wave number.*

We see that the reflection coefficient has a non-linear dependence on the design variables. Horn designs in which $x_1 > x_2$ result in poor mean performance. This is expected since such designs feature a concave horn flare, resulting in a high degree of internal reflection. We denote the design that achieves optimal (minimal) mean reflection coefficient $\mathbf{x}_\mu$, and the design that achieves minimal standard deviation in the reflection coefficient by $\mathbf{x}_\sigma$. The fact that these designs are distinct suggests that it is possible to trade-off between improving the mean performance of the horn design, and reducing the degree of variability in this performance. This type of mean-standard deviation tradeoff will be explored further in Section 2.3.

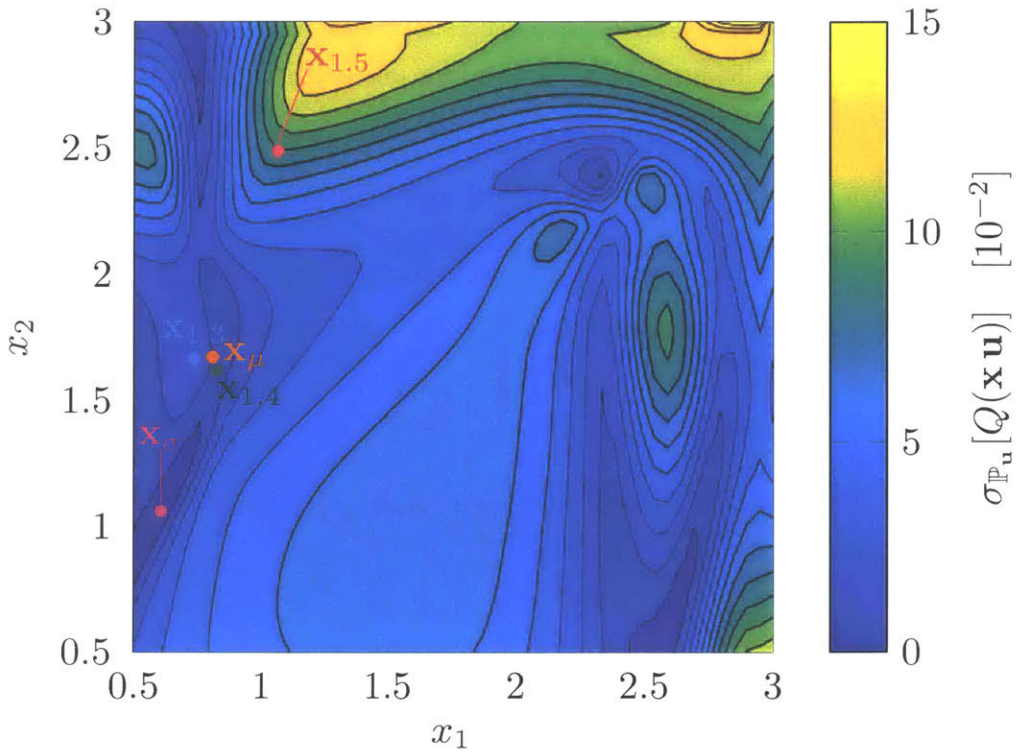


Figure 2-3: *Standard deviation in the QoI over the acoustic horn design space, for a uniformly distributed wave number.*

In addition to these designs, Figures 2-2 and 2-3 also show three point-wise optimal designs. These are the designs that achieve optimal performance at point-wise values of the uncertain parameter, i.e. wave numbers of $\mathbf{u} = k = 1.3$, 1.4, and 1.5, respectively. Figure 2-4 shows the horn flare geometry of each of the aforementioned designs, and the corresponding design performance—given by the QoI—over the range of the uncertain variable.

We see that designs $\mathbf{x}_{1.3}$, $\mathbf{x}_{1.4}$ and $\mathbf{x}_{1.5}$ all exhibit low reflection at the corresponding design wave numbers, but also exhibit reduced off-design performance. We also see from the performance of $\mathbf{x}_{\sigma}$ that it is possible to obtain a horn design with consistent performance over the range of wave numbers, however, this comes with the downside that QoI is relatively high everywhere. In fact, the performance of $\mathbf{x}_{\mu}$ ends up being superior to $\mathbf{x}_{\sigma}$ at almost all wave numbers. In this case we say that $\mathbf{x}_{\mu}$ *stochastically dominates* $\mathbf{x}_{\sigma}$, since it performs better at every point in uncertainty space.

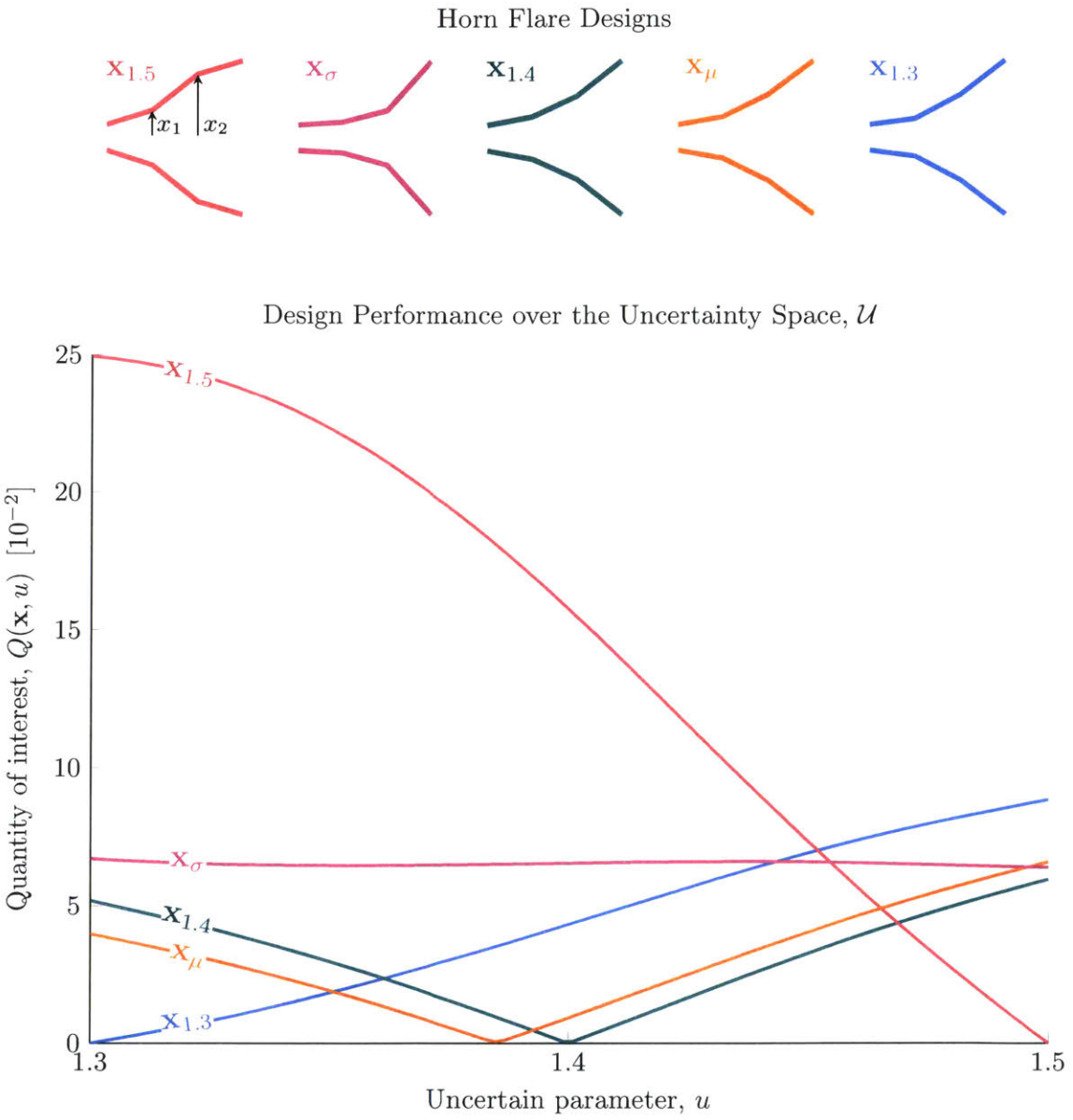


Figure 2-4: *Performance of designs that exhibit optimal mean performance ($\mathbf{x}_\mu$), minimal variation in performance ($\mathbf{x}_\sigma$), or optimal performance at a particular value of the uncertain variable ($\mathbf{x}_{1.3}$, $\mathbf{x}_{1.4}$, $\mathbf{x}_{1.5}$).*

2.2.2 The consequences of partial observability

In this section we apply the SAA approach to the acoustic horn design problem. We simulate T sample draws, each consisting of m random realizations from $\mathbb{P}_{\mathbf{u}}$. For each draw $t = 1, ..., T$, we solve $\mathcal{S}$ using an interior point method[a] to obtain $\mathbf{x}_m$. We then evaluate Z_m, the performance of $\mathbf{x}_m$ under the true distribution $\mathbb{P}_{\mathbf{u}}$, by solving Eqn. 2.3 using 10-point Gaussian quadrature.

In order to evaluate how the methodology performs across all sample draws, we compute the average of Z_m over the T sample draws, denoted by

$$\mu(Z_m) = \mathbb{E}\left[\{Z_m^1, ..., Z_m^T\}\right] = \frac{1}{T}\sum_{t=1}^{T} Z_m^t, \tag{2.15}$$

where Z_m^t denotes the value of Z_m in sample draw t. We evaluate the risk level of the methodology using the notion of conditional value-at-risk (CVaR) (sometimes referred to as expected shortfall (ES) or expected tail loss (ETL)). In our setting, the CVaR at level 10% (CVaR_{10}) is defined as the expected value of the worst 10% of all sample draws. We approximate the true CVaR_{10} using T sample draws. Assuming we have ordered the sample indices so that $Z_m^1, ..., Z_m^T$ is in ascending order, we can define the approximate CVaR_{10} of Z_m, denoted $\phi(Z_m)$, as

$$\phi(Z_m) = \mathbb{E}\left[\{Z_m^t \mid t > 0.9T\}\right]. \tag{2.16}$$

For the sake of comparison, we also compute $\mathbf{x}_\infty$ and Z_∞. Recall that these are the solutions to the full optimization problem $\mathcal{P}$ (Eqn. 2.1), which requires knowledge of the true distribution $\mathbb{P}_{\mathbf{u}}$. This problem is again solved using 10-point quadrature, this time using the empirical distribution support and probabilities as quadrature nodes and weights respectively. We simulate $T = 500$ realizations of random samples. For each random sample we compute an optimized design, and evaluate the mean design performance under the true distribution Z_m, as described above. Figure 2-5 gives a histogram of Z_m values across the 500 sample draws, for sample sizes $m = 5, 10, 20$.

[a] As implemented in the function `fmincon`, included in the MATLAB Optimization Toolbox[22]

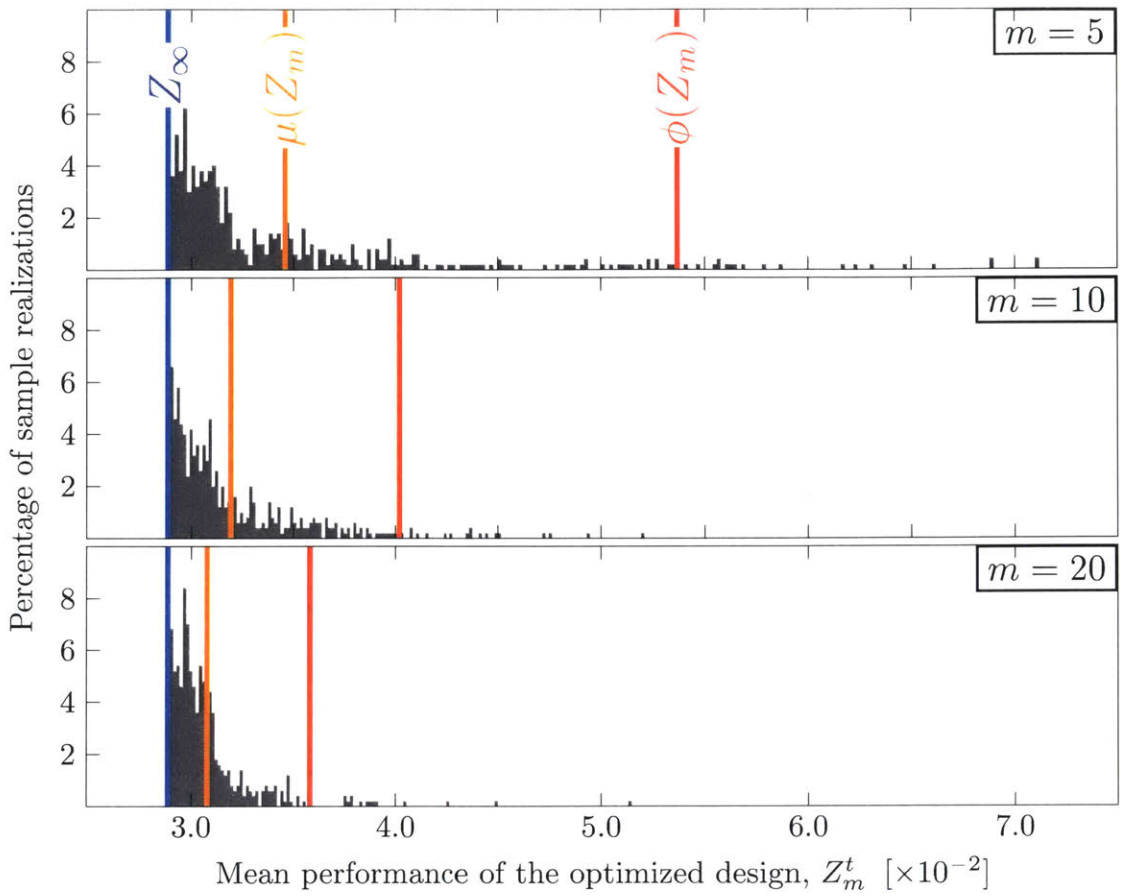


Figure 2-5: *Histograms showing the mean performance of designs computed from $T = 500$ sample draws of size m. The optimal, mean, and $CVaR_{10}$ values over all sample draws are indicated.*

We see that for some sample draws, Z_m is close to the true optimal objective $Z_\infty = 0.0289$. However, the long right tails of the histograms indicate that there are also many sample draws for which Z_m is far from Z_∞. This is especially the case as we decrease the sample size m. In the $m = 5$ case, the CVaR_{10} is $\phi = 0.0550$. This indicates that, in the worst 10% of sample draws, the QoI is over 90% higher (worse) than the true optimum. This drop in performance is the consequence of designing with only partial observability over the distribution of uncertainty.

In order to see how partial observability can lead to such a fall in performance, we investigate how it is manifested in the resulting horn designs. Figure 2-6 displays the performance of the best and worst performing horn designs obtained over all the sample draws of size $m = 5$.

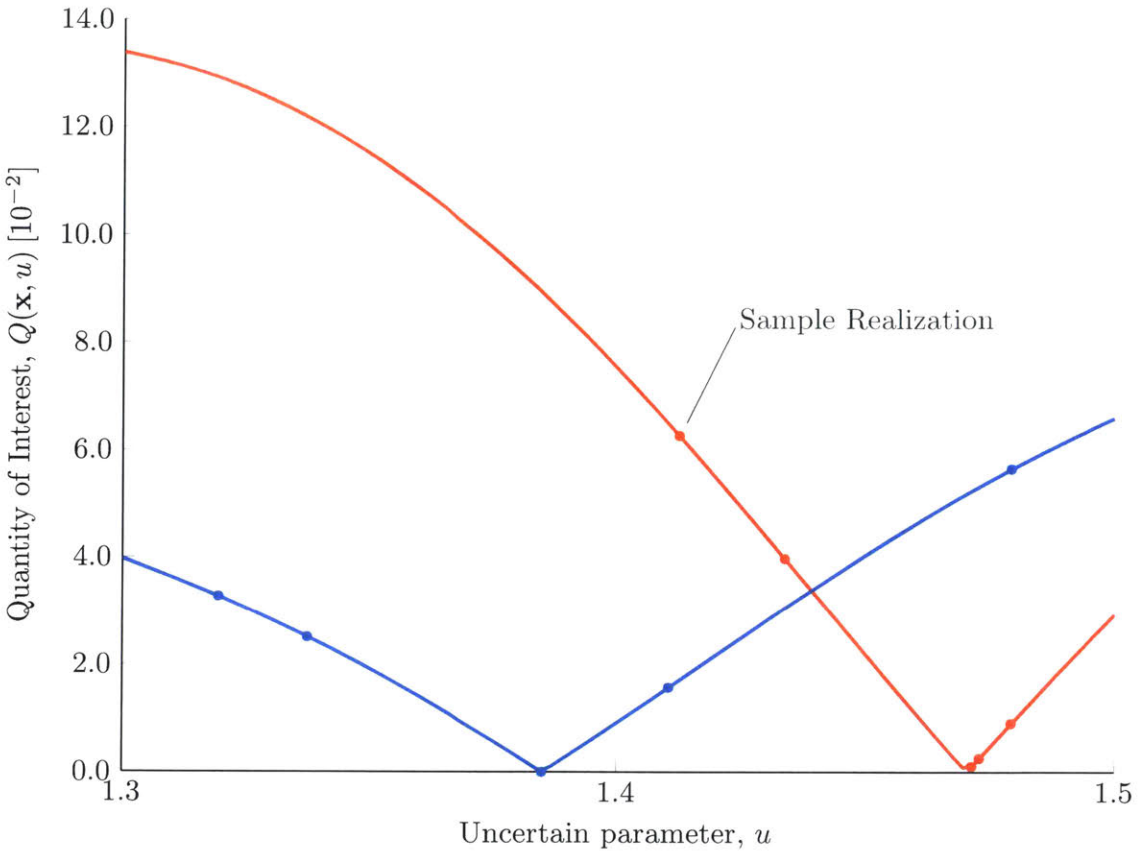


Figure 2-6: *Performance of two designs over the uncertainty space $\mathcal{U}$. These are the designs that exhibit the best and worst mean performance over all designs computed using different sample draws of $m = 5$ realizations of the random variables.*

We see that in the worst sample draw the realizations are clustered around one end of the range of the uncertainty. As a result, the empirical distribution associated with this sample is a poor estimate of the true uniform distribution. The design computed using this sample is over-fitted to the empirical distribution, ultimately leading to poor performance under the true distribution. In contrast, the best sample draw has an empirical distribution that more accurately reflects the underlying uniform distribution. These results show how using the SAA approach and neglecting to acknowledge that we are dealing with partial observability often produces designs that are over-fitted to the sample data.

