Submitted, SIAM Journal on Optimization (SIOPT)
http://www.cds.caltech.edu/~murray/papers/han+13-siopt.html

Convex Optimal Uncertainty Quantification

Shuo Han*[1], Molei Tao[2], Ufuk Topcu[1], Houman Owhadi[3], and Richard M. Murray[3]

[1]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104
[2]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012
[3]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125

November 28, 2013

Abstract

Optimal uncertainty quantification (OUQ) is a framework for numerical extreme-case analysis of stochastic systems with imperfect knowledge of the underlying probability distribution and functions/events. This paper presents sufficient conditions (when underlying functions are known) under which an OUQ problem can be reformulated as a finite-dimensional convex optimization problem.

Key words. convex optimization, uncertainty quantification, duality theory

AMS subject classifications. 90C25, 90C46, 90C15, 60-08

1 Introduction

In many applications, given a cost function $f\colon \mathbb{R}^d \to \mathbb{R}$ that depends on a random variable θ, we would like to compute $\mathbb{E}_{\theta\sim\mathcal{D}}[f(\theta)]$, where $\mathcal{D}$ is the probability distribution of θ. If $\mathcal{D}$ is known exactly, this amounts to a numerical integration problem. However, sometimes we only have access to

*E-mail: hanshuo@seas.upenn.edu. This work was performed while the author was with the California Institute of Technology.

partial information (e.g., moments) about $\mathcal{D}$. This information imposes constraints on $\mathcal{D}$, which we will refer to as *information constraints* throughout the paper. With only information constraints, it is generally impossible to compute the exact value of $\mathbb{E}_{\theta\sim\mathcal{D}}[f(\theta)]$. Instead, we can only hope to obtain a lower or upper bound of $\mathbb{E}_{\theta\sim\mathcal{D}}[f(\theta)]$.

Unfortunately, such bounds are unavailable in closed form except for several special cases, where the bound can be obtained from probability inequalities. To this end, this paper focuses on computing these bounds numerically by solving infinite-dimensional optimization problems over the set of probability distributions that satisfy the information constraints. We follow Owhadi et al. [16] and refer to this problem as *optimal uncertainty quantification* (OUQ) for convenience (note that the actual OUQ framework is more general and is capable of dealing with unknown functions f in additional to unknown probability distributions). In certain cases, it is also called *(generalized) moment problem* or *problem of moments*, since the information constraints often consist of moments of the distribution or can be considered as generalized moments of the distribution.

Although infinite-dimensional, a large class of OUQ problems can be reduced to equivalent finite-dimensional optimization problems having the same optimal value [16]. This reduction operates in several steps and the first one is a generalization of linear programming in spaces of measures [18, 19]. Although this reduction permits the numerical resolution of OUQ problems, these problems, in their reduced form, may be highly constrained and non-convex.

The paper focuses on the first reduction step (over measures, when underlying functions are known) and presents sufficient conditions (Theorem 9) under which an OUQ problem can be reduced to a finite-dimensional *convex* optimization problem. Section 2 introduces the mathematical formulation of OUQ and recalls previous results on equivalent finite-dimensional reduction of OUQ problems when underlying functions are known (and in the absence of independence constraints). Section 3 and 4 present the main results on convex reformulation of OUQ from the primal and the dual form of the original OUQ problem, respectively. Section 5 provides numerical illustrations of the main theoretical result.

Historical Perspective. Among various convex formulations of OUQ, one important special case is when

$$f = I(\theta \in C), \tag{1.1}$$

where $C \subseteq \mathbb{R}^d$ and I is the 0-1 indicator function. Solution to the OUQ problem will yield a sharp bound of the probability $\mathbb{P}(\theta \in C)$ under the given information constraints. This bound is also called Chebyshev-type bound or generalized Chebyshev bound. The earliest theoretical analysis of such bounds can be traced back to the pioneering work by Chebyshev and his student Markov (see Krein [11] for an account of the history of this subject along with substantial contributions by Krein). We also refer to early work by Isii [7, 8, 9] and Marshall and Olkin [13].

As related in Owhadi and Scovel [15, Section 2], OUQ starts from the same mindset and applies it to more complex problems that extend the base space to functions and measures. Instead of developing sophisticated mathematical solutions, OUQ develops optimization problems and reductions, so that their solution may be implemented on a computer, as in Bertsimas and Popescu's [3] convex optimization approach to Chebyshev inequalities, and the decision analysis framework of Smith [20].

Recent works on convex optimization approach to Chebyshev inequalities (motivated by efficient numerical methods such as the interior-point method [14]) also include Lasserre [12], Popescu [17] and Vandenberghe et al. [21] for convex formulations for different classes of sets C in (1.1) (ellipsoids, semi-algebraic sets, etc.).

Besides indicator set functions (1.1), another class of functions f that appear in convex formulations are functions that are both convex and piecewise affine:

$$f(\theta) = \max_{k=1,2,\ldots,K} \{a_k^T \theta + b_k\}, \tag{1.2}$$

where K and $\{a_k, b_k\}_{k=1}^K$ are given constants. This form arises in applications such as stock investment [6] and logistics [2]. It also emerges when f is obtained as the optimal value of a linear program.

Besides above conditions on f, convex formulations have incorporated information constraints in the form of moment constraints (oftentimes these constraints are limited to the mean and covariance). It is known that the feasibility of moment constraints can be represented as a linear matrix inequality on a Hankel matrix consisting of the given moments [5, page 170], which appears in most convex formulations with moment constraints. This allows OUQ problems (over measures without independence constraints) to be cast as semidefinite programs.

This paper shows that such convex formulations can be achieved for the broader class of piecewise concave functions f (this class contains both the 0-1 indicator function and functions of the form (1.2)) and the broader class of information constraints defined by piecewise convex functions (including certain categories of moment constraints).

