

CONVEX OPTIMAL UNCERTAINTY QUANTIFICATION*

SHUO HAN[†], MOLEI TAO[‡], UFUK TOPCU[†], HOUMAN OWHADI[§], AND
RICHARD M. MURRAY[§]

Abstract. Optimal uncertainty quantification (OUQ) is a framework for numerical extreme-case analysis of stochastic systems with imperfect knowledge of the underlying probability distribution. This paper presents sufficient conditions under which an OUQ problem can be reformulated as a finite-dimensional convex optimization problem, for which efficient numerical solutions can be obtained. The sufficient conditions include that the objective function is piecewise concave and the constraints are piecewise convex. In particular, we show that piecewise concave objective functions may appear in applications where the objective is defined by the optimal value of a parameterized linear program.

Key words. convex optimization, uncertainty quantification, duality theory

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

DOI. 10.1137/13094712X

1. Introduction. In many applications, given a cost function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ that depends on a random variable θ in \mathbb{R}^d , 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 partial information (e.g., moments) about \mathcal{D} that can be expressed in the following form:

$$(1.1) \quad \mathbb{E}_{\theta \sim \mathcal{D}}[g(\theta)] \preceq 0, \quad \mathbb{E}_{\theta \sim \mathcal{D}}[h(\theta)] = 0,$$

where g and h are (real) vector-valued functions. Since the constraints appearing in (1.1) are used to specify the available information about \mathcal{D} , we will refer to these constraints as *information constraints* throughout the paper. Using 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 $\mathbb{E}_{\theta \sim \mathcal{D}}[f(\theta)]$.

Generally speaking, 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 such bounds numerically by solving infinite-dimensional optimization problems over the set of probability distributions that satisfy the information constraints. We follow Owhadi et al. [20] and refer to this problem as *optimal uncertainty quantification* (OUQ) for convenience. (Note that the actual OUQ framework presented in [20] is more general and is capable of dealing with unknown functions f in addition to unknown probability distributions.) Formally, an

*Received by the editors December 2, 2013; accepted for publication (in revised form) April 21, 2015; published electronically July 14, 2015. This work was supported in part by NSF grant CNS-0931746 and AFOSR grant FA9550-12-1-0389.

<http://www.siam.org/journals/siopt/25-3/94712.html>

[†]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 (hanshuo@seas.upenn.edu, utopcu@seas.upenn.edu). This work was performed while the first author was with the California Institute of Technology.

[‡]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 (mtao@gatech.edu).

[§]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125 (owhadi@caltech.edu, murray@cds.caltech.edu).

example of OUQ problems is an optimization problem in the following form:

$$(1.2) \quad \underset{\mathcal{D}}{\text{maximize}} \quad \mathbb{E}_{\theta \sim \mathcal{D}} [f(\theta)]$$

$$(1.3) \quad \text{subject to} \quad \mathbb{E}_{\theta \sim \mathcal{D}} [g(\theta)] \preceq 0,$$

$$(1.4) \quad \mathbb{E}_{\theta \sim \mathcal{D}} [h(\theta)] = 0,$$

$$(1.5) \quad \theta \in \Theta \quad \text{almost surely,}$$

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. The functions g and h are (real) vector-valued. The inequality (1.3) is entrywise. Formulas (1.3) and (1.4) correspond to information constraints 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. The set $\Theta \subseteq \mathbb{R}^d$ is the support of the distribution. For brevity, the phrase “almost surely” is dropped later in this paper. Note that the condition that \mathcal{D} is a probability distribution imposes the following implicit constraints:

$$(1.6) \quad \mathbb{E}_{\theta \sim \mathcal{D}} [1] = 1, \quad \mathcal{D} \geq 0.$$

Main result. Despite being infinite-dimensional, a large class of OUQ problems can be reduced to equivalent finite-dimensional optimization problems that have the same optimal value [20]. This reduction operates in several steps, in which the first one is a generalization of linear programming in spaces of measures [22, 23]. Although this reduction permits a numerical solution to OUQ problems, the resulting reduced problems may be highly constrained and nonconvex.

The paper focuses on the first reduction step and presents sufficient conditions (Theorem 1.1) under which an OUQ problem can be reduced to a finite-dimensional *convex* optimization problem. Specifically, we require that the functions f , g , and h satisfy the following conditions:

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

$$(1.7) \quad f(\theta) = \max_{k=1,2,\dots,K} f^{(k)}(\theta),$$

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

2. Each entry of the function $g: \mathbb{R}^d \rightarrow \mathbb{R}^p$ is *piecewise convex*, i.e., each g_i ($i = 1, 2, \dots, p$) can be written as

$$(1.8) \quad g_i(\theta) = \min_{l_i=1,2,\dots,L_i} g_i^{(l_i)}(\theta),$$

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

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

$$(1.9) \quad h(\theta) = A^T \theta + b$$

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

Figure 1 illustrates how piecewise concave and piecewise convex functions look like in dimension one. In general, these functions are neither concave nor convex. Later in section 4, we will give a number of useful examples of piecewise concave/convex functions in the context of OUQ applications.