2.3 Robustness through multi-objective optimization

Directly optimizing for mean performance under the empirical distribution using SAA is one widely-used approach to optimization under uncertainty. A common approach to introducing robustness is to augment the SAA objective by adding a weighted penalty on the variability in the performance of the design. This amounts to casting the problem as a multi-objective optimization (MOO) problem, in which the designer optimizes for mean performance across the sample, while simultaneously minimizing the variability in performance within the sample. This section illustrates the potential benefits of this approach in the partially observable uncertainty setting.

Reducing the variability in performance aims to prevent the optimization from over-fitting to the given sample, thus leading to better performance under the true distribution. To see this, consider the extreme case of a design with perfectly consistent performance over the entire range of the uncertainty. Such a design will exhibit identical performance under both the sample distribution and the true distribution. However, minimizing the variability in performance alone has the downside that it often leads to bad mean performance, as was illustrated in the exploratory design analysis in Section 2.2.1.

In order to optimize for both mean and variability in performance, we introduce the trade-off parameter λ, which governs the relative importance placed on the mean objective versus the standard deviation objective. Note that in the case $\lambda = 0$, the designer optimizes for the mean performance under the empirical distribution, and thus the MOO problem reduces to the SAA problem, $\mathcal{S}$ (Eqn. 2.2). The MOO problem for jointly optimizing the mean and standard deviation in performance over a sample of the uncertain parameters, $\mathbf{u}_1, ..., \mathbf{u}_m$, with associated empirical distribution $\hat{\mathbf{p}}$, can be written as

$$\mathcal{M}(\lambda): \quad \min_{\mathbf{x}\in\mathcal{X}} \quad (1-\lambda)\underbrace{\mathbb{E}_{\hat{\mathbf{p}}}\left[Q(\mathbf{x},\mathbf{u})\right]}_{\text{Mean}} + \lambda\underbrace{\sigma_{\hat{\mathbf{p}}}\left[Q(\mathbf{x},\mathbf{u})\right]}_{\text{Std. Dev.}}, \tag{2.17}$$

where

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbf{p}}}\left[Q(\mathbf{x},\mathbf{u})\right] &= \sum_{i=1}^{m} \hat{p}_i Q(\mathbf{x},\mathbf{u}_i), \\
\sigma_{\hat{\mathbf{p}}}\left[Q(\mathbf{x},\mathbf{u})\right] &= \sqrt{\sum_{i=1}^{m} \hat{p}_i (Q(\mathbf{x},\mathbf{u}_i) - \mathbb{E}_{\hat{\mathbf{p}}}\left[Q(\mathbf{x},\mathbf{u})\right])^2}, \\
\lambda &\in [0,1].
\end{aligned}$$

We solve $\mathcal{M}(\lambda)$ using an interior point method[b], and denote the resulting optimal design by $\mathbf{x}_m$, where the value of λ used will be clear from the context or explicitly denoted using the notation $\mathbf{x}_m|_\lambda$. In this section, we suppose that the fixed truth distribution of the uncertainty is Uniform[1.3, 1.5]. We are interested in how the optimal design, $\mathbf{x}_m$, performs under this true distribution $\mathbb{P}_\mathbf{u}$, rather than the sample distribution $\hat{\mathbf{p}}$. To this end, we define the mean and standard deviation in the performance of the design $\mathbf{x}_m$ under the true distribution, and denote these by

$$Z_m = \mathbb{E}_{\mathbb{P}_\mathbf{u}}\left[Q\left(\mathbf{x}_m,\mathbf{u}\right)\right], \tag{2.18}$$

$$S_m = \sigma_{\mathbb{P}_\mathbf{u}}\left[Q\left(\mathbf{x}_m,\mathbf{u}\right)\right]. \tag{2.19}$$

As in Section 2.2.2, we solve these equations using 10-point Gaussian quadrature. For comparison, we also define the true optimal designs that the designer is only able to obtain if they have full observability over the uncertainty. These are denoted by $\mathbf{x}_\infty|_\lambda$, and are solutions to $\mathcal{M}(\lambda)$, with the empirical distribution $\hat{\mathbf{p}}$ replaced by the true distribution $\mathbb{P}_\mathbf{u}$. The mean and standard deviation in performance of these designs are denoted by $Z_\infty|_\lambda$ and $S_\infty|_\lambda$, and are computed using Eqn. 2.18 and Eqn. 2.19, respectively. Again, the notation "$|_\lambda$" will often be dropped when the value of λ used is clear from the context.

[b] As implemented in the function `fmincon`, included in the MATLAB Optimization Toolbox[22]

The canonical result sought in MOO is the trade-off curve between the two objectives, parameterized by the trade-off parameter λ. In the full observability case this curve is termed the *Pareto Frontier*, and describes the performance of the set of designs that achieve an optimal affine combination of Z_∞ and S_∞. In the partial observability case, the designer is unable to generate the true Pareto Frontier, as they do no not have access to the true distribution $\mathbb{P}_{\mathbf{u}}$. Instead, for each random sample, we can generate a set of designs that achieve an optimal affine combination of sample mean and sample standard deviation, i.e., a set of solutions to $\mathcal{M}(\lambda)$ for $\lambda \in [0, 1]$. Note that being an optimizer of $\mathcal{M}$ does not necessarily guarantee good mean and variance in performance under the true distribution, which are given by Z_m and S_m respectively.

To illustrate this, we solve $\mathcal{M}(\lambda)$ repeatedly, $T = 1000, 600, 200$ times, for different samples of size $m = 5, 10, 20$ respectively. For each resulting optimal design, $\mathbf{x}_m$, we evaluate the mean, Z_m, and standard deviation, S_m, in performance under the true distribution. We repeat this for 20 uniformly spaced values of $\lambda \in [0, 0.95]$. Figure 2-7 shows the mean versus standard deviation trade-off curves for each sample size, with the full observability Pareto Frontier included as a baseline. Recall that the $\mu(\cdot)$ operator computes the average of these values over all of the T realizations (see Eqn. 2.15).

The first thing to note is that the mean-variability curves for a finite sample size, m, do not coincide with the Pareto frontier. As the sample size is decreased, the trade-off curve moves away from the optimal Pareto Frontier, becoming sub-optimal. The shape of the curve also changes as we decrease the sample size. For small m, we see that optimizing for the sample mean alone no longer gives the best mean performance under the true distribution. Instead, shifting some of the objective weight onto the standard deviation objective actually results in an improvement in the true mean performance.

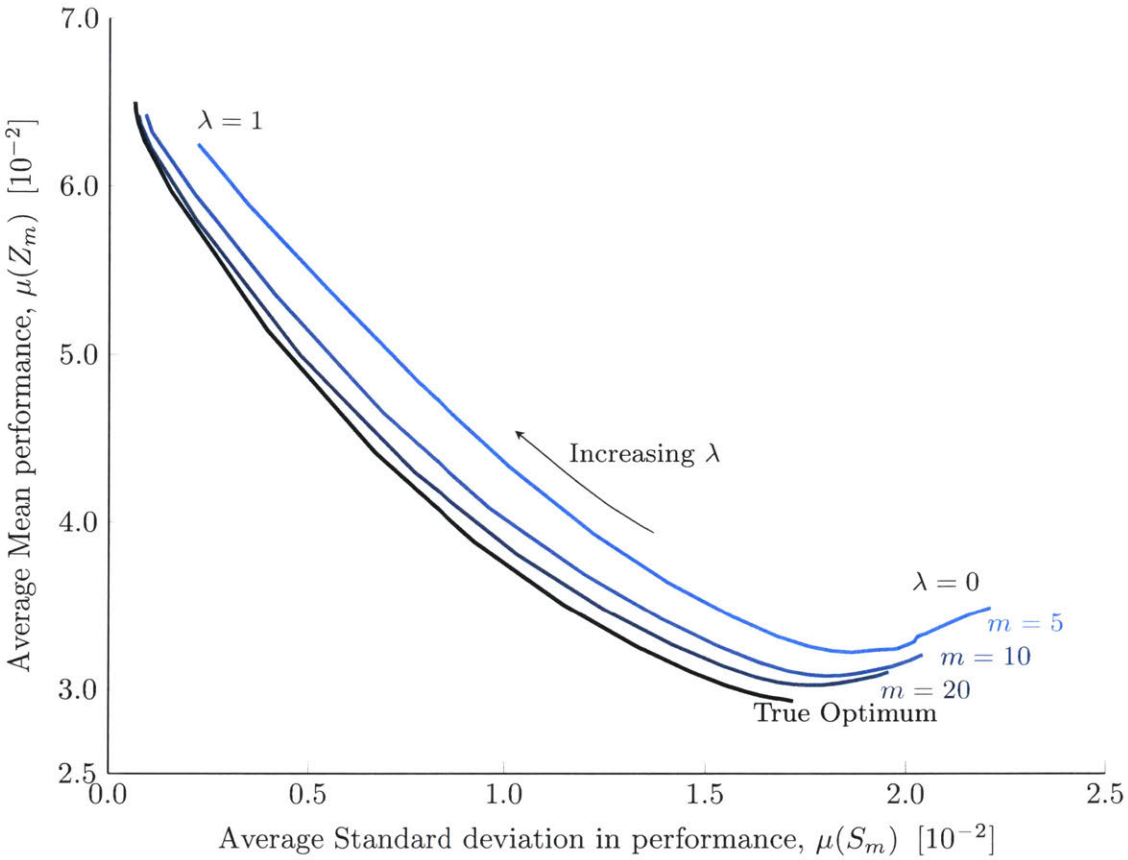


Figure 2-7: *Mean versus standard deviation curves generated using MOO, with varying sample size, for the acoustic horn design problem. Each curve corresponds to a sample size* $m = 5, 10, 20$*, and is computed by averaging over* $T = 1000, 600, 200$ *samples, respectively. Each point on a curve corresponds to a particular value of* $\lambda \in [\,0,\, 1\,]$.

In addition to the average performance over all realizations, we are also interested in how the performance of designs varies between different sample realizations. To investigate this, we analyze the trade-off between average mean performance over all realizations, $\mu(Z_m)$, and the CVaR_{10} in mean performance between realizations, $\phi(Z_m)$ (see Eqn. 2.16). These mean-risk trade-off curves are shown in Figure 2-8.

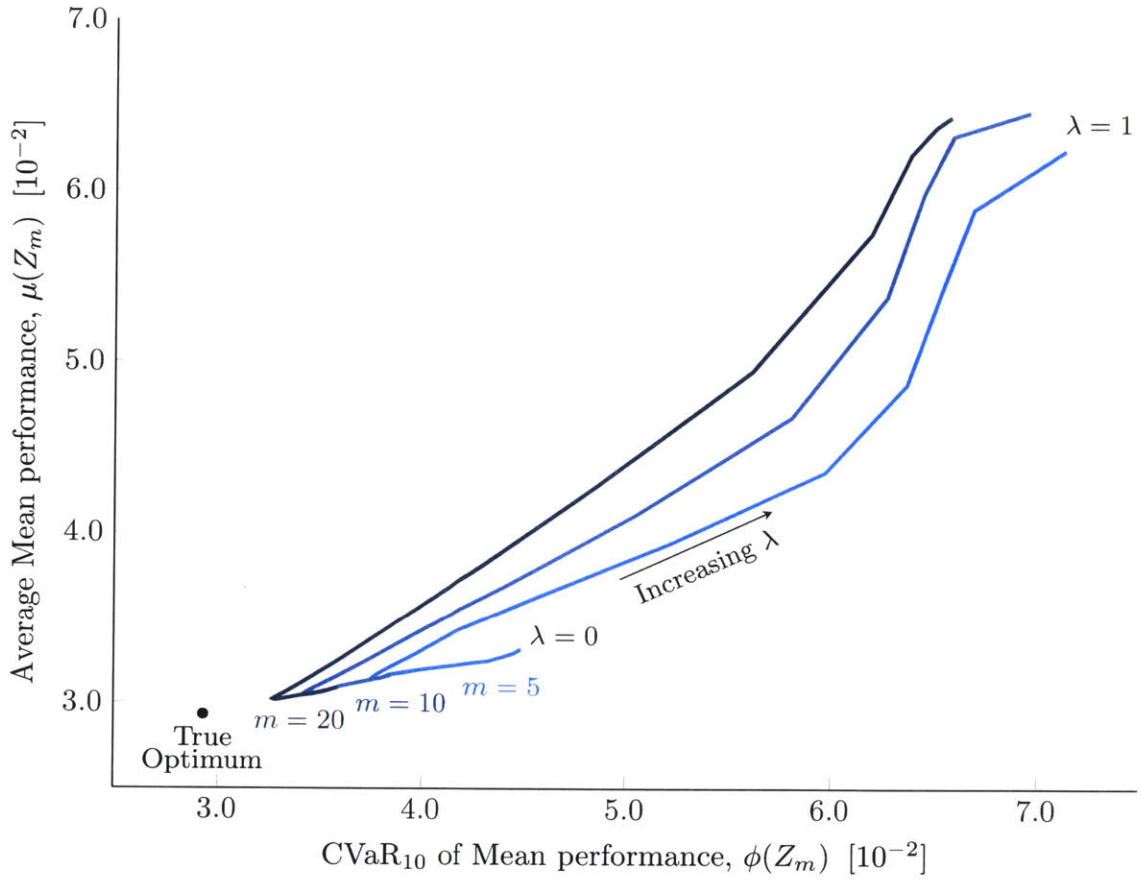


Figure 2-8: *Mean versus $CVaR_{10}$ curves generated using MOO, with varying sample size, for the acoustic horn design problem. Each curve corresponds to a sample size $m = 5, 10, 20$, and is computed by averaging over $T = 1000, 600, 200$ samples, respectively. Each point on a curve corresponds to a particular value of $\lambda \in [0, 1]$.*

We see that increasing λ from zero (i.e., adding some importance to the variability objective) improves both the average and CVaR_{10} in performance over repeated trials. In particular, we see that if the optimal value of λ is chosen for a sample size $m = 5$, the average performance of the resulting designs is better than a sample of size $m = 10$ with $\lambda = 0$. Similarly, samples of size $m = 10$ with the optimal λ parameter outperform samples of size $m = 20$ with $\lambda = 0$. This suggests that introducing a variance reduction objective and optimally selecting the trade-off parameter λ can result in an improvement in the mean performance of designs, in terms of both the average and CVaR_{10} over many sample realizations.

However, increasing λ past a critical value leads to a sharp reduction in the average and CVaR_{10} of mean performance over many sample realizations. To understand

this trend more concretely, we revisit the sample realizations studied in Figure 2-6. Recall that these so-called best and worst realizations produced the best and worst performing designs (respectively) using the SAA approach. Figure 2-9 shows the performance of horn designs found by solving $\mathcal{M}$ for $\lambda = 0, 0.5$, and 0.95, using the worst case sample of size $m = 5$.