2 Optimal uncertainty quantification and finite reduction

Formally, an example of OUQ problems (with known underlying functions and without independence constraints) is an optimization problem of the form:

$$\underset{\mathcal{D}}{\text{maximize}} \quad \mathbb{E}_{\theta\sim\mathcal{D}}\left[f(\theta)\right] \tag{2.1}$$

$$\text{subject to} \quad \mathbb{E}_{\theta\sim\mathcal{D}}[g(\theta)] \preceq 0 \tag{2.2}$$

$$\mathbb{E}_{\theta\sim\mathcal{D}}[h(\theta)] = 0 \tag{2.3}$$

$$\theta \in \Theta \quad \text{almost surely,} \tag{2.4}$$

where θ is a random variable in $\mathbb{R}^d$, and $\mathcal{D}$ is the probability distribution of θ. The function f is real-valued, and g and h are (real) vector-valued in general. The inequality (2.2) denotes entry-wise inequality. The set $\Theta \subseteq \mathbb{R}^d$ is the support of the distribution. For brevity, the notion "almost surely" is dropped later in this paper. Note that the condition that $\mathcal{D}$ is a probability distribution implicitly imposes the constraints:

$$\mathbb{E}_{\theta\sim\mathcal{D}}[1] = 1, \qquad \mathcal{D} \geq 0. \tag{2.5}$$

The functions g and h are used to incorporate available information on $\mathcal{D}$. For example, when g and h consist of powers of θ, it implies that information on the moments of θ is available. Problem (2.1) can also be written in a form without the equality constraint (2.3) and support constraint (2.4). The equality constraint (2.3) can be eliminated by introducing the following inequality constraints:

$$\mathbb{E}_{\theta\sim\mathcal{D}}[h(\theta)] \preceq 0, \qquad \mathbb{E}_{\theta\sim\mathcal{D}}[-h(\theta)] \preceq 0.$$

The support constraint can be shown as equivalent to

$$\mathbb{E}_{\theta\sim\mathcal{D}}[I(\theta \notin \Theta)] \leq 0, \tag{2.6}$$

where I is the 0-1 indicator function:

$$I(E) = \begin{cases} 1 & E = \text{true} \\ 0 & E = \text{false}. \end{cases}$$

In order to prove this, we note that

$$\mathbb{E}_{\theta\sim\mathcal{D}}[I(\theta \notin \Theta)] \geq 0$$

automatically holds due to the fact that I is non-negative. Therefore, the condition (2.6) is equivalent to

$$\mathbb{E}_{\theta\sim\mathcal{D}}[I(\theta \notin \Theta)] = 0,$$

which is equivalent to (2.4). However, we will still use the original form in problem (2.1) to distinguish these constraints from pure inequalities. Throughout the paper, the constraints (2.2)–(2.4) are referred to as *information constraints.*

Interestingly, many inequalities in probability theory can be viewed as OUQ problems. One such example is Markov's inequality, which has its origin in the following problem [11, Page 4] (according to Krein [11], although Chebyshev did solve this problem, it was his student Markov who supplied the proof in his thesis):

> "Given: length, weight, position of the centroid and moment of inertia of a material rod with a density varying from point to point. It is required to find the most accurate limits for the weight of a certain segment of this rod."

Although the statement of the problem assumes knowledge about both the first and second moments (centroid and moment of inertia), Markov has also worked on a similar problem that only assumes knowledge about the first moment, which is given below.

Example 1 (Markov's inequality)**.** Suppose θ is a nonnegative univariate random variable whose probability distribution is unknown, but its mean $\mathbb{E}[\theta] = \mu$ is given. For any $a > 0$, Markov's inequality gives a bound for $\mathbb{P}(\theta \geq a)$ regardless of the probability distribution:

$$\mathbb{P}(\theta \geq a) \leq \mu/a.$$

In the OUQ framework, the problem of obtaining a tight bound for $\mathbb{P}(\theta \geq a)$ becomes

$$\begin{aligned} \underset{\mathcal{D}}{\text{maximize}} \quad & \mathbb{E}_{\theta\sim\mathcal{D}}[I(\theta \geq a)] \\ \text{subject to} \quad & \mathbb{E}_{\theta\sim\mathcal{D}}[\theta] = \mu \\ & \theta \in [0, \infty). \end{aligned}$$

In fact, it can be shown that the optimal value of the above problem is μ/a. Namely, Markov's inequality produces a tight bound.

The perhaps surprising fact is that an OUQ problem can always be reduced to an equivalent finite-dimensional optimization problem that yields the same optimal value. The following theorem is originally due to Rogosinski [18]. See also Shapiro [19] for an extension to conic linear programs and Owhadi et al. [16] for a more general result that allows independence constraints.

Theorem 2 (Finite reduction property, cf. [16, 18, 19])**.** *Let m be the total number of scalar (in)equalities described by g and h in (2.2) and (2.3). The finite-dimensional problem*

$$\begin{aligned}
\underset{\{p_i,\theta_i\}_{i=1}^{m+1}}{\text{maximize}} \quad & \sum_{i=1}^{m+1} p_i f(\theta_i) \\
\text{subject to} \quad & \sum_{i=1}^{m+1} p_i = 1 \\
& p_i \geq 0, \quad i = 1, 2, \ldots, m+1 \\
& \sum_{i=1}^{m+1} p_i g(\theta_i) \preceq 0, \qquad \sum_{i=1}^{m+1} p_i h(\theta_i) = 0 \\
& \theta_i \in \Theta, \quad i = 1, 2, \ldots, m+1
\end{aligned} \tag{2.7}$$

achieves the same optimal value as problem (2.1).

Theorem 2 implies that the optimal value of any OUQ problem can always be achieved by a discrete probability distribution, whose number of Dirac masses depends on the information constraints. In the following, we give a simple example (reproduced from Owhadi et al. [16]) to illustrate this property.