The main result of the paper is that the OUQ problem (1.2) can be solved by considering an equivalent finite-dimensional convex optimization problem, given that

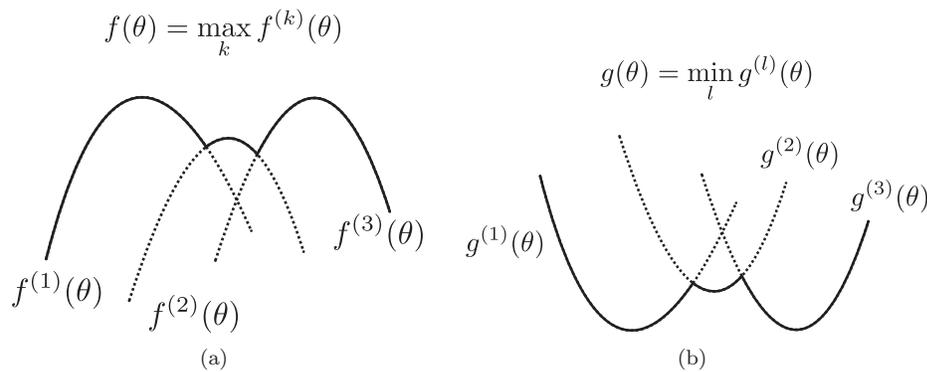


FIG. 1. (a) Piecewise concave and (b) piecewise convex functions in dimension one.

conditions (1.7)–(1.9) are satisfied. For notational simplicity, we will only present the case where g is defined from \mathbb{R}^d to \mathbb{R} (i.e., $p = 1$):

$$(1.10) \quad g(\theta) = \min_{l=1,2,\dots,L} g^{(l)}(\theta), \quad g^{(l)} \text{ is convex.}$$

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

THEOREM 1.1. *The (convex) optimization problem*

$$(1.11) \quad \text{maximize}_{\{p_{kl}, \gamma_{kl}\}_{k,l}} \sum_{k=1}^K \sum_{l=1}^L p_{kl} f^{(k)}(\gamma_{kl}/p_{kl})$$

$$(1.12) \quad \text{subject to} \quad \sum_{k=1}^K \sum_{l=1}^L p_{kl} = 1,$$

$$(1.13) \quad p_{kl} \geq 0, \quad k = 1, 2, \dots, K, \quad l = 1, 2, \dots, L,$$

$$\sum_{k=1}^K \sum_{l=1}^L p_{kl} h(\gamma_{kl}/p_{kl}) = A^T \left(\sum_{k=1}^K \sum_{l=1}^L \gamma_{kl} \right) + b = 0,$$

$$(1.14) \quad \sum_{k=1}^K \sum_{l=1}^L p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) \leq 0$$

achieves the same optimal value as problem (1.2) if the functions f , g , and h satisfy (1.7), (1.10), and (1.9), respectively.

Theorem 1.1 provides sufficient conditions under which an OUQ problem can be solved in a tractable manner, as guaranteed by the convexity of the optimization problem (1.11). The proof of Theorem 1.1 will be given later in section 3. The paper is organized as follows. Section 2 gives a historical perspective of OUQ by reviewing related work and previous results on equivalent finite-dimensional reduction of OUQ problems. Section 3 presents the main result on convex reformulation of OUQ, which can be derived either from the primal or the dual form of the original OUQ problem. Section 4 gives several applications of the convex OUQ formulation. Section 5 provides numerical illustrations of the main theoretical result.

Remarks on the OUQ formulation. Problem (1.2) can also be written in a form without the equality constraint (1.4) and the support constraint (1.5). The equality

constraint (1.4) can be eliminated by introducing the following inequality constraints:

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

The support constraint can be shown as equivalent to

$$(1.15) \quad \mathbb{E}_{\theta \sim \mathcal{D}}[I(\theta \notin \Theta)] \leq 0,$$

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 nonnegative. Therefore, the condition (1.15) is equivalent to

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

which is equivalent to (1.5). However, we will still use the original form in problem (1.2) to distinguish these constraints from pure inequalities. In some cases, problem (1.2) is also called the *(generalized) moment problem* or *problem of moments*, since the information constraints (1.3) and (1.4) often consist of moments of the distribution or can be considered as generalized moments of the distribution.

2. Optimal uncertainty quantification.

2.1. Historical perspective and related work. Among various convex formulations of OUQ, one important special case is when

$$(2.1) \quad f = I(\theta \in C),$$

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 Kreĭn [14] for an account of the history of this subject along with his substantial contributions.) We also refer to early work by Isii [10, 11, 12] and Marshall and Olkin [16].

As related in Owhadi and Scovel [19, 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 [24]. 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 [14, p. 4] (according to Kreĭn [14], 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 obtained an inequality (known as Markov's inequality) that is optimal with respect to the information contained in the first moment.

Example 2.1 (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 [9, p. 311] gives a bound for $\mathbb{P}(\theta \geq a)$ as follows, regardless of the probability distribution of θ :

$$\mathbb{P}(\theta \geq a) \leq \mathbb{E}[\theta]/a.$$

Substituting $\mathbb{E}[\theta] = \mu$ into the above inequality, we obtain

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

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

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

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. Recent works on convex optimization approach to Chebyshev inequalities (motivated by efficient numerical methods such as the interior-point method [18]) also include Lasserre [15], Popescu [21], and Vandenberghe, Boyd, and Comanor [26] for convex formulations for different classes of sets C in (2.1) (e.g., ellipsoids, semialgebraic sets). Compared to these formulations that only consider the case where f can be expressed in the form of (2.1), our formulation extends the objective function f to the more general class of piecewise concave functions as given in (1.7).