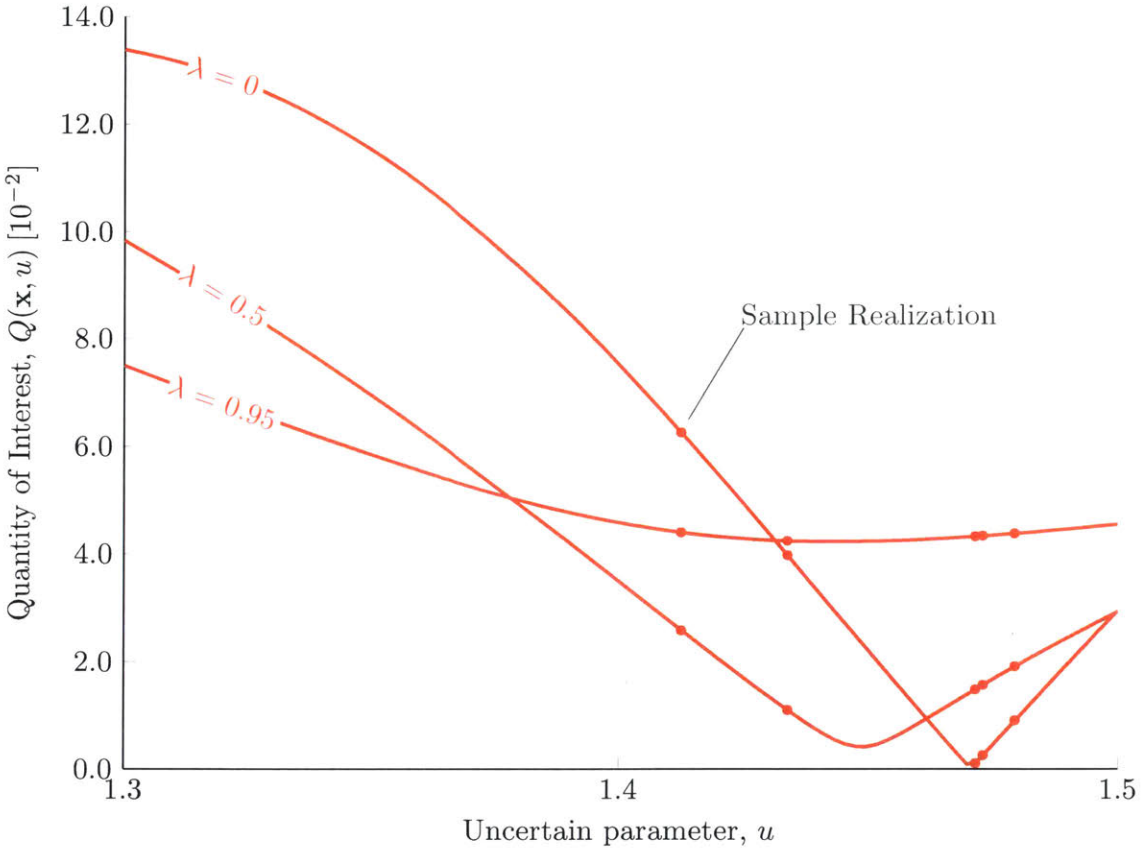


Figure 2-9: *Performance over the uncertainty space, $\mathcal{U}$, of designs computed by solving $\mathcal{M}(\lambda)$, with $\lambda = 0, 0.5, 0.95$ and the worst-case sample of size $m = 5$ realizations.*

Recall that in the $\lambda = 0$ case, $\mathcal{M}$ reduces to $\mathcal{S}$ and the designer optimizes for mean performance at the realized uncertainty values. The resulting design in this case exhibits good performance in the sampled region of the uncertainty space, but poor performance in the region where the uncertain parameter was not realized in the sample (in this case near $u = 1.3$). In this way, the design has been over-fit to the sample. Adding weight to the variance reduction objective gives the $\lambda = 0.5$ de-

sign. This design exhibits worse performance at some realized values of the uncertain parameter, but in return exhibits more consistent performance over the uncertainty space. The net effect is that this design has improved mean performance under the *true* distribution of uncertainty, despite having worse mean performance under the *sample* distribution. However, we see that as λ is increased further to $\lambda = 0.95$, the mean performance under the true distribution suffers, as the variance reduction objective dominates and leads to consistent, but consistently poor performance.

Figure 2-10 gives a plot analogous to Figure 2-9, this time showing the performance of designs found using the *best* possible sample of size $m = 5$ studied in the previous section.

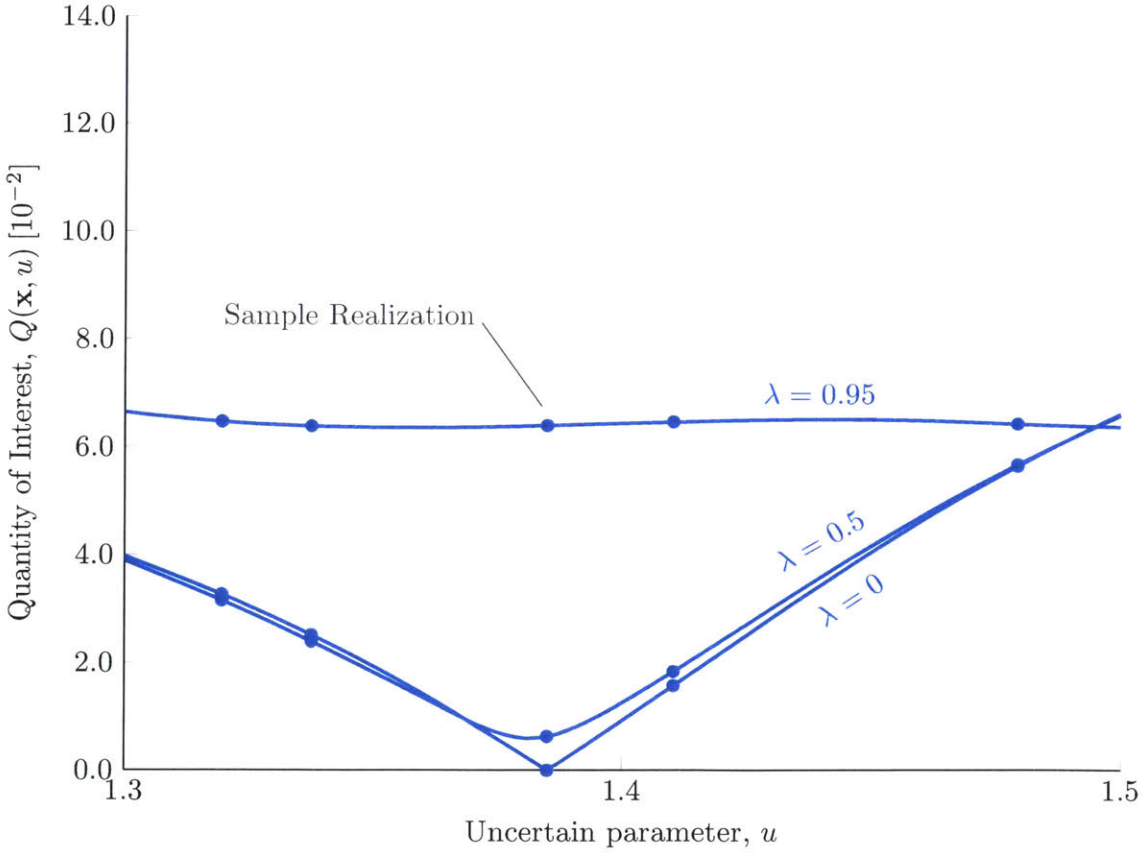


Figure 2-10: *Performance over the uncertainty space $\mathcal{U}$ of designs computed using $\lambda = 0, 0.5, 0.95$ and the best-case sample of size $m = 5$ realizations of the uncertain parameter.*

In this case, adding weight to the variance reduction objective by increasing λ always degrades design performance under the true distribution. However, this re-

duction is minor at moderate values of λ, e.g., in the $\lambda = 0.5$ case shown. This suggests that while the variance reduction objective may be beneficial to design performance when optimizing with a poor sample, it may also harm performance when a favorable sample is used.

From these results, we can conclude that adding a variance reduction objective makes the design methodology robust to poor realizations of the sample. This robustness is achieved by ensuring that the designs we select are able to generalize from the sample distribution to the true distribution. However, a criticism of the MOO approach is that the variance reduction objective does not directly robustify designs against partial information. Instead, it results in designs that are generally conservative, performing consistently at the cost of performing well on average. In particular, if the designer increases λ past a critical value the design will become *overly* conservative, leading to consistently poor design performance. Note that this critical value is not known to the designer a priori. Moreover, it depends on the particular realization of the sample used, so there is no way to guarantee that the value of λ used in the optimization does not exceed the critical value. In this way, the addition of the variance reduction objective can in fact be a threat to mean design performance.

In this section we have illustrated the importance of ensuring that designs are not over-fitted to the empirical distribution given by a random sample. This naturally raises the question of whether we can explicitly seek such designs, rather than achieving this feature as a secondary effect of variance reduction.

CHAPTER 3

DISTRIBUTIONALLY ROBUST DESIGN OPTIMIZATION

In this section we present a principled method for introducing the notion of robustness against variability in sample distributions into the design problem, using the mathematical framework of distributionally robust optimization (DRO). Section 3.1 formulates the problem of finding distributionally robust designs from a given sample of the uncertain parameters. These are designs that achieve good mean performance under the sample distribution, while requiring that this performance be robust to deviations in the distribution of the uncertainty. Section 3.2 discusses how to select a set of probability distributions to robustify the design against, and presents efficient algorithms for solving the resulting distributionally robust design problem.

3.1 Formulation

The central idea behind the distributionally robust approach is to optimize the design while considering a set of possible distributions of the uncertainty, rather than a single distribution.[13,23] The set of distributions that we consider in the optimization is termed the *ambiguity set*,[14] which we denote by $\mathcal{P}$. We seek a design that performs well for all distributions within the ambiguity set. This is achieved by solving a minimax problem to optimize the worst-case expected performance under any distribution within the ambiguity set. Using the notation introduced in Chapter 2.1,

the distributionally robust design optimization problem can be written

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbb{P}\in\mathcal{P}} \quad \mathbb{E}_{\mathbb{P}}\left[Q\left(\mathbf{x},\mathbf{u}\right)\right], \tag{3.1}$$

where $\mathbb{P} \in \mathcal{P}$ denotes probability distributions within the ambiguity set $\mathcal{P}$. In the setting we consider, the designer has access to a sample of m independent realizations of $\mathbf{u}$, randomly drawn from $\mathbb{P}_{\mathbf{u}}$ and denoted $\mathbf{u}_1, \ldots, \mathbf{u}_m$. This sample is used to compute the expectation in the DRO problem (Eqn. 3.1), giving the finite sample DRO problem

$$\mathcal{D}(\mathcal{P}): \qquad \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{p}\in\mathcal{P}} \quad \sum_{i=1}^{m} \hat{p}_i Q(\mathbf{x},\mathbf{u}_i). \tag{3.2}$$

We denote the optimal design found in $\mathcal{D}$ by $\mathbf{x}_m$, where the ambiguity set used in the optimization, $\mathcal{P}$, will be clear from the context, or explicitly denoted using the notation $\mathbf{x}_m|_{\mathcal{P}}$. Solving $\mathcal{D}$ does not require knowledge of the true distribution of uncertainty. Instead it requires the designer to specify the ambiguity set, which is assumed to contain the true distribution with high probability.

The inner maximization problem in $\mathcal{D}$ involves finding the worst-case expectation over all distributions in this ambiguity set. This inner maximization problem gives the worst-case distribution

$$\mathbf{p}^* = \operatorname*{argmax}_{\mathbf{p}\in\mathcal{P}} \mathbb{E}_{\mathbf{p}}\left[Q\left(\mathbf{x},\mathbf{u}\right)\right] = \operatorname*{argmax}_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{m} \hat{p}_i Q(\mathbf{x},\mathbf{u}_i). \tag{3.3}$$

Note that in the context we consider, the ambiguity set contains only discrete probability distributions, which we denote $\mathbf{p}$ (c.f. $\mathbb{P}$ for a general distribution). This ensures that computing the expectation under any distribution in the ambiguity set is tractable given only a black-box computational model. Once the worst case distribution $\mathbf{p}^*$ is found, the outer minimization problem finds the design that achieves the best possible improvement in the mean performance under this distribution. This outer problem now has the same form as the SAA problem, $\mathcal{S}$ (Eqn. 2.2), with the worst-case distribution $\mathbf{p}^*$ taking the place of the empirical distribution $\hat{\mathbf{p}}$, and can be solved using similar optimization methods. A noteworthy difference however, is

that the worst-case distribution within the ambiguity set may differ between designs. Thus if an iterative design optimization method is used, the inner optimization needs to be re-solved at each iteration. Furthermore, it is also worth noting that if the worst-case distribution $\mathbf{p}^*$ is allowed to have many zero entries—as is the case when an L_2-norm ambiguity set is used (see Section 3.2.1)—then the outer optimization problem can become poorly conditioned, and will consequently be sensitive to the initial condition. For the acoustic horn design problem, we are able to overcome this challenge by switching from an interior-point method to a trust-region reflective algorithm[a].

Finally, it is interesting to note that in contrast with the MOO problem, $\mathcal{M}$ (Eqn. 2.17), we are not explicitly optimizing for a reduction in the variation in performance over the uncertainty space. In Section 2.3 we showed that adding a variance reduction objective to the problem does make the resulting designs robust to changes in the distribution of uncertainty. However, we also showed that this can come at the cost of over-conservatism, with the resulting designs often exhibiting consistently poor performance. In the DRO approach we only explicitly optimize for mean performance. We introduce robustness by optimizing for mean performance under all distributions within the ambiguity set. This requirement prevents the optimization from over-fitting the design to a single distribution that, under partial observability, is likely to differ from the true distribution. The DRO formulation often implicitly reduces the variation in performance over the uncertainty space, but more importantly, it guarantees that the resulting design is robust to deviations in the uncertain distribution. The challenge of DRO lies in selecting the ambiguity set, and computing the worst-case distribution in a computationally efficient manner.

[a] As implemented in the function `fmincon`, included in the `MATLAB` Optimization Toolbox[22]

3.2 Constructing the ambiguity set and finding the worst-case distribution

The tractability and success of the distributionally robust approach relies on a careful selection of the ambiguity set $\mathcal{P}$. In the context we consider, the designer has access to a sample of m realizations of the uncertain variables. We define $\Omega^m \subseteq \mathbb{R}^m$ to be the set of all possible vectors assigning probabilities to each of the m realizations, i.e.,

$$\Omega^m = \{\mathbf{p} \in \mathbb{R}^m \ : \ \sum_{i=1}^{m} p_i = 1, p_i \geq 0 \ i = 1, ..., m\}. \tag{3.4}$$

Each element in Ω^m is a vector $\mathbf{p}$, where p_i gives the probability that the random variable $\mathbf{u}$ takes the value $\mathbf{u}_i$, for $i = 1, ..., m$. If we only consider the empirical distribution associated with the sample, then the ambiguity set becomes a singleton, $\mathcal{P} = \{\hat{\mathbf{p}}\}$. This reduces the problem to the SAA problem $\mathcal{S}$ (Eqn. 2.2). In the DRO approach, we acknowledge the fact that the empirical distribution $\hat{\mathbf{p}}$ does not perfectly reflect the true probability distribution. This is done by including additional distributions in the ambiguity set.

To ensure that the ambiguity set contains the true distribution with high probability, we aim to include all distributions that are a plausible reflection of the true distribution. To this end, we define a function $D(\cdot, \cdot) : \Omega^m \times \Omega^m \to \mathbb{R}_{\geq 0}$, that measures the distance (a non-negative scalar) between two discrete probability distributions in the probability space Ω^m. We then construct the ambiguity set $\mathcal{P} \subseteq \Omega^m$ to contain all distributions that are sufficiently similar to the empirical distribution given by the sample data. The allowable distance from the empirical distribution, r, is termed the *radius of ambiguity*. Using these definitions, the ambiguity set constructed using the distance function D, is given by

$$\mathcal{P}_D(\hat{\mathbf{p}}, r) = \{\mathbf{p} \in \Omega^m : D(\mathbf{p}, \hat{\mathbf{p}}) \leq r\}. \tag{3.5}$$

One decision to make when constructing the ambiguity set is the choice of r,

the radius of ambiguity, which determines the size of the set. Increasing the size of the ambiguity set allows the designer to be almost certain that it will include the true distribution. On the other hand, if the ambiguity set becomes too large the optimization will be unable to fully exploit the structure of the data. This leads to over-conservatism and, consequently, poor mean performance of the resulting design. To see this, consider expanding the ambiguity set to include every possible distribution. This gives rise to the usual worst-case optimization problem, since the worst-case distribution will always be one in which all the probability density is assigned to the worst possible realization of the uncertainty. In many real situations this is an unrealistic candidate for the true distribution, and so accounting for this distribution in an optimization for mean performance is generally an overly conservative approach. The effect of the radius of ambiguity on the performance of the resulting DRO designs is explored in Section 4.2.

The other important decision to make when constructing the ambiguity set is the choice of the distance function, $D(\cdot, \cdot)$. A good function is one that accurately measures how plausible a given distribution is, given our known data on the uncertainty. This ensures that as we grow the ambiguity set we preferentially add distributions that are plausible candidates for the true distribution. The following subsections introduce two of the distance functions studied in this work. In each subsection, we discuss the benefits and drawbacks of each choice and present algorithms for solving the inner maximization problem (Eqn. 3.3) that results from each choice.

3.2.1 L_2-norm ambiguity

The first distance function we consider utilizes the L_2 or Euclidean norm between two vectors in Ω^m, and is defined by

$$D_{\mathrm{L2}}(\hat{\mathbf{p}}, \mathbf{p}) = \|\hat{\mathbf{p}} - \mathbf{p}\|_2 = \sqrt{\sum_{i=1}^{m} (\hat{p}_i - p_i)^2}. \tag{3.6}$$

The L_2-norm can be interpreted as a modified χ^2 distance. Consequently, the ambiguity set generated using this distance can be interpreted as the set of probabilities satisfying a goodness of fit test in relation to the empirical distribution (see Ref. 24 for details on this interpretation).