Example 3 ("Seesaw", see also [16])**.** Consider the following OUQ problem for a scalar random variable θ:

$$\begin{aligned}
\underset{\mathcal{D}}{\text{maximize}} \quad & \mathbb{P}(\theta \geq \gamma) \\
\text{subject to} \quad & \mathbb{E}_{\theta\sim\mathcal{D}}[\theta] = 0, \qquad a \leq \theta \leq b,
\end{aligned}$$

where a, b, and γ are constants satisfying $a < 0 \leq \gamma < b$. In order to maximize $\mathbb{P}(\theta \geq \gamma)$, we would want to assign as much probability as possible within $[\gamma, b]$. On the other hand, the condition $\mathbb{E}_{\theta\sim\mathcal{D}}[\theta] = 0$ requires that the probability within $[a, 0]$ and that within $[0, b]$ must be identical. This condition is analogous to a seesaw pivoted at 0 with two end points at a

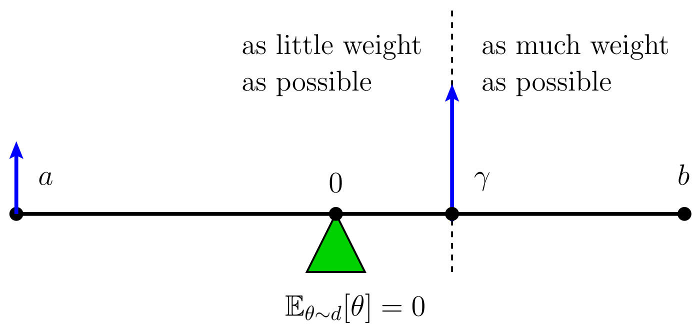

Figure 2.1: "Seesaw" analogy in Example 3 (reproduced from [16]).

and b (Figure 2.1). It is not difficult to see that the best assignment is to put all the probability on the right side at γ (for least leverage) and all the probability on the left side at a (for most leverage). This assignment implies that the optimal distribution can be achieved with a discrete distribution consisting of 2 Dirac masses at a and γ. Indeed, since there is only one scalar constraint, the total number of Dirac masses predicted by Theorem 2 is $1 + 1 = 2$.

3 Convex optimal uncertainty quantification: Primal form

We will now show that the optimization problem (2.1) can be reduced to a finite-dimensional convex problem if the following conditions hold for f, g, and h.

1. The function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *piecewise concave*, i.e., it can be written as

$$f(\theta) = \max_{k=1,2,\ldots,K} f^{(k)}(\theta), \tag{3.1}$$

   where each function $f^{(k)}$ is concave.

2. The function $g\colon \mathbb{R}^d \to \mathbb{R}^p$ is entry-wise *piecewise convex*, i.e., each entry g_i ($i = 1, 2, \ldots, p$) can be written as

$$g_i(\theta) = \min_{l_i=1,2,\ldots,L_i} g_i^{(l_i)}(\theta), \tag{3.2}$$

   where each function $g_i^{(l_i)}$ is convex.

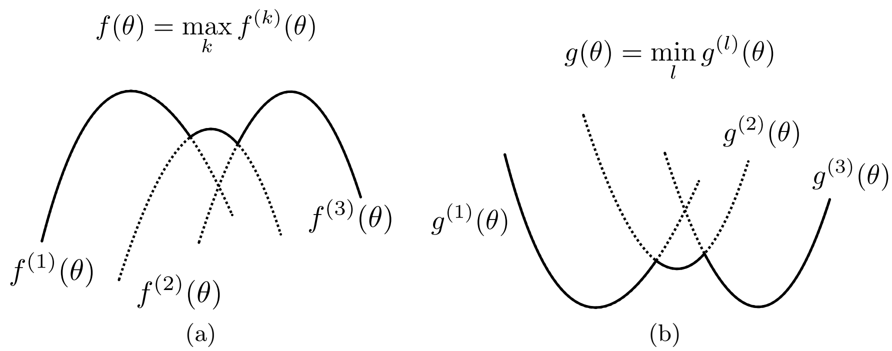

Figure 3.1: (a) Piecewise concave and (b) piecewise convex functions in dimension one.

3. The function $h\colon \mathbb{R}^d \to \mathbb{R}^q$ is affine. Namely, it can be represented as

$$h(\theta) = A^T\theta + b, \tag{3.3}$$

for appropriate choices of $A \in \mathbb{R}^{d\times q}$ and $b \in \mathbb{R}^q$.

Figure 3.1 illustrates how piecewise concave and piecewise convex functions look like in dimension one. In general, these functions are neither concave nor convex. Nevertheless, we list in the following several useful examples of piecewise concave/convex functions. We begin by examples of piecewise concave functions that are potentially useful as objective functions.

Example 4 (Convex and piecewise affine)**.** When each $f^{(k)}$ is affine (hence concave) for each k, the function $f = \max_{k=1,2,\ldots,K}\{f^{(k)}\}$ becomes convex and piecewise affine. This class of functions has been studied by, e.g., Delage and Ye [6] and Bertsimas et al. [2].

Example 5. Consider $f(\theta) = I(\theta \in C)$ for any convex set $C \subseteq \mathbb{R}^d$. This can be used to specify the probability $\mathbb{P}(\theta \in C)$, since $\mathbb{P}(\theta \in C) = \mathbb{E}[I(\theta \in C)]$. The function f can be rewritten as

$$f(\theta) = \max\{0, I_{-\infty}^{1}(\theta \in C)\},$$

where the function I_b^a is defined as

$$I_b^a(E) = \begin{cases} a & E = \text{true} \\ b & E = \text{false} \end{cases}$$

for any $a, b \in \mathbb{R}$. It can be verified that both 0 and $I_{-\infty}^{1}(\theta \in C)$ are concave in θ.

Next we list several examples of piecewise convex functions that can be used as information constraints.

Example 6 (Even-order moments)**.** The random variable θ is univariate and $g(\theta) = \theta^{2q}$ for some positive integer q. This form is a special case where the function g itself is convex.

Example 7. Consider $g(\theta) = I(\theta \notin C)$ for any convex set $C \in \mathbb{R}^d$. The function g can be rewritten as

$$g(\theta) = \min\{1, I_\infty^0(\theta \in C)\}.$$

It can be verified that both functions inside the minimization are convex in θ.