Besides indicator set functions (2.1), another class of functions f that appear in convex formulations are functions that are both convex and piecewise affine:

$$(2.2) \quad f(\theta) = \max_{k=1,2,\dots,K} \{a_k^T \theta + b_k\},$$

where K and $\{a_k, b_k\}_{k=1}^K$ are given constants. This form arises in applications such as stock investment [8] and logistics [2]. Since all affine functions are concave, we can define $f^{(k)} = a_k^T \theta + b_k$ in (1.7), so that objective functions in the form of (2.2) become a special case of our formulation.

Besides the conditions on f , there are convex formulations that incorporate information constraints in the form of moment constraints. Oftentimes, these constraints are limited to the mean and covariance, such as in Delage and Ye [8]. For constraints on moments of arbitrary order, 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, p. 170]. The linear matrix inequality formulation allows OUQ problems to be cast as semidefinite programs if the objective function is the 0-1 indicator function [21] or a linear combination of the moments [5, p. 170]. The same

technique can also be extended to the case of complex moments and functions [6]. Compared to constraints on moments of arbitrary order, our formulation is restricted to inequality constraints on the moments of even order but can handle the broader class of piecewise concave objective functions.

To summarize, this paper shows that a convex reformulation of OUQ problems can be obtained for the broader class of piecewise concave objective functions f , which contains both the 0-1 indicator function (2.1) and functions of the form (2.2) that are considered in previous formulations by others. The information constraints in the formulation are required to be inequalities containing piecewise convex functions, which extend the moment constraints in previous formulations at the expense of relaxing equality to inequality constraints.

2.2. Finite reduction of OUQ problems. 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 proposition is originally due to Rogosinski [22]. See also Shapiro [23] for an extension to conic linear programs and Owhadi et al. [20] for a more general result that allows other kinds of constraints.

PROPOSITION 2.1 (finite reduction property; cf. [20, 22, 23]). *Let m be the total number of scalar (in)equalities described by g and h in (1.3) and (1.4). The finite-dimensional problem*

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

achieves the same optimal value as problem (1.2).

Proposition 2.1 implies that the optimal value of any OUQ problem can always be achieved by a discrete probability distribution with finite support, whose number of Dirac masses depends on the information constraints. In the following, we give a simple example (reproduced from Owhadi et al. [20]) to illustrate this property.

Example 2.2 (“seesaw”; see also [20]). The following example OUQ problem illustrates the fact that the optimal value of an OUQ problem can be achieved by a discrete distribution with finite support; it also shows the connection between the number of information constraints and the number of Dirac masses required in the (optimal) discrete distribution. Consider the following OUQ problem for a scalar random variable θ :

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

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

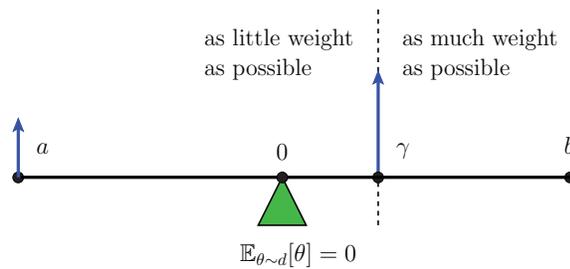


FIG. 2. Seesaw analogy in Example 2.2 (reproduced from [20]).

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 and b (Figure 2). 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 Proposition 2.1 is $1 + 1 = 2$.

3. Convex optimal uncertainty quantification. We will now prove Theorem 1.1, the main result of this paper. Namely, under conditions (1.7)–(1.9), the OUQ problem (1.2) can be reformulated as a finite-dimensional convex optimization problem. We will prove Theorem 1.1 in two ways by considering the primal form and dual form of the OUQ problem, respectively.

3.1. Primal form. 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 3.1. *Suppose that f , g , and h satisfy the conditions (1.7), (1.10), and (1.9), respectively. Then the optimal distribution for problem (1.2) can be achieved by a discrete distribution consisting of at most $K \cdot L$ Dirac masses located at $\{\theta_{kl}\}$ ($k = 1, 2, \dots, K$; $l = 1, 2, \dots, L$). In addition, each θ_{kl} achieves maximum at $f^{(k)}$ and minimum at $g^{(l)}$:*

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

Proof. We will prove the lemma by contradiction. From Proposition 2.1, we know that the optimal probability distribution can be achieved by a discrete distribution with finite support. By way of contradiction, assume that the optimal discrete distribution contains more than $K \cdot L$ Dirac masses. Then, by an argument using the pigeonhole principle, the optimal discrete distribution must contain two Dirac masses located at ϕ_1 and ϕ_2 with probabilities q_1 and q_2 , respectively, such that both ϕ_1 and ϕ_2 achieve maximum at $f^{(k)}$ and minimum at $g^{(l)}$ for some $k \in \{1, 2, \dots, K\}$ and $l \in \{1, 2, \dots, L\}$, i.e.,

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

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

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

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

$$(3.1) \quad qf(\phi) \geq qf^{(k)}(\phi) \geq q_1f^{(k)}(\phi_1) + q_2f^{(k)}(\phi_2) = q_1f(\phi_1) + q_2f(\phi_2),$$

where the second inequality is an application of Jensen’s inequality, and the 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

$$\begin{aligned} qh(\phi) &= q(A^T\phi + b) = A^T(q_1 + q_2)\phi + b(q_1 + q_2) \\ &= A^T(q_1\phi_1 + q_2\phi_2) + b(q_1 + q_2) = q_1h(\phi_1) + q_2h(\phi_2). \end{aligned}$$

The feasibility of the inequality constraint on $\mathbb{E}[g(\theta)]$ can be proved by using a similar argument as in (3.1) by observing that $\mathbb{E}[g(\theta)]$ evaluated under the new distribution will be no larger than that under the original distribution, because

$$qg(\phi) \leq qg^{(l)}(\phi) \leq q_1g^{(l)}(\phi_1) + q_2g^{(l)}(\phi_2) = q_1g(\phi_1) + q_2g(\phi_2).$$

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 3.1 depends only on K and L and is independent from that given by the finite reduction property. By using Lemma 3.1, we can obtain an equivalent convex optimization problem for the original problem (1.2), as given in Theorem 1.1 presented in section 1. The proof of Theorem 1.1 is given as follows.