Using this notion of distance, we define the ambiguity set as the set of all distributions within a given L_2 distance of the sample distribution. We continue to denote the radius of ambiguity by r, where it will be clear from the context that this is measured in the L_2-norm. The L_2-norm ambiguity set is defined as

$$\mathcal{P}_{\mathrm{L2}}(\,\hat{\mathbf{p}},\, r\,) = \{\, \mathbf{p} \in \Omega^m \;:\; D_{\mathrm{L2}}(\hat{\mathbf{p}}, \mathbf{p}) \leq r \,\}\,. \tag{3.7}$$

Note that after a particular value of $r = r_{\max}$, the ambiguity set will contain all possible distributions, i.e. $\mathcal{P}_{\mathrm{L2}}(\hat{\mathbf{p}}, r_{\max}) = \Omega^m$. Increasing r beyond $r_{\max}$ will thus have no effect on $\mathcal{P}_{\mathrm{L2}}$. Note also that $r_{\max}$ depends on m. With these facts in mind, we introduce a normalized form of the radius of ambiguity. This normalization assists in the selection of r, and facilitates size comparisons between ambiguity sets generated using different distance functions. The normalized radius of ambiguity is defined as

$$\bar{r} = r/r_{\max} \in [0, 1]\,. \tag{3.8}$$

The allowable range of the radius of ambiguity can be expressed as $\bar{r} \in [0, 1]$, regardless of m. Note that selecting the $\bar{r}$ parameter for the ambiguity set in the DRO problem, $\mathcal{D}$, is analogous to selecting the parameter λ in the MOO problem, $\mathcal{M}$.

When $\hat{\mathbf{p}}$ corresponds to the empirical distribution for a sample of equally likely realizations, we have the special case that $\hat{\mathbf{p}} = \frac{\mathbb{1}}{m}$, where $\mathbb{1} \in \mathbb{R}^m$ denotes a vector of ones. In this case we can derive $r_{\max}$ analytically:

$$r_{\max}(m) = D_{\mathrm{L2}}(\hat{\mathbf{p}}, \mathbf{e_1}) = \sqrt{\left(\frac{1}{m} - 1\right)^2 + \sum_{i=2}^{m} \left(\frac{1}{m}\right)^2} = \sqrt{1 - \frac{1}{m}}, \tag{3.9}$$

where the degenerate distribution $\mathbf{e}_i \in \Omega^m$ has $e_i = 1$, and all other entries set to 0.

In order to gain an intuition for the shape of the L_2 ambiguity set, Figure 3-1 illustrates $\mathcal{P}_{\text{L2}} \subseteq \Omega^3$, for $\bar{r} = 0.25, 0.5, 0.75$, plotted in barycentric co-ordinates. It

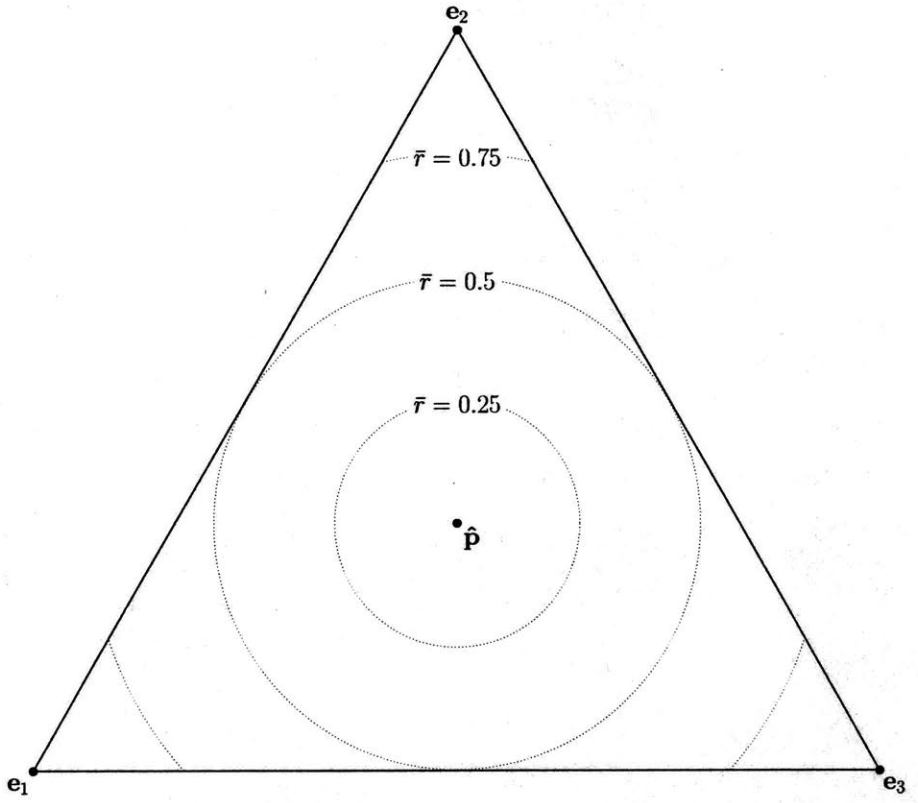


Figure 3-1: *Plot of L_2 ambiguity sets in 3-dimensional barycentric co-ordinates.*

is interesting to note that at a certain value of $\bar{r}$, $\mathcal{P}_{\text{L2}}$ reaches the boundary of the probability simplex Ω^m. From this point on, $\mathcal{P}_{\text{L2}}$ will include distributions that assign a probability of zero to some values of the uncertainty that have been realized in the sample. A criticism of the L_2-norm distance is that as the ambiguity set grows, it will include such distributions before other distributions that are more plausible, i.e. those that assign all realizations in the sample a non-zero probability.

Under this choice of distance function, the inner maximization problem given by Eq. (3.3) involves finding the distribution $\mathbf{p}^* \in \mathcal{P}_{\text{L2}}$ that generates the worst possible expected performance of a given design. Given a design, $\mathbf{x}$, a sample of the uncertain parameters, $\mathbf{u}_i$, for $i = 1, ..., m$, and a radius of ambiguity r, this can be formulated as a convex maximization problem, involving a linear objective function subject to

linear and second-order cone constraints:

$$\begin{aligned} \max_{\mathbf{p}} \quad & \sum_{i=1}^{m} p_i Q(\mathbf{x}, \mathbf{u}_i) \\ \text{s.t.} \quad & \|\hat{\mathbf{p}} - \mathbf{p}\|_2 \leq r \\ & \mathbf{p} \in \Omega^m \end{aligned} \tag{3.10}$$

Much of the appeal of using an ambiguity set defined by the L_2-norm is that this problem can be solved in closed-form using Algorithm 3-1. This algorithm was originally derived in Ref. 24, for the limited case where $\hat{\mathbf{p}} = \frac{\mathbb{1}}{m}$. The algorithm presented here is an extension applicable to a general $\hat{\mathbf{p}}$. The relaxation of this restriction is useful if, for example, the designer opts to bin the realizations and create the sample distribution out of bin frequencies, since this will mean that $\hat{\mathbf{p}} \neq \frac{\mathbb{1}}{m}$ in general. For details on the derivation of the algorithm see Ref. 24.

Algorithm 3-1: *Compute the worst-case probability distribution in a given L2 ball.*

Input:

Discrete reference distribution $\hat{\mathbf{p}}$,
Function values at distribution support $Q(\mathbf{x}, \mathbf{u}_i) \equiv Q_i$,
Radius of L2 ambiguity set r.

Output:

Worst-case discrete distribution $\mathbf{p}^*$, on the same support as $\hat{\mathbf{p}}$.

1: $m = \text{length}(\hat{\mathbf{p}})$
2: $K = \{1, 2, ..., m\}$
3: **while** $|K| > 1$ **do**
4: $\quad k = m - |K|$
5: $\quad \bar{Q} = \dfrac{1}{(m-k)} \sum_{i \in K} Q_i$
6: $\quad s = \sqrt{\dfrac{1}{(m-k)} \sum_{i \in K} (Q_i^2 - \bar{Q}^2)}$
7: $\quad$ **if** $k = 0$ **then**
8: $\quad\quad p_i = \hat{p}_i + \dfrac{Q_i - \bar{Q}}{\sqrt{m}\, s} r, \; i \in K$
9: $\quad$ **else**
10: $\quad\quad$ **for** $i \in K$ **do**
11: $\quad\quad\quad p_i = \hat{p}_i + \dfrac{1}{(m-k)} \left(\sum_{i \notin K} \hat{p}_i + \sqrt{(m-k)(r^2 - \sum_{i \notin K} \hat{p}_i^2) - (\sum_{i \notin K} \hat{p})^2} \dfrac{Q_i - \bar{Q}}{s} \right)$
12: $\quad\quad$ **end for**
13: $\quad\quad p_i = 0, \; i \notin K$
14: $\quad$ **end if**
15: $\quad$ **if** $p_i \geq 0 \quad \forall i \in K$ **then**
16: $\quad\quad \mathbf{p}^* = \mathbf{p}$
17: $\quad\quad$ **return** $\mathbf{p}^*$
18: $\quad$ **else**
19: $\quad\quad$ Find critical $j \in K$. This is the last index of $p_i < 0$ to become positive as we decrease r. This can be done by setting $p_i = 0$ in line 10 and solving for r.
20: $\quad\quad$ Set $K = K \setminus \{j\}$.
21: $\quad$ **end if**
22: **end while**
23: $p_i = 0, \quad i \notin K$
24: $p_i = 1, \quad i \in K$
25: $\mathbf{p}^* = \mathbf{p}$
26: **return** $\mathbf{p}^*$

3.2.2 K-L divergence ambiguity

The second function that we study for comparing the degree of similarity between two discrete probability distributions is the K-L divergence (or relative entropy). The K-L divergence has been used extensively in the literature to construct ambiguity sets for DRO.[15–18] The K-L divergence between the reference distribution, $\hat{\mathbf{p}}$, and a distribution (of equal dimension), $\mathbf{p}$, is defined by

$$D_{\mathrm{KL}}(\hat{\mathbf{p}}, \mathbf{p}) = \sum_{i=1}^{m} \hat{p}_i \log\left(\frac{\hat{p}_i}{p_i}\right). \tag{3.11}$$

The K-L divergence has two cases that require special consideration. In this work, we adopt the conventions that for any a,

$$D_{\mathrm{KL}}(0, a) = 0 \log \frac{0}{a} = 0, \tag{3.12}$$

$$D_{\mathrm{KL}}(a, 0) = a \log \frac{a}{0} = \infty. \tag{3.13}$$

The first convention (3.12) is adopted because $\lim_{x \to 0} x \log(x) = 0$. The second convention (3.13) ensures that a distribution that assigns a zero probability to a realization that was observed in the sample will result in an infinite K-L divergence. This ensures that such distributions will not be included in the K-L divergence ambiguity set. This is a favorable contrast with the L_2-norm ambiguity set, which can include distributions that assign realizations observed in the sample a likelihood of zero.

Since the K-L divergence function is not symmetric, the ordering of the arguments matters. In this work we choose the reference distribution $\hat{\mathbf{p}}$ as the first argument when constructing the K-L ambiguity set. This is motivated by a recent paper by Van Parys et al., which showed that this choice is optimal (see Ref. 16 for more details).

We define the K-L ambiguity set in a way that is analogous to the L_2-norm ambiguity set. We denote the radius of ambiguity by r, where it will be clear from the context when this is measured in terms of the K-L divergence (as opposed to the

L_2-norm), and define

$$\mathcal{P}_{\mathrm{KL}}(\,\hat{\mathbf{p}},\; r\,) = \{\; \mathbf{p} \in \Omega^m \;:\; D_{\mathrm{KL}}(\hat{\mathbf{p}}, \mathbf{p}) \leq r \;\}\,. \tag{3.14}$$

As in the previous section we define a normalized radius of ambiguity $\bar{r}$, which we can use to directly compare the size of the K-L divergence ambiguity set with the size of an L_2-norm ambiguity set. However in this case there is no well-defined maximum radius $r_{\max}$. In fact, we could increase r indefinitely, with the boundaries of $\mathcal{P}_{\mathrm{KL}}$ getting asymptotically close to the probability simplex Ω^m. Instead of normalizing by the maximum radius as we did in the L_2-norm case, we instead define the normalized K-L divergence radius of ambiguity to be

$$\bar{r} = \frac{\max_{\mathbf{p} \in \mathcal{P}_{\mathrm{KL}}} \|\mathbf{p} - \hat{\mathbf{p}}\|_\infty}{1 - \frac{1}{m}} \in [0, 1]. \tag{3.15}$$

This is essentially a normalized L_∞-norm radius. The maximum L_∞ norm gives the maximum difference that any single probability is allowed to vary, over all distributions within the ambiguity set. Note that there are other valid ways to parameterize the size of the ambiguity set, e.g., by measuring the volume proportion of Ω^m contained in the ambiguity set. However, care should be taken that the chosen parameterization is consistent across different dimensions, m, and distance function D. In order to gain an intuition for the shape of the K-L ambiguity set, Figure 3-2 illustrates $\mathcal{P}_{\mathrm{KL}} \subseteq \Omega^3$, for $\bar{r} = 0.25, 0.5, 0.75$, plotted in barycentric co-ordinates.

Given a design $\mathbf{x}$, a sample of the uncertain parameters $\mathbf{u}_i$, for $i = 1, ..., m$, and a K-L divergence radius of ambiguity r, the inner maximization problem given by Eq. (3.3) can be formulated as a convex maximization problem. This is done by introducing new variables $z_i, i = 1, ..., m$ and performing simple algebraic manipulation in order to reformulate the K-L divergence constraint to yield linear and exponential

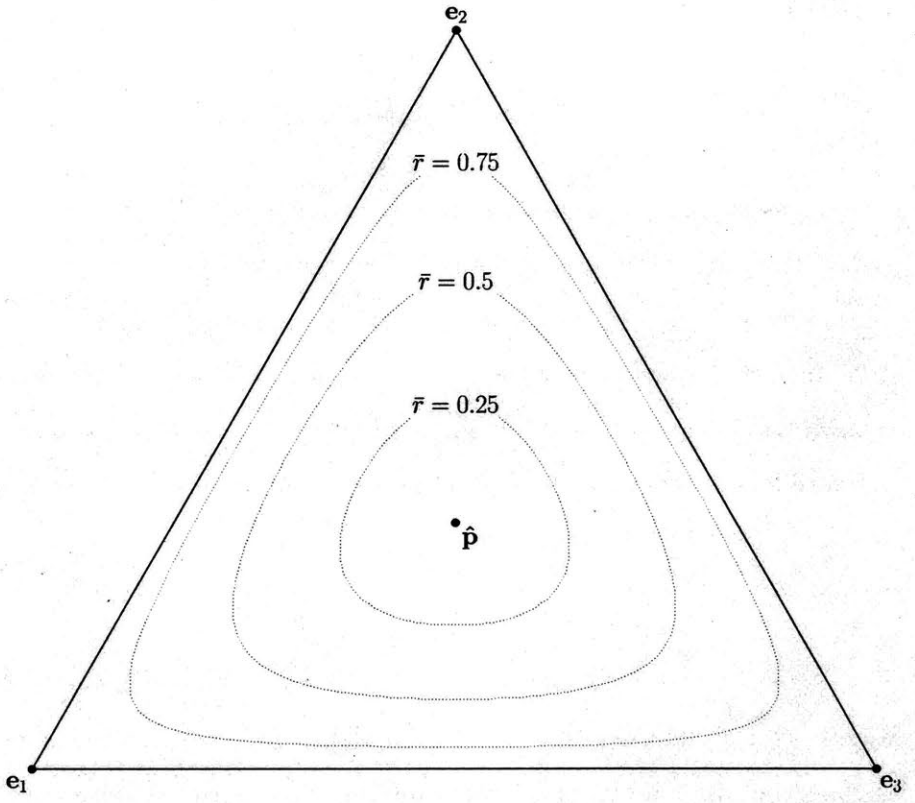


Figure 3-2: *Plot of K-L ambiguity sets in 3-dimensional barycentric co-ordinates.*

cone type constraints:

$$\sum_{i=1}^{m} \hat{p}_i \log\left(\frac{\hat{p}_i}{p_i}\right) \leq r \iff \begin{cases} p_i \geq \hat{p}_i \exp\left(\dfrac{-z_i}{\hat{p}_i}\right), & i = 1, \ldots, m \\ \displaystyle\sum_{i=1}^{m} z_i \leq r \end{cases} \tag{3.16}$$

Under this reformulation, the maximization can be written as

$$\begin{aligned} \max_{\mathbf{p},\mathbf{z}} \quad & \sum_{i=1}^{m} p_i Q(\mathbf{x}, \mathbf{u}_i) \\ \text{s.t.} \quad & p_i \geq \hat{p}_i \exp\left(\frac{-z_i}{\hat{p}_i}\right), \quad i = 1, \ldots, m \\ & \sum_{i=1}^{m} z_i \leq r \\ & \mathbf{p} \in \Omega^m \end{aligned} \tag{3.17}$$

The problem can be solved in this form by many commercial convex optimization packages, for example the CVX package,[25,26] which uses a successive approximation method. However, since this problem needs to be solved at every iteration of the design optimization, the computational cost of a brute force optimization such as this can quickly accumulate. To achieve further speedup, the optimization problem in Eqn. 3.17 has been analyzed in detail in the literature,[27,28] and has been shown to be reducible to a scalar root-finding problem. Algorithm 3-2 outlines how to solve the the optimization problem using this approach, which is far more computationally efficient than solving Eqn. 3.17 directly. For details on the derivation of the algorithm, as well as a discussion on how best to solve the root finding problem (line 14 of the algorithm) we refer the reader to Reference 28. Note that although the authors of that work suggest using Newton's method, it was found for our problem that a bisection method with initial bracket

$$[\nu_L, \nu_R] = \left[\max_{i \in \bar{K}} Q_i + \epsilon, \max_{i \in \bar{K}} Q_i + 100\right], \tag{3.18}$$

where ϵ is a small positive number, proved to be efficient and robust.