We will now show that, under conditions (3.1)–(3.3), an OUQ problem can be reformulated as a finite-dimensional convex optimization problem. For notational simplicity, we will only present the case when g is defined from $\mathbb{R}^d$ to $\mathbb{R}$ (i.e., $p = 1$):

$$g(\theta) = \min_{l=1,2,\ldots,L} g^{(l)}(\theta), \qquad g^{(l)} \text{ is convex.} \tag{3.4}$$

The proof can be easily generalized to the case of $p > 1$.

According to the finite reduction property, it suffices to use a finite number of Dirac masses to represent the optimal distribution. In particular, due to the special form of the objective function and constraints, these Dirac masses satisfy a useful property as given in the following lemma.

Lemma 8. *Suppose f, g, and h satisfy the conditions (3.1), (3.4), and (3.3), respectively. Then the optimal distribution for problem (2.1) can be achieved by a discrete distribution consisting of at most $K \cdot L$ Dirac masses located at $\{\theta_{kl}\}$ ($k = 1, 2, \ldots, K$; $l = 1, 2, \ldots, L$). In addition, each θ_{kl} achieves maximum at $f^{(k)}$ and minimum at $g^{(l)}$:*

$$f(\theta_{kl}) = f^{(k)}(\theta_{kl}), \qquad g(\theta_{kl}) = g^{(l)}(\theta_{kl}).$$

Proof. Suppose that, for certain k and l, the optimal distribution contains two Dirac masses located at ϕ_1 and ϕ_2 with probabilities q_1 and q_2, respectively, whereas both ϕ_1 and ϕ_2 achieve maximum at $f^{(k)}$ and minimum at $g^{(l)}$, i.e.,

$$\begin{aligned} f(\phi_1) &= f^{(k)}(\phi_1), \qquad & f(\phi_2) &= f^{(k)}(\phi_2), \\ g(\phi_1) &= g^{(l)}(\phi_1), \qquad & g(\phi_2) &= g^{(l)}(\phi_2). \end{aligned}$$

Consider a new Dirac mass whose probability q and location ϕ are given by

$$q = q_1 + q_2, \qquad \phi = \frac{q_1\phi_1 + q_2\phi_2}{q_1 + q_2}.$$

It can be verified that replacing the two previous Dirac masses (q_1, ϕ_1) and (q_2, ϕ_2) with this new Dirac mass (q, ϕ) will still yield a valid probability distribution (i.e., the probability masses sum up to 1). Moreover, the new distribution will give an objective $\mathbb{E}[f(\theta)]$ that is no smaller than the previous one, since

$$qf(\phi) \geq qf^{(k)}(\phi) \geq q_1 f^{(k)}(\phi_1) + q_2 f^{(k)}(\phi_2) = q_1 f(\phi_1) + q_2 f(\phi_2), \tag{3.5}$$

where the second inequality is an application of Jensen's inequality and last equality uses the fact that ϕ_1 and ϕ_2 achieve maximum at $f^{(k)}$.

On the other hand, it can be shown that the new distribution will remain as a feasible solution. The equality constraint on $\mathbb{E}[h(\theta)]$ remains feasible, because

$$\begin{aligned} qh(\phi) &= q(A^T\phi + b) = A^T(q_1 + q_2)\phi + b(q_1 + q_2) \\ &= A^T(q_1\phi_1 + q_2\phi_2) + b(q_1 + q_2) = q_1 h(\phi_1) + q_2 h(\phi_2). \end{aligned}$$

The feasibility of the inequality constraint on $\mathbb{E}[g(\theta)]$ can be proved by using a similar argument as in (3.5) by observing that $\mathbb{E}[g(\theta)]$ evaluated under the new distribution will be no larger than that under the original distribution, because

$$qg(\phi) \leq qg^{(l)}(\phi) \leq q_1 g^{(l)}(\phi_1) + q_2 g^{(l)}(\phi_2) = q_1 g(\phi_1) + q_2 g(\phi_2).$$

Therefore, the two old Dirac masses can be replaced by the new single one without affecting optimality, from which the uniqueness of θ_{kl} follows. $\square$

The number of Dirac masses given by Lemma 8 depends only on K and L, and is independent from that given by the finite reduction property. By using Lemma 8, we can obtain an equivalent convex optimization problem for the original problem (2.1), as given in the following theorem.

Theorem 9. *The (convex) optimization problem*

$$\begin{aligned}
\underset{\{p_{kl},\gamma_{kl}\}_{k,l}}{\text{maximize}} \quad & \sum_{k,l} p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) && (3.6)\\
\text{subject to} \quad & \sum_{k,l} p_{kl} = 1 && (3.7)\\
& p_{kl} \geq 0, \quad \forall k, l && (3.8)\\
& \sum_{k,l} p_{kl} h(\gamma_{kl}/p_{kl}) = A^T \left(\sum_{k,l} \gamma_{kl} \right) + b = 0 \\
& \sum_{k,l} p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) \leq 0 && (3.9)
\end{aligned}$$

achieves the same optimal value as problem (2.1) if the functions f, g, and h satisfy (3.1), (3.4), and (3.3), respectively.

Proof. According to Lemma 8, we can optimize over a new set of Dirac masses whose probability weights and locations are $\{p_{kl}, \theta_{kl}\}$. The requirement that the set of Dirac masses forms a valid probability distribution imposes the constraints (3.7) and (3.8). Under the new set of Dirac masses, the objective function can be rewritten as

$$\mathbb{E}\left[f(\theta)\right] = \sum_{k,l} p_{kl} f(\theta_{kl}) = \sum_{k,l} p_{kl} f^{(k)}(\theta_{kl}),$$

where the second equality uses the fact that θ_{kl} achieves maximum at $f^{(k)}$. As will be shown later, this step is critical since f is generally not concave, but $\sum_{k,l} p_{kl} f^{(k)}(\theta_{kl})$ is concave. Similarly, we have

$$\mathbb{E}[g(\theta)] = \sum_{k,l} p_{kl} g(\theta_{kl}) = \sum_{k,l} p_{kl} g^{(l)}(\theta_{kl}).$$

The final form can be obtained by introducing new variables $\gamma_{kl} = p_{kl}\theta_{kl}$ for all k and l and choosing to optimize over $\{p_{kl}, \gamma_{kl}\}$ instead of $\{p_{kl}, \theta_{kl}\}$. Each term in the sum in the objective function

$$\sum_{k,l} p_{kl} f^{(k)}(\gamma_{kl}/p_{kl})$$

is a perspective transform of a concave function $f^{(k)}$ and hence is concave [5, page 39]. Therefore, the objective function is concave. Likewise, the term

$$\sum_{k,l} p_{kl} g^{(l)}(\gamma_{kl}/p_{kl})$$

is convex, and corresponding constraint (3.9) is also convex. All other constraints are affine and do not affect convexity. In conclusion, the final optimization problem is a finite-dimensional convex problem and is equivalent to the original problem (2.1) due to Lemma 8. □

Meanwhile, there are a couple straightforward extensions to Theorem 9.