Proof of Theorem 1.1. According to Lemma 3.1, we can optimize over a new set of Dirac masses whose probability weights and locations are $\{p_{kl}, \theta_{kl}\}$ ($k = 1, 2, \dots, K$, $l = 1, 2, \dots, L$). The requirement that the set of Dirac masses forms a valid probability distribution imposes the constraints (1.12) and (1.13). Under the new set of Dirac masses, the objective function can be rewritten as

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

where the second equality uses the fact that θ_{kl} achieves maximum at $f^{(k)}$. Unless otherwise noted, the range of the summation $\sum_{k,l}$ over the indices k and l is given by $k = 1, 2, \dots, K$ and $l = 1, 2, \dots, L$. 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, p. 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 (1.14) is also convex. All the 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 (1.2) due to Lemma 3.1. \square

Remark. The perspective transformation $p_{kl} f^{(k)}(\gamma_{kl}/p_{kl})$ of $f^{(k)}$, which appears in the objective function (1.11), maintains the convexity of $f^{(k)}$. For certain forms of $f^{(k)}$, the perspective transformation can be directly handled by numerical optimization software. For example, when $f^{(k)}(x) = -x^T P x$ (where P is a positive semidefinite matrix) is a quadratic form, the perspective transformation becomes $p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) = -\gamma_{kl}^T P \gamma_{kl}/p_{kl}$, which can be handled directly by the `quad_over_lin()` function included in CVX [7].

Meanwhile, there are a couple straightforward extensions to Theorem 1.1.

Multiple inequality constraints. Lemma 3.1 and Theorem 1.1 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 1.1.

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

$$(3.2) \quad \Theta = \bigcup_{s=1}^S C^{(s)},$$

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 (1.5) is equivalent to

$$(3.3) \quad \mathbb{E}[I(\theta \notin \Theta)] \leq 0,$$

as presented at the beginning of section 2. When Θ satisfies (3.2), we have

$$(3.4) \quad I(\theta \notin \Theta) = \min\{1, I_\infty^0(\theta \in C^{(1)}), \dots, I_\infty^0(\theta \in C^{(S)})\},$$

which is piecewise convex.

If the support constraint $\theta \in \Theta$ (or equivalently (3.3)) is added to an OUQ problem where f and g satisfy (1.7) and (1.10), 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 (3.3) 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 (3.4). Recall that $p_{kl} \geq 0$ for all k and l . Then we have

$$(3.5) \quad p_{kl} = 0, \quad \gamma_{kls}/p_{kls} \in C^{(s)}, \quad s = 1, 2, \dots, S$$

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

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

for certain constants $A^{(s)}$ and $b^{(s)}$ (cf. [4]). Then the support constraint (3.5) becomes affine constraints

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

for all k and l .

3.2. Dual form. The convex reformulation of OUQ that appears in Theorem 1.1 can also be derived from the Lagrange dual problem of (1.2). Derivation from the dual problem is particularly useful in *distributionally robust stochastic programming*, as studied by Delage and Ye [8] and Mehrotra and Zhang [17]. First, the Lagrangian of problem (1.2) can be written as

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

The last two terms are due to the constraints (1.6). From the Lagrangian, the Lagrange dual can be derived as

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

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

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

we can obtain the the dual problem as

$$(3.6) \quad \underset{\lambda, \nu, \mu}{\text{minimize}} \quad \mu$$

$$(3.7) \quad \text{subject to} \quad f(\theta) - \lambda^T g(\theta) - \nu^T h(\theta) - \mu \leq 0 \quad \forall \theta,$$

$$\lambda \succeq 0,$$

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 [10, 11, 12], strong duality holds so that we can solve the dual problem (3.6) to obtain the optimal value of problem (1.2). Analysis on strong duality can also be found in Karlin and Studden [13], Akhiezer and Krein [1], Smith [24], and Shapiro [23].

The inequality constraint (3.7) implies that the optimal solution $(\lambda^*, \nu^*, \mu^*)$ must satisfy

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

which allows us to eliminate the inequality constraint (3.7) and rewrite problem (3.6) as

$$(3.8) \quad \begin{aligned} & \text{minimize} && \max_{\lambda, \nu} [f(\theta) - \lambda^T g(\theta) - \nu^T h(\theta)] \\ & \text{subject to} && \lambda \succeq 0. \end{aligned}$$

As it turns out, Theorem 1.1 can also be proved from the dual form (3.8). Similar to section 3.1, 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 3.2. *Given a set of real-valued functions $\{f^{(k)}\}_{k=1}^K$, the optimal value of the optimization problem*

$$(3.9) \quad \begin{aligned} & \text{maximize} && \sum_{k=1}^K p_k f^{(k)}(\theta_k) \\ & \text{subject to} && \sum_{k=1}^K p_k = 1, \quad p_k \geq 0, \quad k = 1, 2, \dots, K, \end{aligned}$$

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

Proof. Denote the optimal value of problem (3.9) as OPT and

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

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

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

is a feasible solution of problem (3.9), 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 (3.9). Then we have

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

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

We are now ready to prove Theorem 1.1 from the dual problem (3.8).