Algorithm 3-2: *Compute worst-case distribution within a KL ball*

Input:

Discrete reference distribution $\hat{\mathbf{p}}$,

Function values at distribution support $Q(\mathbf{x}, \mathbf{u}_i) \equiv Q_i$,

Radius of KL ambiguity set r.

Output:

Worst-case discrete distribution $\mathbf{p}^*$, on the same support as $\hat{\mathbf{p}}$.

$m = \text{length}(\hat{\mathbf{p}})$

$K = \{i : \hat{p}_i = 0\}$

$\tilde{K} = \{i : \hat{p}_i > 0\}$

Define $f(\nu) = \sum_{i \in \tilde{K}} \hat{p}_i \log(\nu - Q_i) + \log\left(\sum_{i \in \tilde{K}} \frac{\hat{p}_i}{\nu - Q_i}\right)$, for $\nu > \max_{i \in \tilde{K}} Q_i$

for $i \in K$ **do**

    $p_i = 0$

end for

$I^* = K \cap \text{argmax}_i \, Q_i$

if $\exists\, k \in I^*$ **such that** $f(Q_k) < r$ **then** ▷ If multiple k exist, choose any one

    $\nu = Q_k$

    $r = 1 - \exp(f(\nu) - r)$

    $p_k = r$

else

    Find ν such that $f(\nu) = r$ using a line search

    $r = 0$

end if

for $i \in \tilde{K}$ **do**

    $q_i = \frac{\hat{p}_i}{\nu - Q_i}$

end for

for $i \in \tilde{K}$ **do**

    $p_i = \frac{(1 - r) q_i}{\sum_{i=1}^{m} q_i}$

end for

$\mathbf{p}^* = \mathbf{p}$

return $\mathbf{p}^*$

CHAPTER 4

PERFORMANCE OF THE DISTRIBUTIONALLY ROBUST APPROACH

This chapter explores various aspects of the performance of the DRO design methodology outlined in the previous section. In particular, we investigate three important aspects of the approach by applying it to the acoustic horn design problem presented in Section 2.2.1. In Section 4.1 we investigate how the method trades off mean performance with the risk level in performance, and also compare and contrast the performance of distance functions used to construct the ambiguity set. In Section 4.2 we explore the issue of sizing the ambiguity set. In particular, we investigate the relationship between the size of the ambiguity set and the performance of the resulting designs, and how this relationship depends on how accurately the sample data reflects the true distribution of uncertainty. Finally, Section 4.3 analyzes how the performance of the DRO approach depends on the distribution of the underlying uncertainty.

4.1 Mean-risk tradeoff

In Section 3.2, we discussed how enlarging the ambiguity set trades off mean performance over all sample draws, versus robustness in performance between draws. In this section, we study how effectively the DRO methodology is able to make this trade-off, for each of the distance functions studied.

We do this using computational experiments on the acoustic horn design problem described in Section 2.2.1. In particular, we draw $T = 1000, 600, 200$ independent random samples of size $m = 5, 10, 20$ (respectively) from $\mathbb{P}_{\mathbf{u}}$. Note that in an effort to minimize the effect of random sampling when comparing the different methods, we use the same samples as those used to study the MOO approach. For each sample, we compute distributionally robust designs $\mathbf{x}_m$ by solving problem $\mathcal{D}$. We do this for the L_2-norm ambiguity set, using 22 different values of the radius of ambiguity spanning the range $\bar{r} \in [0, 1]$. This experiment is repeated (again using the same samples), but this time using the K-L divergence ambiguity set, and 21 uniformly spaced values of the radius of ambiguity spanning $\bar{r} \in [0, 1]$. For each design we evaluate the mean of the QoI under the true distribution, and denote this by Z_m (see Eqn. 2.3).

We are interested in how effective and robust the DRO methodology is at producing designs that perform well under the true distribution, when given only m realizations of the uncertain variables to use in the optimization. To this end, we compute the mean performance, $\mu(Z_m)$, and the CVaR_{10} or risk level in performance, $\phi(Z_m)$, over the T sample draws (as discussed in Section 2.3). Figure 4-1 below plots the mean-risk trade-off curves for the DRO methodology using both L_2-norm and K-L divergence ambiguity sets. The analogous mean-risk curves for the MOO approach (from Figure 2-8) are also shown for comparison.

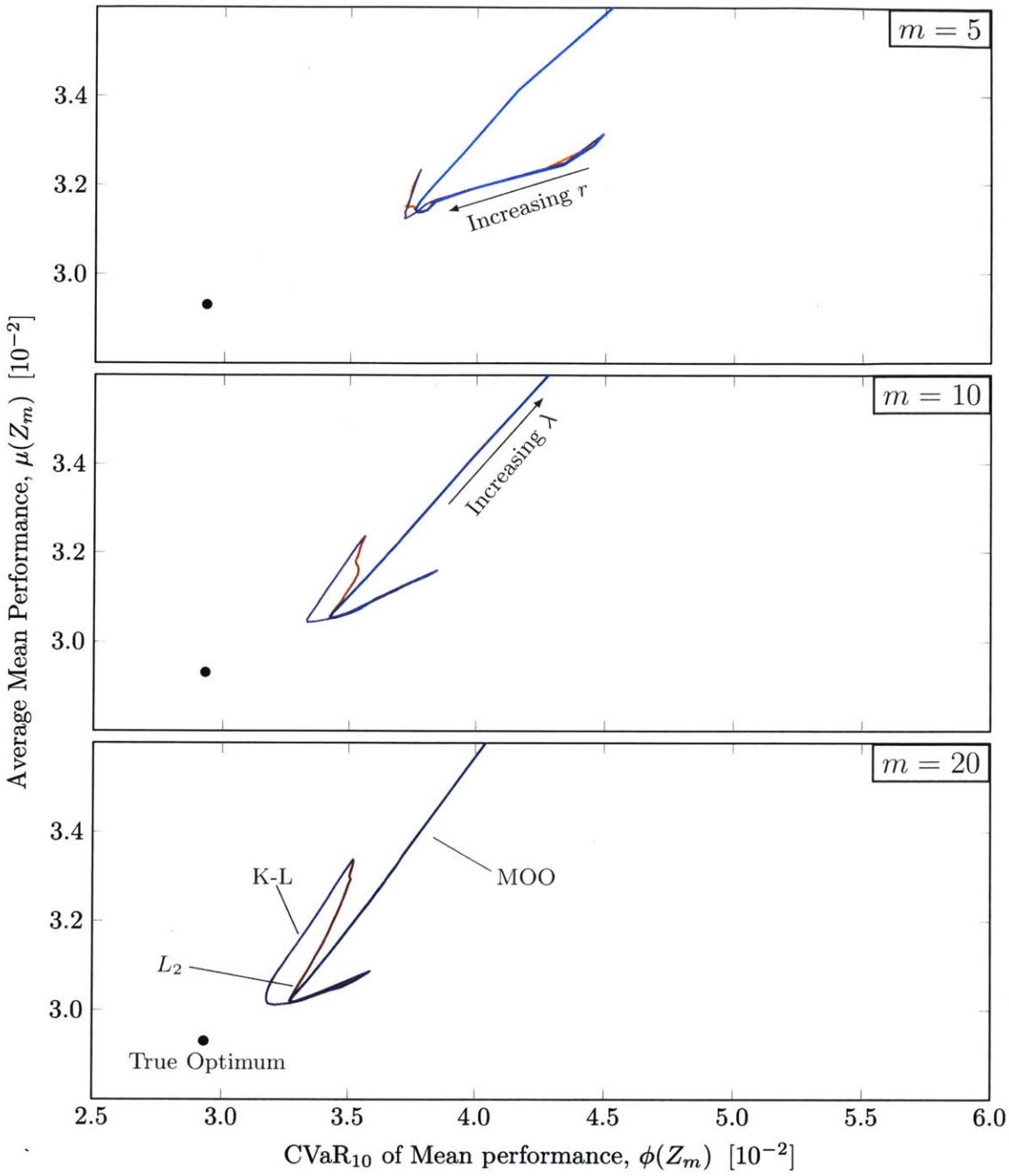


Figure 4-1: *Mean-Risk tradeoff curves for designs computed using sample draws of size* $m = 5, 10, 20$ *and the DRO approach with* L_2*-norm and K-L divergence ambiguity sets. Also shown are the corresponding curves for the MOO approach, as well as the true optimal performance.*

The first thing to note is that when the radius of ambiguity is zero, the DRO and MOO problems, $\mathcal{D}$ and $\mathcal{M}$ respectively, are both equivalent to the SAA problem, $\mathcal{S}$. Naturally, this means that the mean-risk trade-off curves for all methods coincide

here. As we increase the radius of ambiguity, $\bar{r}$, we increase the amount of distributional robustness enforced in the designs, which is analogous to increasing λ in the MOO approach. We see that the mean-risk tradeoff curves for the DRO have qualitatively similar behavior to their MOO counterparts. At first, adding robustness leads to an improvement in both the mean and risk objectives, but after a critical radius of ambiguity is reached, increasing the radius further results in worse performance in terms of both mean and risk. However, it is worth noting that this drop in performance is bounded much more favorably when using the DRO approach than with the MOO approach. This means that designs found using the largest ambiguity sets significantly outperform the designs found using the MOO approach with $\lambda = 0.95$.

The mean-risk curve corresponding to the K-L divergence ambiguity set reaches the critical point at a better mean and risk level than the mean-risk curves for either the L_2-norm ambiguity set or the MOO approach. This means that if the optimal values of λ and r are chosen for each sample size and ambiguity set combination, the DRO approach using a K-L divergence ambiguity set will, on average, produce designs with the best mean performance and lowest risk level in performance.

We define the optimality gap for a methodology as the difference between the performance of designs found using that methodology (averaged over T sample realizations), and the true optimal design performance, which the designer could obtain if they had full observability over the uncertainty. The SAA approach exhibits the largest optimality gap, and is thus used as a baseline. Table 4-1 presents the optimal mean and risk levels attained by each method, for each sample size. Also shown is the corresponding percentage reduction in the optimality gap (when compared with the SAA baseline) achieved by each method.

We see that for all of the sample sizes we studied, the DRO method with a K-L divergence ambiguity set results in the greatest percentage reduction in the optimality gap. We also see that the percentage reductions decrease as the sample size is increased, indicating that the benefit of the DRO approach is greatest at small sample sizes.

		Mean (%)		CVaR(%)	
m=5	SAA	3.32	(-)	4.49	(-)
	MOO	3.14	(41.62)	3.76	(45.91)
	L2	3.14	(41.66)	3.72	(48.23)
	KL	3.13	(45.58)	3.72	(48.47)
m=10	SAA	3.16	(-)	3.85	(-)
	MOO	3.05	(39.61)	3.42	(44.41)
	L2	3.05	(40.11)	4.42	(44.60)
	KL	3.04	(43.04)	3.33	(53.97)
m=20	SAA	3.09	(-)	3.59	(-)
	MOO	3.02	(35.63)	3.27	(46.52)
	L2	3.02	(35.81)	3.27	(45.20)
	KL	3.01	(38.49)	3.18	(58.82)

Table 4-1: *Best average mean and average $CVaR_{10}$ values attained by each method for sample sizes of $m = 5, 10, 20$. Also shown is the percentage reduction in the optimality gap relative to the SAA result.*

4.2 Sizing the ambiguity set

For a given problem, there is an optimal ambiguity set size that would result in an optimal mean-risk tradeoff; however in practice the designer will not know this optimal radius of ambiguity. In this section we investigate how the size of the ambiguity set affects the performance of the resulting designs, with the goal of uncovering trends that can inform the designer on a good choice of ambiguity set size. In particular, we study how the optimal size of the ambiguity set and the performance of the resulting design are influenced by the degree to which the empirical distribution associated with the sampled uncertainty reflects the true distribution.

We introduce the notion of statistical distance between a sample and the true distribution from which the sample was drawn. For this analysis, we choose the Kolmogorov-Smirnov (K-S) statistic as the measure of statistical distance. In our setting, the K-S statistic is defined as the maximum difference between the empirical cumulative distribution function (CDF), and the (continuous) CDF of the true

distribution, $\mathbb{P}_{\mathbf{u}}$. We denote the K-S statistic by κ. A sample that has an empirical distribution closely resembling $\mathbb{P}_{\mathbf{u}}$ will be close to $\kappa = 0$, while a sample with an empirical distribution that does not well resemble $\mathbb{P}_{\mathbf{u}}$ will be close to $\kappa = 1$.

For each of the samples used in the previous section, the K-S statistic was computed, and the realizations were sorted in order of increasing κ. The realizations were then separated into 7 equally sized bins, with each bin characterized by the mean K-S statistic of all realizations within the bin. In what follows, we will focus primarily on the K-L divergence ambiguity set. Analogous results for the L_2-norm ambiguity set, along with a discussion on how these differ from the K-L divergence case, are presented in Appendix A.

Figure 4-2 presents the mean performance of designs computed using samples of size $m = 5$. We average the mean performance of designs across the realizations in each K-S statistic bin, and plot average mean design performance versus the mean κ of the bin, and the normalized radius of ambiguity used in the optimization, $\bar{r}$. We first focus on Figure 4-2(a), which shows the relationship between design performance and the size of the ambiguity set, for three different levels of κ. We see that as κ increases, the resulting design performance generally gets worse. At all levels of κ, increasing the size of the ambiguity set from $\bar{r} = 0$ results in an improvement in design performance until a critical ambiguity set size is reached, at which point further increase results in worse design performance. Note that this was the trend observed in Section 4.1. However, we now see an additional trend: as κ increases and the sample becomes a worse representation of the true distribution, the critical ambiguity set size also increases. A higher value of κ is also associated with a less severe drop in performance when the critical ambiguity set size is exceeded. This means that overestimating the critical ambiguity set size affects low κ realizations more than high κ realizations.

We now turn our attention to Figure 4-2(b), which shows the relationship between the K-S statistic, κ, and design performance for designs found using three different sizes of the ambiguity set. Again we see the general trend that an increase in κ results in worse design performance. We also see that a relatively small increase in

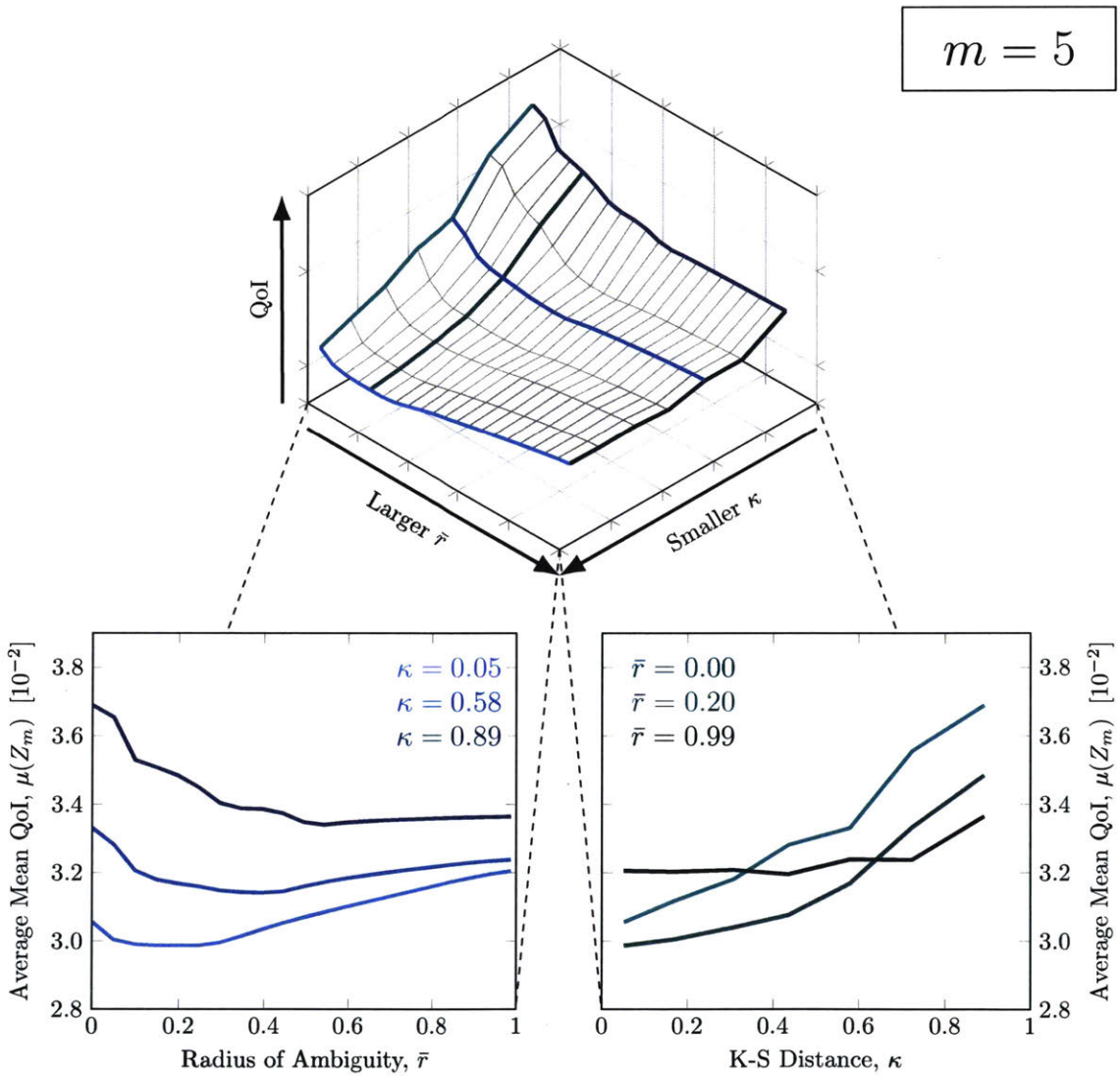
the ambiguity set size from $\bar{r} = 0$ to $\bar{r} = 0.2$ results in improved mean performance across all levels of κ. This means that using a non-zero ambiguity set size always resulted in improved design performance, regardless of whether or not the sample used was a good representation of the true distribution. Increasing the size of the ambiguity set further continues to improve performance for samples with high κ, but now this comes at the cost of worse performance for samples that have small κ. A result of this is that using a large radius of ambiguity results in more consistent design performance across different levels of κ.