Multiple inequality constraints. Lemma 8 and Theorem 9 can be generalized to the case of $p > 1$ based on a similar proof, except that the number of Dirac masses becomes $K \cdot \prod_{i=1,2,\dots,p} L_i$. It can be shown that, among all Dirac masses, there is at most one Dirac mass θ^* that achieves maximum at $f^{(k)}$ and minimum at $g_1^{(l_1)}, g_2^{(l_2)}, \dots, g_p^{(l_p)}$ for any given indices k and $\{l_i\}_{i=1}^p$:

$$f(\theta^*) = f^{(k)}(\theta^*), \quad g_1(\theta^*) = g_1^{(l_1)}(\theta^*), \quad \dots, \quad g_p(\theta^*) = g_p^{(l_p)}(\theta^*).$$

The corresponding convex optimization problem can be formed by following a similar procedure as in the proof of Theorem 9.

Support constraints. It is possible to impose certain types of constraints on the support Θ without affecting convexity. Specifically, we can allow

$$\Theta = \bigcup_{s=1}^{S} C^{(s)}, \tag{3.10}$$

where each $C^{(s)} \subseteq \mathbb{R}^d$ is a convex set. In order to show that the corresponding OUQ problem remains convex, we use the fact that the support constraint (2.4) is equivalent to

$$\mathbb{E}[I(\theta \notin \Theta)] \leq 0, \tag{3.11}$$

as presented at the beginning of Section 2. When Θ satisfies (3.10), we have

$$I(\theta \notin \Theta) = \min\{1, I_\infty^0(\theta \in C^{(1)}), \dots, I_\infty^0(\theta \in C^{(S)}\}, \tag{3.12}$$

which is piecewise convex.

If the support constraint $\theta \in \Theta$ (or equivalently (3.11)) is added to an OUQ problem where f and g satisfy (3.1) and (3.4), the number of Dirac masses becomes $K \cdot L \cdot (S+1)$. Denote these Dirac masses as $\{p_{kl}, \gamma_{kl}\}_{k,l}$ and $\{p_{kls}, \gamma_{kls}\}_{k,l,s}$ ($k = 1, 2, \dots, K$; $l = 1, 2, \dots, L$; $s = 1, 2, \dots, S$). Then inequality constraint (3.11) becomes

$$\sum_{k,l} \left[p_{kl} + \sum_{s=1}^{S} p_{kls} I_\infty^0(\gamma_{kls}/p_{kls} \in C^{(s)}) \right] \leq 0,$$

where the first term appears due to the constant 1 in (3.12). Recall that $p_{kl} \geq 0$ for all k and l. Then we have

$$p_{kl} = 0, \qquad \gamma_{kls}/p_{kls} \in C^{(s)}, \quad s = 1, 2, \ldots, S \tag{3.13}$$

for all k and l. The fact $p_{kl} = 0$ implies that the actual number of Dirac masses needed in this case is $K \cdot L \cdot S$.

In particular, when each $C^{(s)}$ is a convex polytope, the corresponding support constraint has a simpler form. It is known that any convex polytope can be represented as an intersection of affine halfspaces. Namely, each $C^{(s)}$ can be expressed as:

$$C^{(s)} = \{\theta \colon A^{(s)}\theta \preceq b^{(s)}\}$$

for certain constants $A^{(s)}$ and $b^{(s)}$ (cf. [4]). Then the support constraint (3.13) becomes affine constraints

$$A^{(s)}\gamma_{kls} \preceq p_{kls} b^{(s)}, \quad s = 1, 2, \ldots, S$$

for all k and l.

4 Convex optimal uncertainty quantification: Dual form

The same convex reformulation of OUQ can be derived from the Lagrange dual problem of (2.1). First of all, the Lagrangian of problem (2.1) can be written as

$$\begin{aligned} L = &\int f(\theta)\mathcal{D}(\theta)\, d\theta - \lambda^T \int g(\theta)\mathcal{D}(\theta)\, d\theta - \nu^T \int h(\theta)\mathcal{D}(\theta)\, d\theta \\ &+ \int \lambda_p(\theta)\mathcal{D}(\theta)\, d\theta + \mu\left(1 - \int \mathcal{D}(\theta)\, d\theta\right). \end{aligned}$$

The last two terms are due to the constraints (2.5). From the Lagrangian, the Lagrange dual can be derived as

$$\sup_{\mathcal{D}} L = \begin{cases} \mu & f(\theta) - \lambda^T g(\theta) - \nu^T h(\theta) + \lambda_p(\theta) - \mu = 0 \text{ for all } \theta \\ \infty & \text{otherwise.} \end{cases}$$

By including the conditions on the Lagrange multipliers, i.e.,

$$\lambda \succeq 0, \qquad \lambda_p(\theta) \geq 0, \quad \forall\theta,$$

we can obtain the the dual problem as

$$\begin{aligned}
\underset{\lambda,\nu,\mu}{\text{minimize}} \quad & \mu && (4.1)\\
\text{subject to} \quad & f(\theta) - \lambda^T g(\theta) - \nu^T h(\theta) - \mu \leq 0, \quad \forall\theta && (4.2)\\
& \lambda \succeq 0,
\end{aligned}$$

which is a linear program with an infinite number of constraints (also known as a semi-infinite linear program). Under standard constraint qualifications, as shown by Isii [7, 8, 9], strong duality holds so that we can solve the dual problem (4.1) to obtain the optimal value of problem (2.1). Analysis on strong duality can also be found in Karlin and Studden [10], Akhiezer and Krein [1], Smith [20], and Shapiro [19].