Proof of Theorem 1.1. For convenience, we define the objective function in (3.8) as

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

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

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

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 3.2,

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

where $\{p_{kl}\}$ ($k = 1, 2, \dots, K$; $l = 1, 2, \dots, L$) 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 3.1, 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 (3.8) 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 (3.8) achieves the same optimal value as the following problem, obtained by exchanging the order of maximization and minimization:

$$(3.10) \quad \begin{aligned} & \underset{\{p_{kl}, \gamma_{kl}\}}{\text{maximize}} && \min_{\lambda \geq 0, \nu} \sum_{k,l} \left[p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) - \lambda p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) - \nu^T p_{kl} h(\gamma_{kl}/p_{kl}) \right] \\ & \text{subject to} && \sum_{k,l} p_{kl} = 1, \quad p_{kl} \geq 0 \quad \forall k, l. \end{aligned}$$

Using the fact

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

we can further rewrite problem (3.10) as problem (1.11) in Theorem 1.1. \square

Remark. The dual form (3.8) is often used to derive a numerically favorable solution for distributionally robust stochastic programming, in which an optimization problem in the following form is being solved:

$$(3.11) \quad \underset{x}{\text{minimize}} \quad \max_{\mathcal{D} \in \Delta_{\mathcal{D}}} \mathbb{E}_{\theta \sim \mathcal{D}} [f(x, \theta)],$$

where the set $\Delta_{\mathcal{D}}$ is defined by the information constraints (1.3)–(1.5). Using the dual form (3.8), it can be seen that the distributionally robust stochastic programming problem (3.11) can be rewritten as

$$(3.12) \quad \begin{aligned} &\underset{x, \lambda, \nu}{\text{minimize}} \quad \max_{\theta} [f(x, \theta) - \lambda^T g(\theta) - \nu^T h(\theta)] \\ &\text{subject to} \quad \lambda \succeq 0. \end{aligned}$$

If f is convex in x , then the objective function in (3.12) is a pointwise maximum of convex functions of x , λ , and ν , and hence problem (3.12) is convex. In some cases (e.g., Delage and Ye [8], Mehrotra and Zhang [17]), the objective function in (3.12) can be rewritten using linear matrix inequalities to permit a more efficient numerical solution.

4. Applications. In this section, we illustrate the applications of the convex OUQ framework through a number of examples. In particular, we show that piecewise concave functions are expressive enough in modeling the objective function for several different applications.

4.1. Piecewise convex functions as information constraints. In the next, we list several examples of piecewise convex functions that can be used as information constraints.

Example 4.1 (even-order moments). Consider the case where the random variable θ is univariate and the function g in the inequality constraint is given by $g(\theta) = \theta^{2q}$ (for some positive integer q). It can be verified that g is convex, which is a special case of piecewise convex functions.

Example 4.2. Consider $g(\theta) = I(\theta \notin C)$ for any convex set $C \in \mathbb{R}^d$. The function g can be used to specify the probability $\mathbb{P}(\theta \notin C)$, since $\mathbb{P}(\theta \notin C) = \mathbb{E}[I(\theta \notin C)]$. The function g can be rewritten as

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

where the function I_b^a is defined as

$$(4.1) \quad I_b^a(E) = \begin{cases} a, & E = \text{true}, \\ b, & E = \text{false} \end{cases}$$

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

4.2. Piecewise concave functions as objectives. We begin with two simple examples of piecewise concave functions that are used as objective functions in OUQ.

Example 4.3 (convex and piecewise affine). When $f^{(k)}$ is affine (hence concave) for each $k \in \{1, 2, \dots, K\}$, the function $f = \max_{k=1,2,\dots,K} \{f^{(k)}\}$ becomes convex and

piecewise affine. This class of functions has been studied by, e.g., Delage and Ye [8] and Bertsimas et al. [2].

Example 4.4. Consider $f(\theta) = I(\theta \in C)$ for any convex set $C \subseteq \mathbb{R}^d$. The function f 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 definition of $I_{-\infty}^1$ follows from (4.1). It can be verified that both 0 and $I_{-\infty}^1(\theta \in C)$ are concave in θ , and hence f is piecewise concave.

Aside from the two simple examples presented above, we present in the following a very important form of objective functions that is also piecewise concave.

Example 4.5 (optimal value of parameterized linear programs). This form of objective functions is defined by the optimal value of a parameterized linear program whose constraints contain nonlinear terms of the random variable θ . The following is the linear program under consideration:

$$(4.2) \quad \begin{aligned} & \underset{x}{\text{minimize}} && c^T x \\ & \text{subject to} && Ax \preceq u(\theta), \quad Hx = v, \end{aligned}$$

where $x, c \in \mathbb{R}^n, v \in \mathbb{R}^q, A \in \mathbb{R}^{m \times n}, H \in \mathbb{R}^{q \times n}$, and each component of the function $u: \mathbb{R}^p \rightarrow \mathbb{R}^m$ is convex in θ .