Figure 4-3 repeats this analysis for a sample size of $m = 10$. Figure 4-3(a) shows that at this sample size a small increase in the ambiguity set size away from zero is still associated with an improvement in performance for any κ. However, when compared with the $m = 5$ case, the critical ambiguity set size is smaller for all levels of κ. This has two implications. Firstly, the performance improvement attainable by increasing the ambiguity set size from zero to the critical value is less than for $m = 5$. Secondly, the maximum penalty in performance caused by exceeding the critical ambiguity set size is more severe. As a result of this second implication, an increase in κ is no longer always associated with worse performance. Instead, we see that optimizing using a large ambiguity set can, in fact, perform better on data that is a worse representation of the true distribution, than on data that represents the true distribution well. Figure 4-3(b) shows the relationship between κ and performance for samples of size $m = 10$. We observe qualitatively similar behavior to the $m = 5$ case. The only notable difference at this sample size is that as the radius of ambiguity increases beyond the critical value, the penalty in performance for small κ is greater, while the improvement in performance for high κ is smaller.

The trends observed in increasing the sample size from $m = 10$ to $m = 20$ are similar to those seen between $m = 5$ and $m = 10$. Figure 4-4 shows the results of the analysis for a sample size $m = 20$. In Figure 4-4(a) we see that the critical ambiguity set size is again decreased for all levels of κ, when compared with the $m = 10$ case. It is worth noting however, that the critical ambiguity set size is still greater than zero, indicating that a small ambiguity set will still outperform the SAA approach at

this sample size. In Figure 4-4(b) we see that using a small ambiguity set ($\bar{r} = 0.1$) leads to a small improvement for small κ, and a moderate improvement for large κ. Increasing the size of the ambiguity set a small amount beyond this (to $\bar{r} = 0.3$) we observe the same trend as in the $m = 10$ case, whereby worse design performance for small κ is traded off in favor of better performance for high κ. However, in contrast to the $m = 5$ and $m = 10$ cases, this trend does not continue to the maximum ambiguity set size. Instead, increasing the size beyond $\bar{r} = 0.3$ results in a drop in performance across all levels of κ

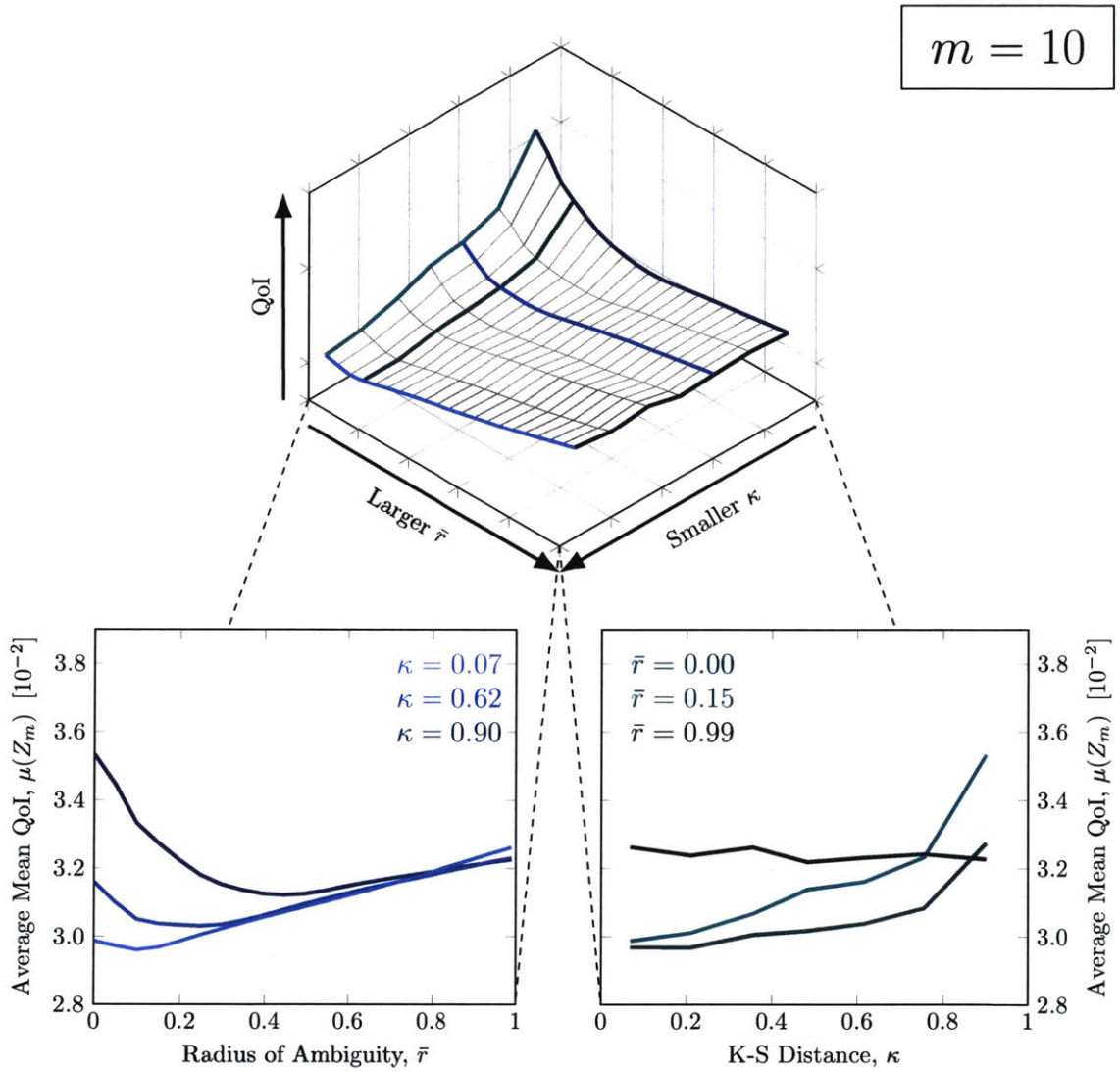
These trends can be summarized to form two key conclusions for the acoustic horn design problem. Firstly, if the designer wishes to optimize the worst-case performance of the design, they should focus on the worst samples (high κ), as these are the samples that result in the worst performing designs. To do this they should tailor the size of the ambiguity set to these realizations by using a relatively large ambiguity set size. In this way the designer will hedge against samples with high κ, at the cost of a drop in performance if the sample draw has low κ. Secondly, as the amount of data available increases, the designer should generally decrease the size of the ambiguity set. In the $m = 20$ case in particular, care should be taken not to increase the size of the ambiguity set too much, as this can result in a drop in performance across all levels of κ. It should be stressed that the trends and relationships shown in this section are for the acoustic horn design problem. Such trends will be problem dependent, as they will depend on the shape of the function defining the quantity of interest, i.e., $Q(\mathbf{x}, \mathbf{u})$. Note in particular that there may be some problems for which designing using samples with a large κ do not result in over-fitting, e.g., if the quantity of interest is constant with respect to the uncertain variables.



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

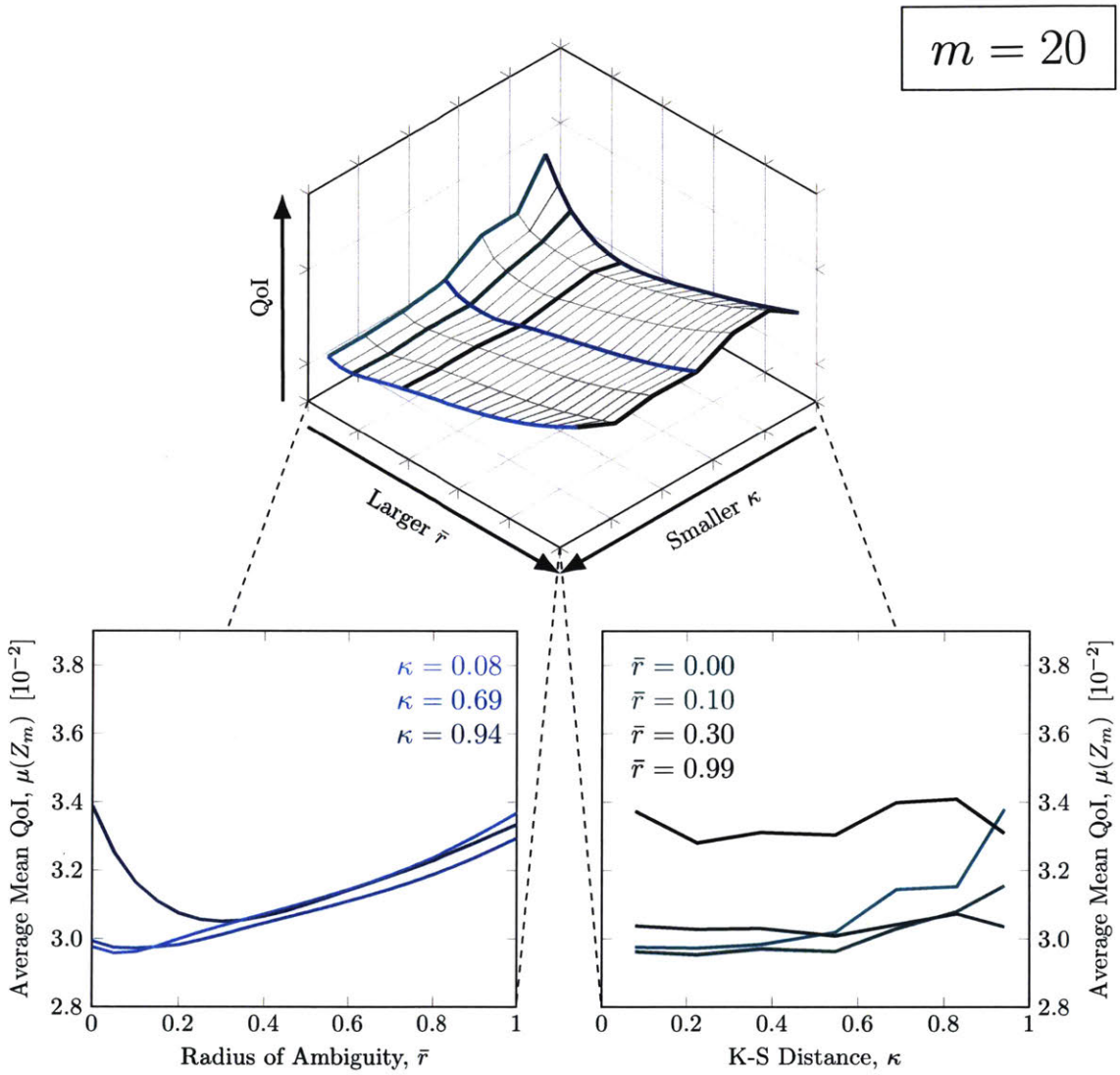
Figure 4-2: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using uniformly distributed samples of size $m = 5$ and the DRO approach with a K-L divergence ambiguity set.*



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

Figure 4-3: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using uniformly distributed samples of size $m = 10$ and the DRO approach with a K-L divergence ambiguity set.*



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

Figure 4-4: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using uniformly distributed samples of size $m = 20$ and the DRO approach with a K-L divergence ambiguity set.*

4.3 Effect of the underlying distribution

In this section we investigate how the shape of the probability distribution governing the uncertain parameters affects the performance of the DRO approach for design under partially observable uncertainty. When applying the methodology to the acoustic horn model problem in the previous sections, we used a uniform fixed truth distribution (Eqn. 2.14). In this section, we repeat the previous experiments, this time with a new truth distribution, given by:

$$\mathbf{u} \sim \mathbb{P}_u = \text{Normal}(1.4, 0.0577). \tag{4.1}$$

The mean and standard deviation of the normal distribution were chosen to be equal to those of the uniform distribution used in the previous sections. The PDF's of the two distributions are shown in Figure 4-5.

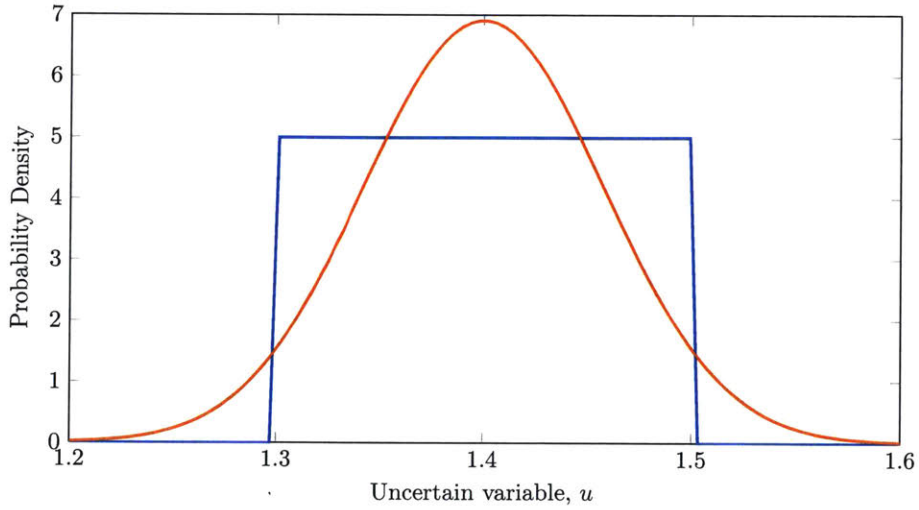


Figure 4-5: *Comparison between the uniform and normal probability density functions used in the acoustic horn design experiments.*

The experiment outlined in Section 4.1 was repeated using the normal distribution. Results for the K-L divergence ambiguity set are presented in this section, while analogous results for the L_2-norm ambiguity set can be found in Appendix A. The performance of designs under the true distribution is evaluated using 10-point Gauss-Hermite quadrature.

Figure 4-6 shows the average mean performance vs. the CVaR_{10} in mean performance over all T samples. The analogous mean-risk curves for the uniform distribution (as seen in Figure 4-1) are shown for comparison. Also shown is the performance of the true optimal designs, found as solutions to the full problem $\mathcal{P}$ (Eqn. 2.1) using the uniform and normal distributions respectively.

We see that the mean-risk curves for the two distributions are qualitatively similar. Under the normally distributed uncertainty, increasing the radius of ambiguity from zero still results in an improvement in both mean performance across the samples, and the CVaR_{10} across the samples. However we see that in comparison to the uniformly distributed uncertainty, the critical radius–after which a further increase is detrimental–occurs much sooner. As a result, the potential performance gain by using the DRO approach compared to the SAA approach (which is equivalent to $r = 0$) is smaller for the normally distributed uncertainty.

We also repeat the analysis of Section 4.2, this time using the normally distributed uncertainty. Figure 4-7 shows the relationship between the radius of ambiguity, the K-S distance of the sample, and the performance of the resulting designs for the case of normally distributed uncertainty. The sample size used is $m = 10$. We are interested in comparing these results with the analogous results for the uniform distribution in Figure 4-3. We see that for the normally distributed uncertainty the critical radius of ambiguity is smaller. As a result, the improvement in performance gained by increasing the radius of ambiguity from zero to the critical radius of ambiguity is also smaller, and the maximum penalty for exceeding this critical radius is greater.

These results show that for the acoustic horn design problem we study, the DRO methodology is still able to outperform the SAA approach under normally distributed uncertainty, but to a lesser extent than under uniformly distributed uncertainty. A full study including additional distributions is required to understand the cause of this, and to generalize these results enough to make claims about which distributions of uncertainty are most amenable to the DRO methodology.