The inequality constraint (4.2) implies that the optimal solution $(\lambda^*, \nu^*, \mu^*)$ must satisfy

$$\mu^* = \max_{\theta} \left[f(\theta) - \lambda^{*T} g(\theta) - \nu^{*T} h(\theta) \right],$$

which allows us to eliminate the inequality constraint (4.2) and rewrite problem (4.1) as

$$\begin{aligned}
\underset{\lambda,\nu}{\text{minimize}} \quad & \max_{\theta} \left[f(\theta) - \lambda^T g(\theta) - \nu^T h(\theta) \right] && (4.3)\\
\text{subject to} \quad & \lambda \succeq 0.
\end{aligned}$$

As it turns out, Theorem 9 can also be proved from the dual form (4.3). Similar to Section 3, we will only prove for the case of $p = 1$ for notational convenience. First, we present the following lemma for later use in the proof.

Lemma 10. *Given a set of real-valued functions $\{f^{(k)}\}_{k=1}^K$, the optimal value of the optimization problem*

$$\begin{aligned}
\underset{\{p_k,\theta_k\}_{k=1}^K}{\text{maximize}} \quad & \sum_{k=1}^{K} p_k f^{(k)}(\theta_k) && (4.4)\\
\text{subject to} \quad & \sum_{k=1}^{K} p_k = 1, \qquad p_k \geq 0, \quad k = 1, 2, \ldots, K
\end{aligned}$$

is $\max_\theta \max_{k=1,2,\ldots,K}\{f^{(k)}(\theta)\}$.

Proof. Denote the optimal value of problem (4.4) as OPT and

$$\theta^* = \arg\max_{\theta} \max_{k=1,2,\ldots,K} \{f^{(k)}(\theta)\}, \qquad k^* = \arg \max_{k=1,2,\ldots,K} \{f^{(k)}(\theta^*)\}.$$

Then we have OPT $\geq f^{(k^*)}(\theta^*) = \max_\theta \max_{k=1,2,\ldots,K}\{f^{(k)}(\theta)\}$, since

$$p_k = \begin{cases} 1 & k = k^* \\ 0 & \text{otherwise,} \end{cases} \qquad \theta_k = \theta^*, \quad \forall k$$

is a feasible solution of problem (4.4), and its corresponding objective value is $f^{(k^*)}(\theta^*)$. On the other hand, suppose $\{p_k^*, \theta_k^*\}_{k=1}^K$ is the optimal solution of problem (4.4). Then we have

$$\begin{aligned} \text{OPT} &= \sum_{k=1}^K p_k^* f^{(k)}(\theta_k^*) \leq \sum_{k=1}^K \left[p_k^* \max_{k=1,2,\ldots,K} \left\{ f^{(k)}(\theta_k^*) \right\} \right] \\ &= \left(\sum_{k=1}^K p_k^* \right) \cdot \max_{k=1,2,\ldots,K} \left\{ f^{(k)}(\theta_k^*) \right\} = \max_{k=1,2,\ldots,K} \left\{ f^{(k)}(\theta_k^*) \right\} \\ &\leq \max_{k=1,2,\ldots,K} \left\{ \max_\theta f^{(k)}(\theta) \right\} = \max_\theta \max_{k=1,2,\ldots,K} \left\{ f^{(k)}(\theta) \right\}. \end{aligned}$$

Therefore, we have OPT $= \max_\theta \max_{k=1,2,\ldots,K}\{f^{(k)}(\theta)\}$. □

We are now ready to prove Theorem 9 from the dual problem (4.3).

Proof. (Theorem 9) For convenience, we define the objective function in (4.3) as

$$L(\lambda, \nu) = \max_\theta \left[f(\theta) - \lambda g(\theta) - \nu^T h(\theta) \right],$$

where λ is now reduced to a scalar in the case of $p = 1$. Recall that

$$f(\theta) = \max_{k=1,2,\ldots,K} f^{(k)}(\theta), \qquad g(\theta) = \min_{l=1,2,\ldots,L} g^{(l)}(\theta).$$

Because $\lambda \geq 0$, we have

$$L(\lambda, \nu) = \max_\theta \max_{k,l} \left\{ f^{(k)}(\theta) - \lambda g^{(l)}(\theta) - \nu^T h(\theta) \right\}$$

and, by Lemma 10,

$$L(\lambda, \nu) = \max_{\{p_{kl}, \theta_{kl}\}_{k,l}} \sum_{k,l} p_{kl} \left[f^{(k)}(\theta_{kl}) - \lambda g^{(l)}(\theta_{kl}) - \nu^T h(\theta_{kl}) \right],$$

where $\{p_{kl}\}$ ($k = 1, 2, \ldots, K$; $l = 1, 2, \ldots, L$) need to satisfy $\sum_{k,l} p_{kl} = 1$ and $p_{kl} \geq 0$ for all k and l. Similar to the previous proof in Section 3, we introduce new variables $\gamma_{kl} = p_{kl}\theta_{kl}$, and rewrite $L(\lambda, \nu)$ as

$$L(\lambda, \nu) = \max_{\{p_{kl}, \gamma_{kl}\}_{k,l}} \sum_{k,l} \left[p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) - \lambda p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) - \nu^T p_{kl} h(\gamma_{kl}/p_{kl}) \right].$$

Next, because $f^{(k)}$ is concave and $g^{(l)}$ is convex for all k and l, and h is affine, if problem (4.3) is feasible, then the optimal solution is a saddle point of

$$\sum_{k,l} \left[p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) - \lambda p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) - \nu^T p_{kl} h(\gamma_{kl}/p_{kl}) \right].$$