For any fixed θ , denote the optimal value of problem (4.2) by $f(\theta)$. We assume that there exists a nonempty set $\Theta \subseteq \mathbb{R}^p$ such that problem (4.2) is feasible for all $\theta \in \Theta$, i.e., $f(\theta) < \infty$ for all $\theta \in \Theta$. We will show that f is piecewise concave on Θ . Namely, f can be expressed as

$$f(\theta) = \max_{k \in \{1, 2, \dots, K\}} f^{(k)}(\theta)$$

for some K and concave functions $\{f^{(k)}\}_{k=1}^K$. First, we consider the Lagrange dual problem of problem (4.2), which is given as follows:

$$(4.3) \quad \begin{aligned} & \underset{\lambda, \mu}{\text{maximize}} && -\lambda^T u(\theta) - \mu^T v \\ & \text{subject to} && A^T \lambda + H^T \mu + c = 0, \quad \lambda \succeq 0. \end{aligned}$$

Since problem (4.2) is feasible, strong duality between problems (4.2) and (4.3) holds, so that $f(\theta)$ is also the optimal value of problem (4.3). Notice that problem (4.3) is a linear program for any given θ . Since the optimal value of a linear program can always be achieved at a vertex of the constraint set, we can rewrite f as

$$f(\theta) = \max_{k \in \{1, 2, \dots, K\}} \{-\lambda_k^T u(\theta) - \mu_k^T v\},$$

where $\{(\lambda_k, \mu_k)\}_{k=1}^K$ is the set of vertices of the polytope

$$\{(\lambda, \mu): A^T \lambda + H^T \mu + c = 0, \lambda \succeq 0\}.$$

Since $\lambda_k \succeq 0$ for all $k \in \{1, 2, \dots, K\}$, and each component of u is convex, we know that $-\lambda_k^T u - \mu_k^T v$ is concave, which implies that f is piecewise concave.

In the following, we give two applications in which the objective function is defined through the optimal value of a linear program.

Revenue maximization with stochastic supplies. We start with a simple application that cannot be handled by previous convex formulations of OUQ. Consider a scenario where a merchant would like to estimate the expected revenue from selling (divisible) goods to K potential customers, whereas the quantity of the goods follows an unknown probability distribution. We assume that the merchant can only choose to sell the goods exclusively to one of the K 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, the 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.

We denote by θ the total quantity of goods. We denote by $f^{(k)}(\theta) \in \mathbb{R}$ the corresponding revenue of selling θ quantity of goods to customer k . In the following, we shall assume that $f^{(k)}$ is concave. Then, the maximum revenue of selling θ quantity of goods is given by

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

which is piecewise concave by definition. We can also write f (as a function of θ) as the optimal value of the following optimization problem:

$$(4.4) \quad \begin{array}{ll} \text{minimize} & \mu \\ & \mu \\ \text{subject to} & \mu \geq f^{(k)}(\theta), \quad k = 1, 2, \dots, K. \end{array}$$

It can be verified that problem (4.4) has the same form as problem (4.2), which also implies that f is piecewise concave.

DC optimal power flow with stochastic demands. Consider a power network modeled as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where \mathcal{N} is the set of nodes, and \mathcal{E} is the set of edges. We assume that a subset $\mathcal{S} \subseteq \mathcal{N}$ of the nodes are capable of generating electric power to supply other loads in the network. The power generation at any node $i \in \mathcal{S}$ is denoted by $s_i \in \mathbb{R}$. For convenience, we assume $s_i = 0$ for all $i \in \mathcal{N} \setminus \mathcal{S}$. We assume that the power demand at any node $i \in \mathcal{N}$ is stochastic, and the demand can be modeled by a function $d_i(\theta)$ that depends on some random variable θ . As an example, the power demand that comes from the usage of air conditioning depends on the ambient temperature, which can be modeled as a random variable. Specifically, we can define $\theta \in \mathbb{R}^{|\mathcal{N}|}$ as the vector of ambient temperatures at different nodes (corresponding to different geographical regions), so that d_i depends on the temperature θ_i at node i . We shall assume that the demand function d_i is concave (in fact, also monotonic) in θ_i for all $i \in \mathcal{N}$.

We consider a simplified DC power flow model of the network adopted from Stott, Jardim, and Alsac [25]. Denote by $q_{ij} \in \mathbb{R}$ the power flow from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$. The power flow is determined by the difference in the voltage angles $\alpha_i, \alpha_j \in \mathbb{R}$ of i and j as follows:

$$(4.5) \quad q_{ij} = B_{ij}(\alpha_i - \alpha_j),$$

where $B_{ij} = B_{ji}$ is the susceptance of the transmission line between i and j . The amount of power flow is also limited by the transmission capacity $\bar{q}_{ij} \in \mathbb{R}_+$ of the line

connecting nodes i and j :

$$(4.6) \quad -\bar{q}_{ij} \leq q_{ij} \leq \bar{q}_{ij}.$$

The net supply at node $i \in \mathcal{N}$ is given by

$$(4.7) \quad p_i = \sum_{j:(j,i) \in \mathcal{E}} q_{ji} + s_i.$$

We require that there should be enough net supply to satisfy the demand at any node $i \in \mathcal{N}$, which implies

$$(4.8) \quad d_i(\theta) \leq p_i, \quad i \in \mathcal{N}.$$

The goal of the DC optimal power flow problem is to minimize the total generation cost while respecting the constraints (4.5)–(4.8). We assume that the generation cost c_i at node i can be modeled by a convex and piecewise affine function:

$$c_i(s_i) = \max_{r=1,2,\dots,R_i} \{a_{ir}^T s_i + b_{ir}\},$$

where R_i and $\{a_{ir}, b_{ir}\}_{r=1}^{R_i}$ are given constants. Then, the DC optimal power flow problem can be cast as an optimization problem as follows:

$$\begin{aligned} & \text{minimize} && \sum_{i \in \mathcal{S}} c_i(s_i) && (\text{over } s_i, q_{ij}, \alpha_i, p_i) \\ & \text{subject to} && (4.5)\text{--}(4.8). \end{aligned}$$