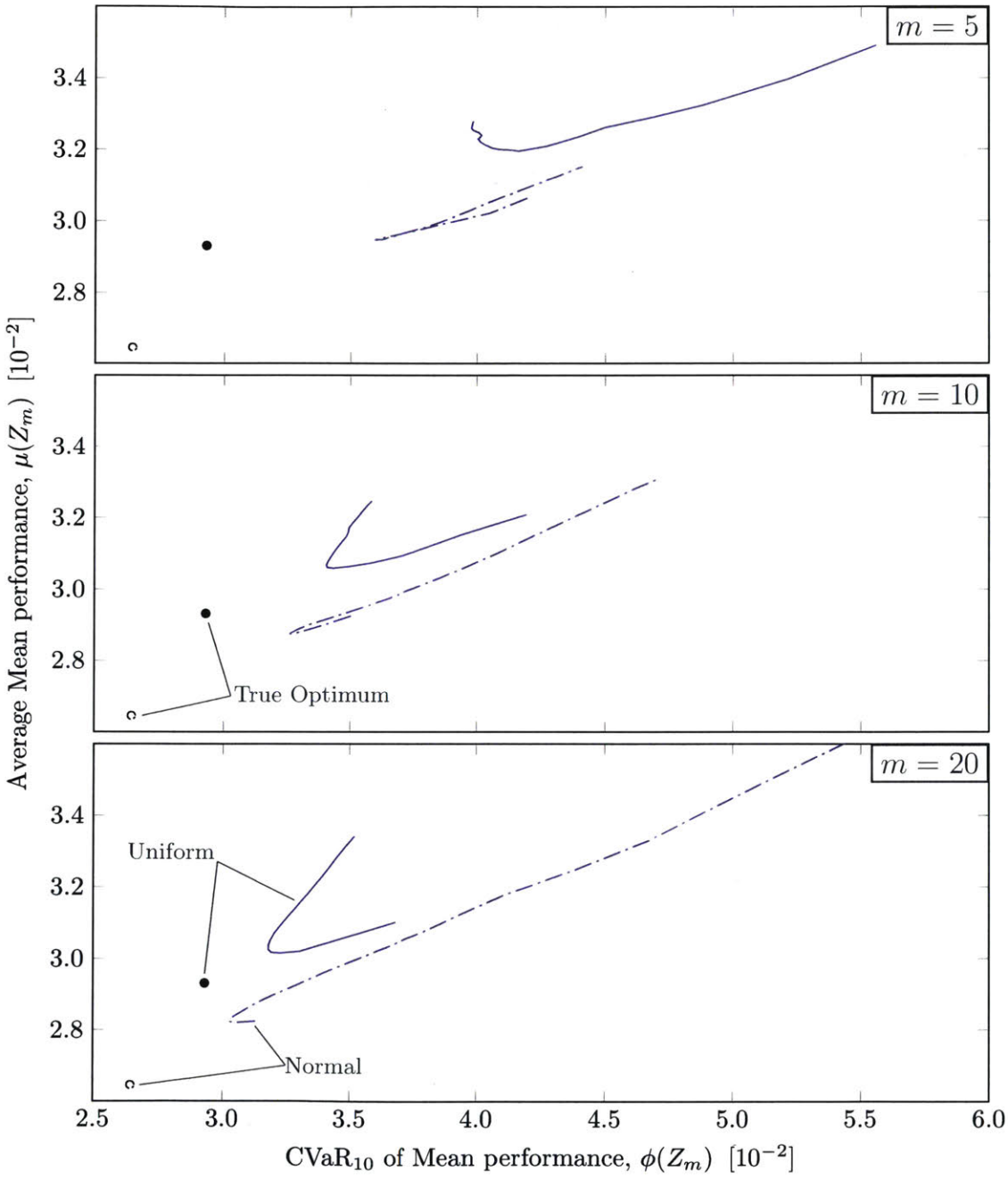
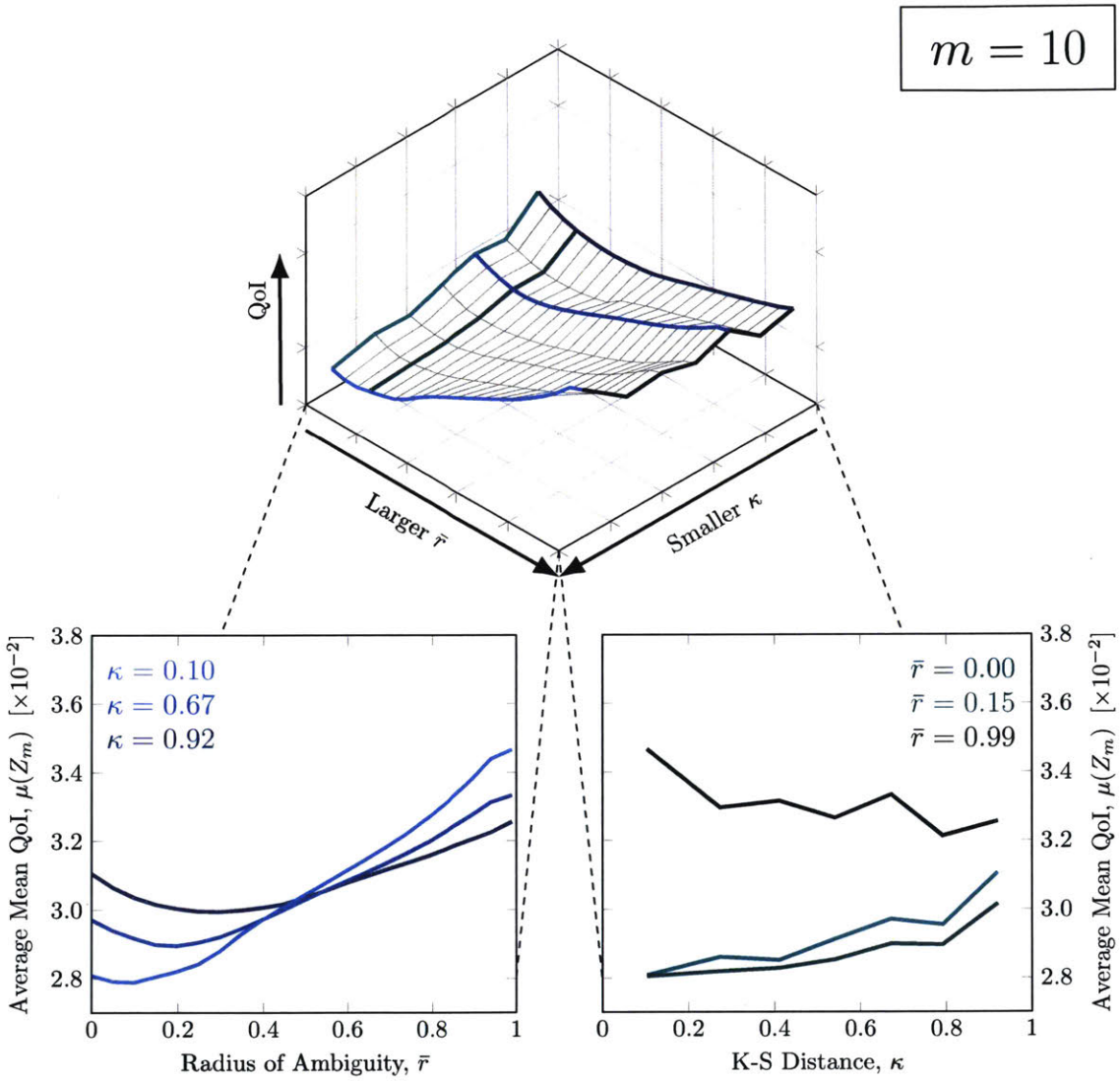


Figure 4-6: *Comparison between mean-risk trade-off curves for designs computed using normally distributed and uniformly distributed uncertainty. Results are for sample sizes $m = 5, 10, 20$ and the DRO approach with K-L divergence ambiguity set.*



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

Figure 4-7: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using normally distributed samples of size $m = 10$ and the DRO approach with a K-L divergence ambiguity set.*

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

This thesis has shown the potential benefits of a distributionally robust approach to design optimization under partially observable uncertainty. Our approach has been shown to outperform both SAA and MOO based approaches on the model design problem we studied. The formulation has also been shown to be computationally tractable, as the algorithms presented in this thesis require a practically identical computational budget to widely used existing methods.

Although the distributionally robust approach in this paper has been demonstrated on a single problem, the formulation and algorithms make no assumptions about the system in question. The approach is therefore applicable to a wide range of design problems. The effects of uncertainty on the quantity of interest vary depending on the system in question. Planned future work involves testing the distributionally robust approach on additional design problems, to explore whether certain characteristics of the system of interest influence the performance of the method. Similarly, characteristics of the underlying distribution of the uncertainty could influence the performance of the method. In this thesis we compared the performance of the method for uniformly and a normally distributed uncertainty, but more distributions will need to be investigated before reliable insights can be gained.

There is also room for future improvements to the distributionally robust approach itself, namely, the exploration of different formulations of the ambiguity set. In this thesis the ambiguity set is formulated using either the L_2-norm or the K-L divergence. Both of these formulations only take advantage of the reference probability

vector $\hat{\mathbf{p}}$ to determine which probability vectors are similar enough to the reference distribution to be included in the ambiguity set. They do not incorporate the sample realizations $\mathbf{u}_1, \ldots, \mathbf{u}_m$. Notions of distance such as the Wasserstein distance (also known as the earth-movers distance) utilize both the empirical distribution and the sample realizations. This has been used successfully to construct ambiguity sets in the literature.[29,30] Future work could involve investigating whether ambiguity sets constructed using the Wasserstein distance outperform those constructed using the L_2-norm or K-L in the DRO design formulation.

In this thesis we considered the setting in which the designer has access to a sample of realized values of the uncertain parameters. Another avenue for future work is to extend the distributionally robust design formulation to alternative settings. For example, the formulation could be applied to the situation where the designer has knowledge about the moments of the distribution of uncertain parameters. This would likely involve constructing the ambiguity sets based on moment constraints, which has also been explored previously in the literature.[13]

APPENDIX A

ADDITIONAL RESULTS FOR THE L_2-NORM AMBIGUITY SET

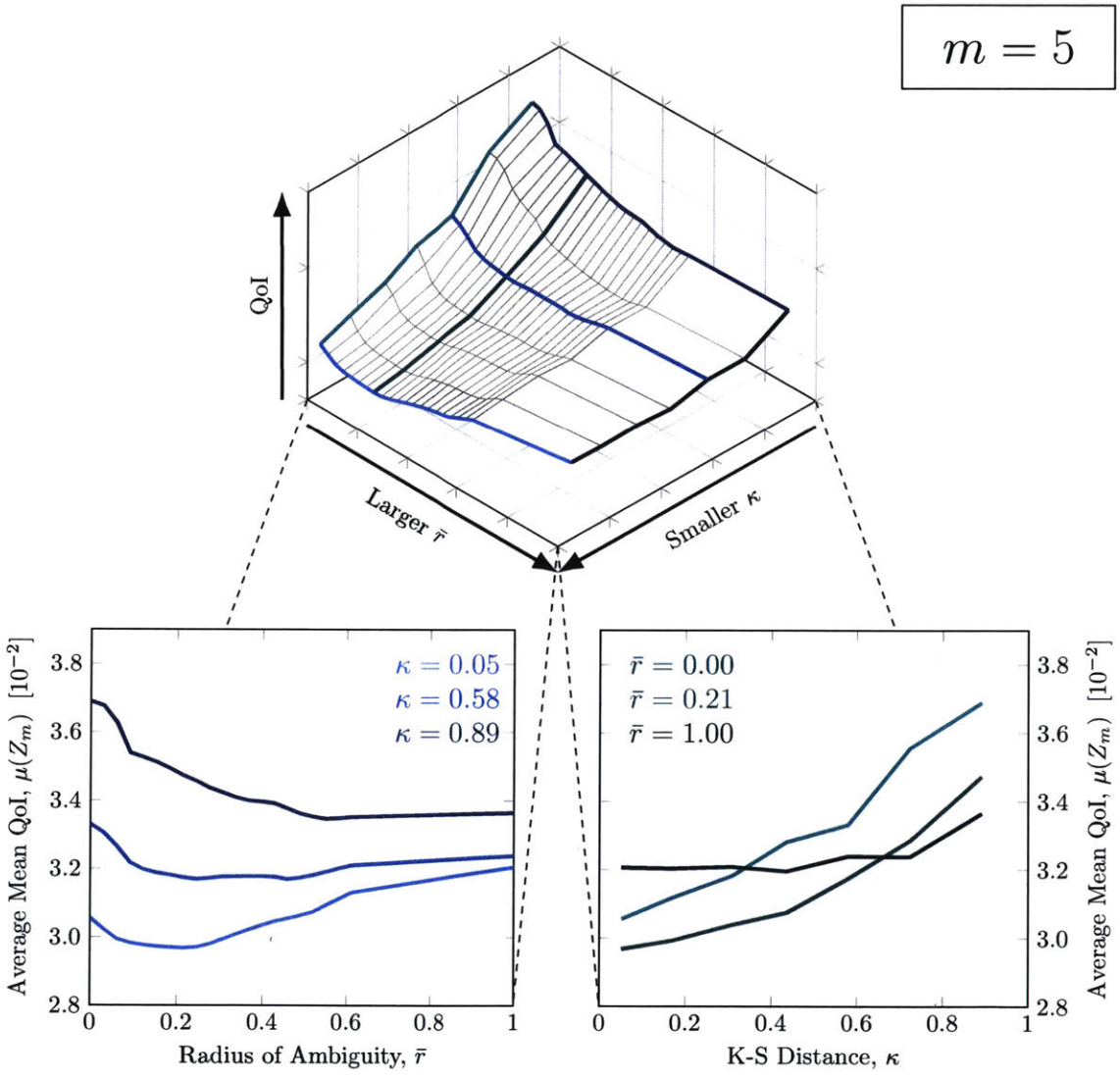
This appendix contains additional results, showcasing the DRO approach with L_2-norm ambiguity, applied to the acoustic horn design problem. These results are analogous to those presented in Sections 4.2 and 4.3 for the K-L divergence ambiguity set. For details on the experimental method behind these results, we refer the reader to those sections. In this appendix we highlight key similarities and differences in performance between the L_2-norm and K-L divergence ambiguity sets.

The next three figures show the relationship between the K-S distance of the sample, κ, the normalized radius of the L_2-norm ambiguity set, $\bar{r}$, and the performance of the resulting designs, in terms of the mean QoI over the uncertainty space. Figure A-1 shows the $m = 5$ case, Figure A-2 shows the $m = 10$ case, and Figure A-3 shows $m = 20$ results. In this study we suppose the uncertain variables are uniformly distributed (Eqn. 2.14). Note that the reason for the large jump in $\bar{r}$ shown in the data is that at the second highest value of $\bar{r}$, the ambiguity set already contains over 95% of the probability space Ω^m (over 99.9% in the $m = 10$ and $m = 20$ cases). Thus the result is not expected to change significantly between the two highest values of $\bar{r}$ shown.

The conclusions to be drawn are essentially the same as for the K-L divergence ambiguity. In each sample size studied, we see that increasing the ambiguity set a small amount results in improved performance compared with the $\bar{r} = 0$ (SAA) case.

After a critical value of $\bar{r}$ is reached, increasing it further results in worse performance. The critical value of $\bar{r}$ appears to decrease as the sample size m is increased. As a result, the potential for improvement in performance is greatest for small m, while the maximum penalty for exceeding the critical value of $\bar{r}$ is greatest for large m. In the case of the L_2-norm ambiguity set, the fall in performance after the critical radius *appears* steeper than for the K-L divergence ambiguity set. However, this is simply an effect of the parameterization of $\bar{r}$. Note that if we had not computed the results for $\bar{r} = 1$, and simply defined the second highest value of r to correspond to $\bar{r} = 1$, then the results would look almost identical to those computed using the K-L divergence ambiguity.