Therefore, problem (4.3) achieves the same optimal value as the following problem, obtained by exchanging the order of maximization and minimization:

$$\begin{array}{ll} \underset{\{p_{kl},\gamma_{kl}\}}{\text{maximize}} & \min_{\lambda \geq 0, \nu} \sum_{k,l} \left[p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) - \lambda p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) - \nu^T p_{kl} h(\gamma_{kl}/p_{kl}) \right] \\ \text{subject to} & \sum_{k,l} p_{kl} = 1, \qquad p_{kl} \geq 0, \quad \forall k, l. \end{array} \tag{4.5}$$

Using the fact

$$\begin{aligned} &\min_{\lambda \geq 0, \nu} \sum_{k,l} \left[p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) - \lambda p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) - \nu^T p_{kl} h(\gamma_{kl}/p_{kl}) \right] \\ &\quad = \begin{cases} \sum_{k,l} p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) & \sum_{k,l} p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) \leq 0, \ \sum_{k,l} p_{kl} h(\gamma_{kl}/p_{kl}) = 0 \\ -\infty & \text{otherwise}, \end{cases} \end{aligned}$$

we can further rewrite problem (4.5) as problem (3.6) in Theorem 9. □

5 Numerical examples

This section demonstrates the use of convex OUQ through two examples. The first example compares bounds obtained by asymmetric and incomplete information using convex OUQ and the algorithm introduced by Bertsimas and Popescu [3]. The second example presents a scenario where convex OUQ is (up to the authors knowledge) the only applicable convex formulation.

Bound on the tail of Gaussian distribution. This example applies convex OUQ and the algorithm introduced by Bertsimas and Popescu [3] in order to compute an upper bound of $\mathbb{P}(\theta \geq a)$ (where a is a given constant) in presence of incomplete and asymmetric information.

To apply convex OUQ we assume that we are given the constraints

$$\mathbb{E}[\theta] = M_1, \qquad \mathbb{E}[\theta^2] \leq M_2, \qquad \mathbb{E}[|\theta|] \leq M_1^+.$$

To apply Bertsimas and Popescu [3] (which has not been designed to incorporate the constraint $\mathbb{E}[|\theta|] \leq M_1^+$) we assume that we are given the constraints

$$\mathbb{E}[\theta^q] = M_q, \quad q = 1, 2, \ldots, Q$$

for some fixed Q. All the constants, including $\{M_q\}_{q=1}^Q$ and M_1^+, are computed from the standard Gaussian distribution $\mathcal{N}(0, 1)$.

Table 1 lists the results obtained from the two algorithms. As a baseline of comparison, it also lists the exact value of $\mathbb{P}(\theta \geq x)$, which can be evaluated from the complementary error function. The algorithm by Bertsimas and Popescu eventually gives a tighter bound by using more moment-based information, since it is capable of incorporating equality constraints. On the other hand, when higher moment ($Q \geq 6$) information is unavailable, convex OUQ gives a better bound by being able to handle more types of constraints such as $\mathbb{E}[|\theta|] \leq M_1^+$.

Method	Upper bound for $\mathbb{P}(\theta \geq a)$
B & P ($Q = 2$)	0.6400
B & P ($Q = 4$)	0.6074
B & P ($Q = 6$)	0.4964
convex OUQ ($Q = 2$)	0.5319
Exact value	0.2266

Table 1: Upper bounds obtained by convex OUQ and the algorithm in Bertsimas and Popescu (B & P) [3]. The exact value is also listed for reference. In all results, the constant $a = 0.75$.

Revenue estimation. Consider a scenario where a merchant would like to estimate the expected revenue from selling (divisible) goods to multiple customers. Each customer offers a different price, and the merchant tries to maximize the revenue by selling to the one who offers the best price. In addition, price from each customer drops as the quantity increases, which makes the revenue a nonlinear function of the quantity. This model can capture the situation in which the customer gradually loses interest in purchasing as the quantity increases. Eventually, a customer stops purchasing beyond a certain maximum quantity, at which point the merchant can no longer increase the revenue by selling to that customer.

For simplicity, this example considers 3 customers so that the revenue f can be defined as

$$f(\theta) = \max_{k=1,2,3} f^{(k)}(\theta),$$

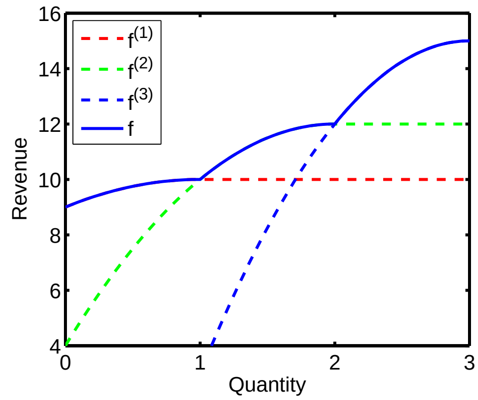

Figure 5.1: Solid: best revenue from all customers. Dashed: revenue from individual customers. The parameters are: $a = (-1, -2, -3)$, $b = (1, 2, 3)$, and $c = (10, 12, 15)$.

where $f^{(k)}$ is the revenue due to customer k. We choose

$$f^{(k)}(\theta) = \begin{cases} a_k(\theta - b_k)^2 + c_k & \theta \le b_k \\ c_k & \theta > b_k, \end{cases}$$

where $a_k < 0$ models the rate at which the price from customer k drops, b_k is the maximum quantity that customer k is willing to purchase, and c_k is maximum revenue from customer k. It can be verified that each $f^{(k)}$ is concave, and thus f is piecewise concave. The functions f and each $f^{(k)}$ are plotted in Figure 5.1 based on the parameters used in this example.

The information constraints include constraints on the first and second moments, i.e.,

$$\mathbb{E}[\theta] = \mu, \qquad \mathbb{E}[\theta^2] \le \mu^2 + \sigma^2,$$

and tail probabilities

$$\mathbb{P}(\theta \le \theta_L) \le \delta_L, \qquad \mathbb{P}(\theta \ge \theta_H) \le \delta_H.$$

As we mentioned in Example 7, both $I(\theta \le \theta_L)$ and $I(\theta \ge \theta_H)$ are piecewise convex, and each can be constructed from 2 convex functions. Therefore, we have $K = 3$, $L_1 = 2$, and $L_2 = 2$. Consequently, the total number of Dirac masses is $K \cdot L_1 \cdot L_2 = 12$.