We can rewrite the above optimization problem as a linear program by introducing additional slack variables (also denoted by c_i , in an abuse of notation):

$$(4.9) \quad \begin{aligned} & \text{minimize} && \sum_{i \in \mathcal{S}} c_i && (\text{over } c_i, s_i, q_{ij}, \alpha_i, p_i) \\ & \text{subject to} && (4.5)\text{--}(4.8) \\ & && c_i \geq a_{ir}^T s_i + b_{ir}, \quad i \in \mathcal{S}, \quad r = 1, 2, \dots, R_i. \end{aligned}$$

It can be verified that problem (4.9) has the same form as problem (4.2), and hence its optimal value is a piecewise concave function of θ .

5. Numerical examples. This section demonstrates the numerical aspect 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 follows from the application in revenue maximization (with stochastic supplies) given in section 4; in particular, it presents a scenario where convex OUQ is (to the authors’ knowledge) the only applicable convex formulation, since previous formulations do not handle arbitrary piecewise concave objective functions.

Bound on the tail of Gaussian distributions. 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 the presence of incomplete and asymmetric information.

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 $a = 0.75$. The CPU time is measured on a laptop computer equipped with a dual-core 2.5 GHz Intel Core i5 processor and 4 GB of RAM. All the optimization problems are solved in MATLAB (R2012b) using CVX (Version 2.1, Build 1079) with the Mosek solver (Version 7.0.0.106).

Method	Upper bound for $\mathbb{P}(\theta \geq a)$	CPU time (in seconds, 100 runs)
B & P ($Q = 2$)	0.6400	8.29
B & P ($Q = 4$)	0.6074	11.05
B & P ($Q = 6$)	0.4964	14.21
convex OUQ ($Q = 2$)	0.5319	11.50
Exact value	0.2266	—

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

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

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

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

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 $\mathbb{P}(\theta \geq x)$ as computed by numerically evaluating the integral

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\theta^2/2) d\theta.$$

It can be seen that the above integral is related to the complementary error function $\text{erfc}(\cdot)$ as follows:

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\theta^2/2) d\theta = \frac{1}{2} \text{erfc}(x/\sqrt{2}),$$

where the complementary error function $\text{erfc}(\cdot)$ is defined as

$$\text{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-\theta^2) d\theta.$$

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 $\mathbb{E}[|\theta|] \leq M_1^+$. In terms of computational complexity, the algorithm by Bertsimas and Popescu requires solving a semidefinite program, whereas our convex OUQ formulation only requires solving a second-order cone program. It can be seen from Table 1 that the computational time of our convex OUQ formulation is also comparable to the algorithm by Bertsimas and Popescu.

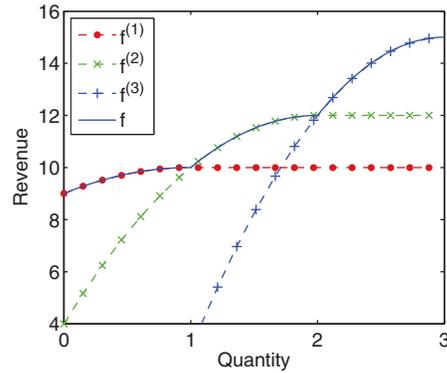


FIG. 3. Solid: best revenue from all customers. Dashed: revenue from individual customers. The parameters are $a = (-1, -2, -3)$, $b = (1, 2, 3)$, and $c = (10, 12, 15)$.

Revenue estimation. This numerical example follows from problem (4.4) presented in section 4. For simplicity, this example considers three customers, so that the revenue f can be defined as

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

where $f^{(k)}$ is the revenue due to customer k , and θ is the (random) quantity of the goods. We choose

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

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 3 based on the parameters used in this example.

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

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

and tail probabilities

$$\mathbb{P}(\theta \leq \theta_L) \leq \delta_L, \quad \mathbb{P}(\theta \geq \theta_H) \leq \delta_H.$$

These information constraints specify the first and second order statistics on the (random) quantity of the goods θ , as well as the probabilities that the quantity θ goes below a given lower bound θ_L or above a given upper bound θ_H , respectively. As we mentioned in Example 4.2, both $I(\theta \leq \theta_L)$ and $I(\theta \geq \theta_H)$ 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 $K \cdot L_1 \cdot L_2 = 12$.