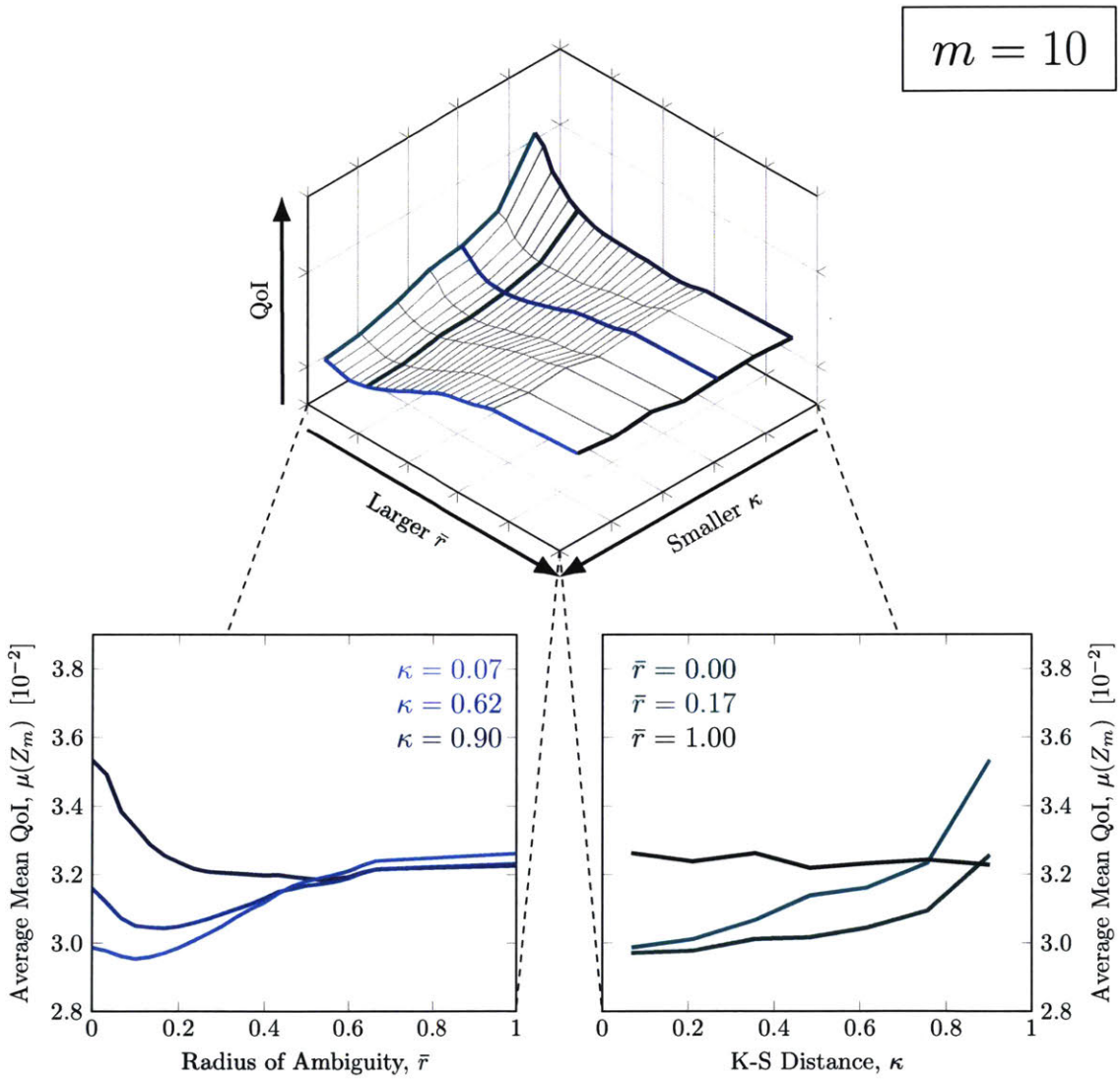
Figure A-4 compares the results found using uniformly and normally distributed uncertainties, using the L_2-norm ambiguity set. This figure is analogous to Figure 4-6, which showed similar results for the K-L divergence ambiguity. Again, we see that the results for the L_2-norm ambiguity set are similar to those of the K-L divergence ambiguity set. The maximum improvement gained by using the DRO approach over the SAA approach (corresponding to $r = 0$) appears to be smaller for normally distributed uncertainty. As was argued in Section 4.3, a wider study of different underlying distributions is needed to identify the cause of this effect.



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

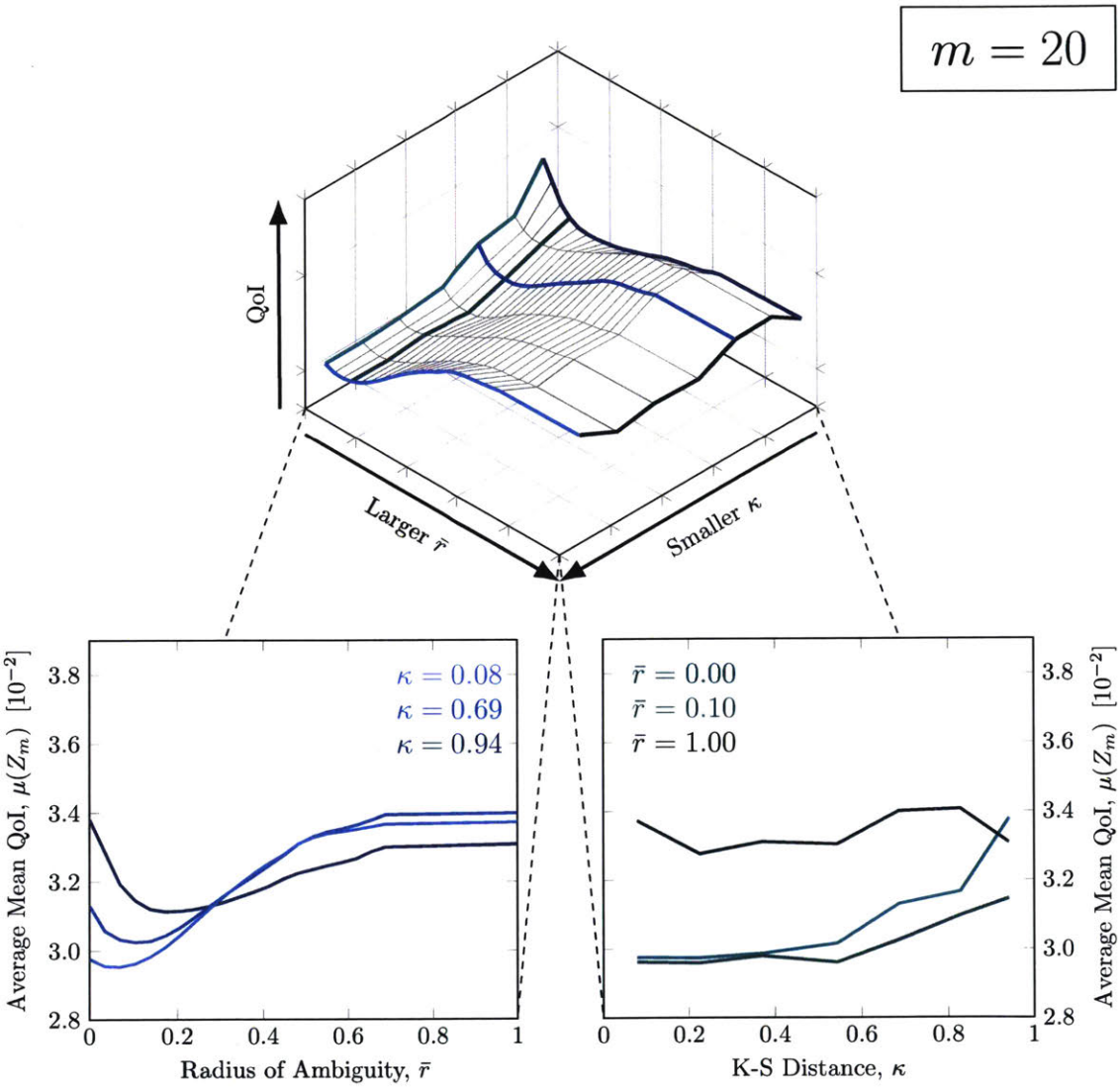
Figure A-1: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using uniformly distributed samples of size $m = 5$ and the DRO approach with a L_2-norm ambiguity set.*



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

Figure A-2: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using uniformly distributed samples of size $m = 10$ and the DRO approach with a L_2-norm ambiguity set.*



(a) Performance vs. Size of Ambiguity Set

(b) Performance vs. K-S distance of Sample

Figure A-3: *Surface plot showing the relationship between the K-S statistic of a sample, the ambiguity set size, and the resulting design performance using uniformly distributed samples of size $m = 20$ and the DRO approach with a L_2-norm ambiguity set.*

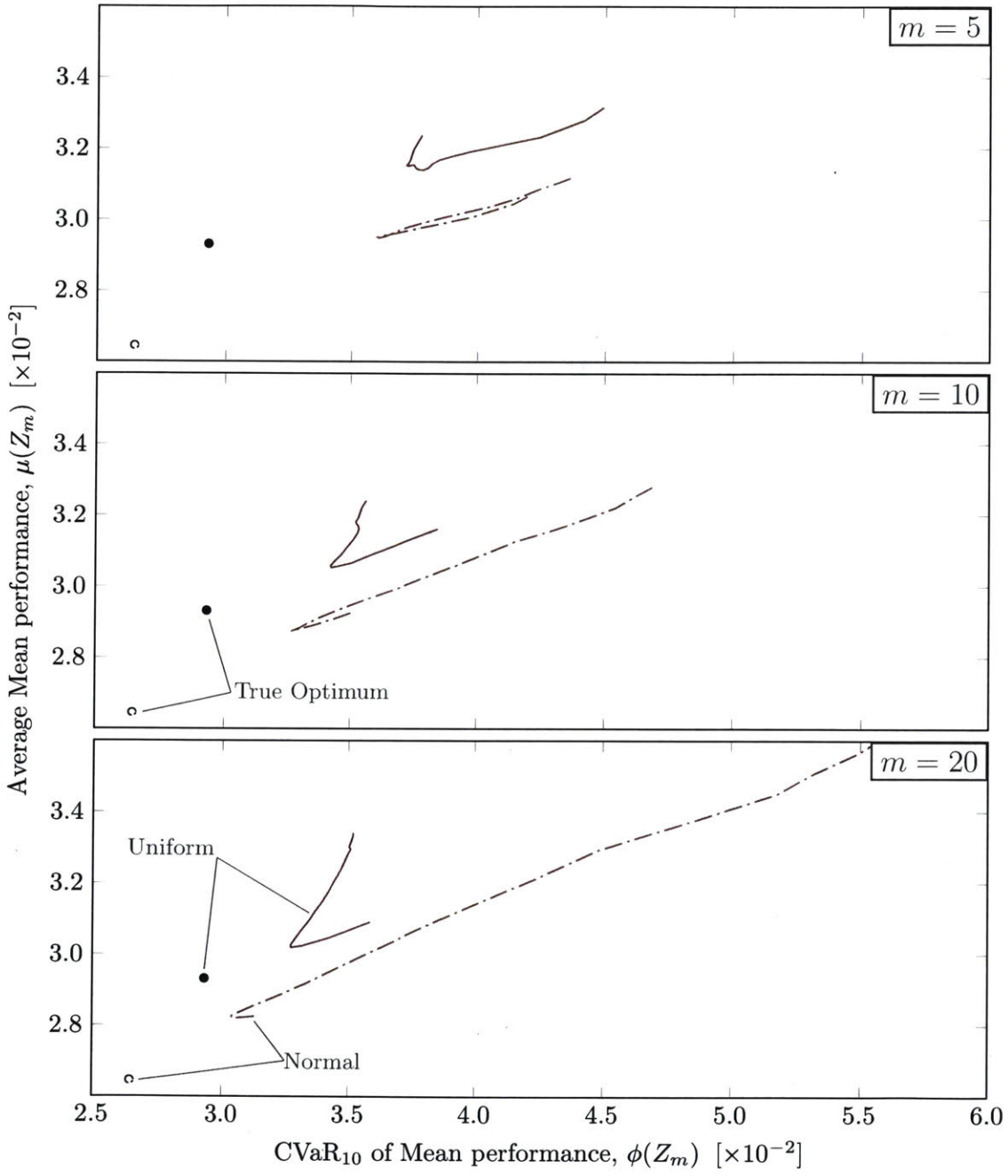


Figure A-4: *Comparison between mean-risk trade-off curves for designs computed using normally distributed and uniformly distributed uncertainty. Results are for sample sizes $m = 5, 10, 20$ and the DRO approach with L_2-norm ambiguity sets.*

BIBLIOGRAPHY

[1] Ng, L. W., Huynh, D. P., and Willcox, K., "Multifidelity Uncertainty Propagation for Optimization Under Uncertainty," *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2012, p. 5602.

[2] Kennedy, M. C. and O'Hagan, A., "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425–464.

[3] Beyer, H.-G. and Sendhoff, B., "Robust Optimization–A Comprehensive Survey," *Computer Methods in Applied Mechanics and Engineering*, Vol. 196, No. 33, 2007, pp. 3190–3218.

[4] Du, X. and Chen, W., "Efficient Uncertainty Analysis Methods for Multidisciplinary Robust Design," *AIAA journal*, Vol. 40, No. 3, 2002, pp. 545–581.

[5] Du, X. and Chen, W., "Methodology for Managing the Effect of Uncertainty in Simulation-Based Design," *AIAA Journal*, Vol. 38, No. 8, 2000, pp. 1471–1478.

[6] Keane, A. and Nair, P., *Computational Approaches for Aerospace Design: The Pursuit of Excellence*, John Wiley & Sons, 2005.

[7] Helton, J. C., Johnson, J. D., and Oberkampf, W. L., "An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions," *Reliability Engineering & System Safety*, Vol. 85, No. 1, 2004, pp. 39–71.

[8] Rao, S. and Cao, L., "Optimum Design of Mechanical Systems Involving Interval Parameters," *Transactions-American Society of Mechanical Engineers Journal of Mechanical Design*, Vol. 124, No. 3, 2002, pp. 465–472.

[9] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A., *Robust Optimization*, Princeton University Press, 2009.

[10] Zaman, K., Rangavajhala, S., McDonald, M. P., and Mahadevan, S., "A Probabilistic Approach for Representation of Interval Uncertainty," *Reliability Engineering & System Safety*, Vol. 96, No. 1, 2011, pp. 117–130.

[11] Smets, P., "Probability, Possibility, Belief: Which and Where?" *Quantified Representation of Uncertainty and Imprecision*, Springer, 1998, pp. 1–24.

[12] Goh, J. and Sim, M., "Distributionally Robust Optimization and its Tractable Approximations," *Operations Research*, Vol. 58, No. 4-part-1, 2010, pp. 902–917.

[13] Delage, E. and Ye, Y., "Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems," *Operations research*, Vol. 58, No. 3, 2010, pp. 595–612.

[14] Wiesemann, W., Kuhn, D., and Sim, M., "Distributionally Robust Convex Optimization," *Operations Research*, Vol. 62, No. 6, 2014, pp. 1358–1376.

[15] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G., "Robust Solutions of Optimization Problems Affected by Uncertain Probabilities," *Management Science*, Vol. 59, No. 2, 2013, pp. 341–357.

[16] Van Parys, B. P., Esfahani, P. M., and Kuhn, D., "From Data to Decisions: Distributionally Robust Optimization is Optimal," *arXiv preprint arXiv:1704.04118*, 2017.

[17] Hu, Z. and Hong, L. J., "Kullback-Leibler Divergence Constrained Distributionally Robust Optimization," *Optimization Online*, 2013.

[18] Calafiore, G. C., "Ambiguous Risk Measures and Optimal Robust Portfolios," *SIAM Journal on Optimization*, Vol. 18, No. 3, 2007, pp. 853–877.

[19] Ng, L. W. and Willcox, K. E., "Multifidelity Approaches for Optimization Under Uncertainty," *International Journal for Numerical Methods in Engineering*, Vol. 100, No. 10, 2014, pp. 746–772.

[20] Eftang, J. L., Huynh, D., Knezevic, D. J., and Patera, A. T., "A Two-step Certified Reduced Basis Method," *Journal of Scientific Computing*, Vol. 51, No. 1, 2012, pp. 28–58.

[21] Udawalpola, R. and Berggren, M., "Optimization of an Acoustic Horn with Respect to Efficiency and Directivity," *International Journal for Numerical Methods in Engineering*, Vol. 73, No. 11, 2008, pp. 1571–1606.

[22] "MATLAB Optimization Toolbox," Version 7.6, MATLAB 2017a, MathWorks, Natick, MA, USA.

[23] Xu, H., Caramanis, C., and Mannor, S., "A Distributional Interpretation of Robust Optimization," *Mathematics of Operations Research*, Vol. 37, No. 1, 2012, pp. 95–110.

[24] Philpott, A., de Matos, V., and Kapelevich, L., "Distributionally Robust SDDP," *Preprint available at http://www.epoc.org.nz/papers/DROPaperv52.pdf*.

[25] Grant, M. and Boyd, S., "CVX: Matlab Software for Disciplined Convex Programming, Version 2.1," `http://cvxr.com/cvx`, March 2014.

[26] Grant, M. and Boyd, S., "Graph Implementations for Nonsmooth Convex Programs," *Recent Advances in Learning and Control*, edited by V. Blondel, S. Boyd, and H. Kimura, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110, `http://stanford.edu/~boyd/graph_dcp.html`.

[27] Nilim, A. and El Ghaoui, L., "Robust Control of Markov Decision Processes with Uncertain Transition Matrices," *Operations Research*, Vol. 53, No. 5, 2005, pp. 780–798.

[28] Filippi, S., Cappé, O., and Garivier, A., "Optimism in Reinforcement Learning and Kullback-Leibler divergence," *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, IEEE, 2010, pp. 115–122.

[29] Esfahani, P. M. and Kuhn, D., "Data-Driven Distributionally Robust Optimization using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations," *arXiv preprint arXiv:1505.05116*, 2015.

[30] Shafieezadeh-Abadeh, S., Esfahani, P. M., and Kuhn, D., "Distributionally Robust Logistic Regression," *Advances in Neural Information Processing Systems*, 2015, pp. 1576–1584.