The corresponding convex optimization problem that solves for an upper

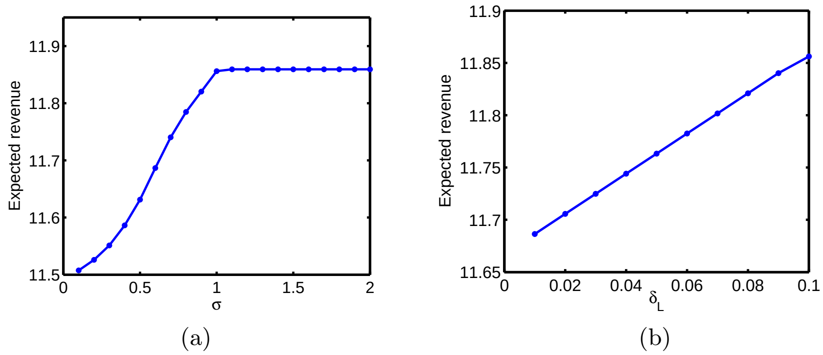

Figure 5.2: Upper bound of the expected revenue computed by convex OUQ. (a) Dependence on the standard deviation σ; (b) Dependence on the (one-side) tail probability δ_L.

bound of the expected revenue $\mathbb{E}[f(\theta)]$ is

$$\begin{aligned} \underset{\{p_{k,l_1,l_2},\gamma_{k,l_1,l_2}\}_{k,l_1,l_2}}{\text{maximize}} \quad & \sum_{k,l_1,l_2} p_{k,l_1,l_2} f^{(k)}(\gamma_{k,l_1,l_2}/p_{k,l_1,l_2}) \\ \text{subject to} \quad & \sum_{k,l_1,l_2} p_{k,l_1,l_2} = 1, \qquad p_{k,l_1,l_2} \geq 0, \quad \forall k, l_1, l_2 \\ & \sum_{k,l_1,l_2} \gamma_{k,l_1,l_2} = \mu \\ & \sum_{k,l_1,l_2} \gamma_{k,l_1,l_2}^2 / p_{k,l_1,l_2} \leq \mu^2 + \sigma^2 \\ & p_{k,1,l_2} \leq \delta_L, \quad \gamma_{k,2,l_2} \geq p_{k,2,l_2}\theta_L, \quad \forall k, l_2 \\ & p_{k,l_1,1} \leq \delta_H, \quad \gamma_{k,l_1,2} \leq p_{k,l_1,2}\theta_H, \quad \forall k, l_1. \end{aligned}$$

Figure 5.2 shows the effect of changing the second moment and the tail probabilities. As expected, loosening the constraints (i.e., increasing either σ or δ_L) increases the upper bound of $\mathbb{E}[f(\theta)]$. In the case of changing the second moment (Figure 5.2a), the upper bound stops increasing beyond a certain point, which implies that the information constraint on the second moment is no longer active.

6 Conclusions

Although optimal uncertainty quantification (OUQ) problems can in general be reduced to finite dimensional optimization problems, the reduced problems may be highly constrained and non-convex. This paper introduces the following new sufficient conditions under which an OUQ problem can be reformulated as a convex optimization problem: (1) the objective is piecewise concave, (2) the inequality information constraints are piecewise convex, and (3) the equality constraints are affine. Constraints on the support of the probability distribution can also be incorporated, as long as the support is a finite union of convex sets. This proof of this result is based on two different approaches starting with the primal and the dual forms of the original OUQ problem.

Acknowledgement. The work is supported jointly by the National Science Foundation (NSF) under grant CNS-0931746 and the Air Force Office of Scientific Research (AFOSR) under grant number FA9550-12-1-0389.

References

[1] N. I. Akhiezer and M. G. Kreĭn. *Some Questions in the Theory of Moments*, volume 2 of *Translations of Mathematical Monographs*. American Mathematical Society, 1962.

[2] D. Bertsimas, X. V. Doan, K. Natarajan, and C.-P. Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.

[3] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804, 2005.

[4] D. Bertsimas and J. N. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific, Belmont, Massachusetts, 1997.

[5] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[6] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[7] K. Isii. On a method for generalizations of Tchebycheff's inequality. *Annals of the Institute of Statistical Mathematics*, 10(2):65–88, 1959.

[8] K. Isii. On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14(1):185–197, 1962.

[9] K. Isii. Inequalities of the types of Chebyshev and Cramér–Rao and mathematical programming. *Annals of the Institute of Statistical Mathematics*, 16(1):277–293, 1964.

[10] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics*, volume 376. Interscience Publishers New York, 1966.

[11] M. G. Kreĭn. The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals and their further development. In E. B. Dynkin, editor, *Eleven papers on Analysis, Probability, and Topology*, American Mathematical Society Translations, Series 2, Volume 12, pages 1–122. American Mathematical Society, New York, 1959.

[12] J. B. Lasserre. Bounds on measures satisfying moment conditions. *The Annals of Applied Probability*, 12(3):1114–1137, 2002.

[13] A. W. Marshall and I. Olkin. Multivariate Chebyshev inequalities. *The Annals of Mathematical Statistics*, pages 1001–1014, 1960.

[14] Y. Nesterov and A. S. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*, volume 13. SIAM, 1994.

[15] H. Owhadi and C. Scovel. Brittleness of Bayesian inference and new Selberg formulas. 2013. Preprint at arXiv:1304.7046.

[16] H. Owhadi, C. Scovel, T. Sullivan, M. McKerns, and M. Ortiz. Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345, 2013.

[17] I. Popescu. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657, 2005.

[18] W. Rogosinski. Moments of non-negative mass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 245(1240):1–27, 1958.

[19] A. Shapiro. On duality theory of conic linear problems. *Nonconvex Optimizaiton and Its Applications*, 57:135–155, 2001.

[20] J. E. Smith. Generalized Chebychev inequalities: Theory and applications in decision analysis. *Operations Research*, 43(5):807–825, 1995.

[21] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52, 2007.