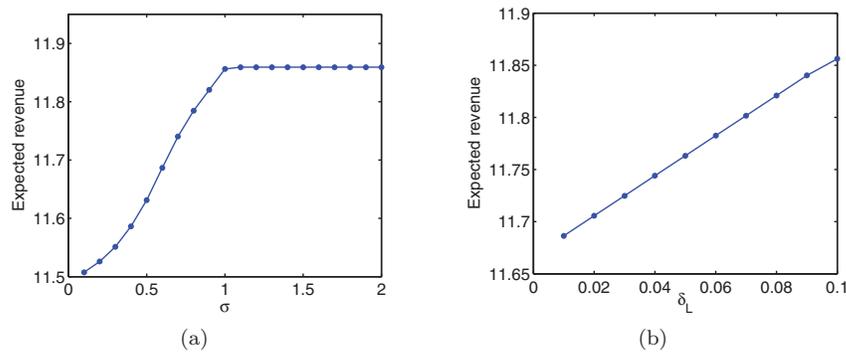


FIG. 4. Upper bound of the expected revenue computed by convex OUQ. (a) Dependence on the standard deviation σ . (b) Dependence on the (one-side) tail probability δ_L .

The corresponding convex optimization problem that solves for an upper bound of the expected revenue $\mathbb{E}[f(\theta)]$ is

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

Figure 4 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 4(a)), 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. 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. We prove the result based on two different approaches, which start with the primal and the dual forms of the original OUQ problem, respectively. Through a number of examples, we also illustrate the use of the convex OUQ formulation in several different applications, such as estimating the maximum revenue of selling goods to customers and the operational cost of power systems.

REFERENCES

- [1] N. I. AKHIEZER AND M. G. KREĬN, *Some Questions in the Theory of Moments*, Transl. Math. Monogr. 2, AMS, Providence, RI, 1962.
- [2] D. BERTSIMAS, X. V. DOAN, K. NATARAJAN, AND C.-P. TEO, *Models for minimax stochastic linear optimization problems with risk aversion*, Math. Oper. Res., 35 (2010), pp. 580–602.
- [3] D. BERTSIMAS AND I. POPESCU, *Optimal inequalities in probability theory: A convex optimization approach*, SIAM J. Optim., 15 (2005), pp. 780–804.
- [4] D. BERTSIMAS AND J. N. TSITSIKLIS, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, 1997.
- [5] S. P. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- [6] C. I. BYRNES AND A. LINDQUIST, *A convex optimization approach to generalized moment problems*, in Control and Modeling of Complex Systems, K. Hashimoto, Y. Oishi, and Y. Yamamoto, eds., Trends Math., Birkhäuser Boston, Cambridge, 2003, pp. 3–21.
- [7] CVX RESEARCH, *CVX: Matlab Software for Disciplined Convex Programming, Version 2.0*. <http://cvxr.com/cvx>, 2012.
- [8] E. DELAGE AND Y. YE, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res., 58 (2010), pp. 595–612.
- [9] G. R. GRIMMETT AND D. R. STIRZAKER, *Probability and Random Processes*, 3rd ed., Oxford University Press, Oxford, 2001.
- [10] K. ISII, *On a method for generalizations of Tchebycheff's inequality*, Ann. Inst. Statist. Math., 10 (1959), pp. 65–88.
- [11] K. ISII, *On sharpness of Tchebycheff-type inequalities*, Ann. Inst. Statist. Math., 14 (1962), pp. 185–197.
- [12] K. ISII, *Inequalities of the types of Chebyshev and Cramér–Rao and mathematical programming*, Ann. Inst. Statist. Math., 16 (1964), pp. 277–293.
- [13] S. KARLIN AND W. J. STUDDEN, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Vol. 376, Interscience, New York, 1966.
- [14] 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 Eleven Papers on Analysis, Probability, and Topology, E. B. Dynkin, ed., Amer. Math. Soc. Transl. Ser. 2 12, AMS, New York, 1959, pp. 1–122.
- [15] J. B. LASSERRE, *Bounds on measures satisfying moment conditions*, Ann. Appl. Probab., 12 (2002), pp. 1114–1137.
- [16] A. W. MARSHALL AND I. OLKIN, *Multivariate Chebyshev inequalities*, Ann. Math. Stat., 31 (1960), pp. 1001–1014.
- [17] S. MEHROTRA AND H. ZHANG, *Models and algorithms for distributionally robust least squares problems*, Math. Program., 146 (2013), pp. 123–141.
- [18] Y. NESTEROV AND A. S. NEMIROVSKII, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.
- [19] H. OWHADI AND C. SCOVEL, *Brittleness of Bayesian inference and new Selberg formulas*, Commun. Math. Sci., to appear.
- [20] H. OWHADI, C. SCOVEL, T. SULLIVAN, M. MCKERNS, AND M. ORTIZ, *Optimal uncertainty quantification*, SIAM Rev., 55 (2013), pp. 271–345.
- [21] I. POPESCU, *A semidefinite programming approach to optimal-moment bounds for convex classes of distributions*, Math. Oper. Res., 30 (2005), pp. 632–657.
- [22] W. W. ROGOSINSKI, *Moments of non-negative mass*, Proc. R. Soc. Lond. Ser. A, 245 (1958), pp. 1–27.
- [23] A. SHAPIRO, *On duality theory of conic linear problems*, Nonconvex Optim. Appl., 57 (2001), pp. 135–155.
- [24] J. E. SMITH, *Generalized Chebyshev inequalities: Theory and applications in decision analysis*, Oper. Res., 43 (1995), pp. 807–825.
- [25] B. STOTT, J. JARDIM, AND O. ALSAÇ, *DC power flow revisited*, IEEE Trans. Power Syst., 24 (2009), pp. 1290–1300.
- [26] L. VANDENBERGHE, S. BOYD, AND K. COMANOR, *Generalized Chebyshev bounds via semidefinite programming*, SIAM Rev., 49 (2007), pp. 52–64.