Convex optimal uncertainty quantification: Algorithms and a case study in energy storage placement for power grids study

Shuo Han, Ufuk Topcu, Molei Tao, Houman Owhadi, Richard M. Murray*

September 23, 2012

Abstract

How does one evaluate the performance of a stochastic system in the absence of a perfect model (i.e. probability distribution)? We address this question under the framework of optimal uncertainty quantification (OUQ), which is an information-based approach for worst-case analysis of stochastic systems. We are able to generalize previous results and show that the OUQ problem can be solved using convex optimization when the function under evaluation can be expressed in a polytopic canonical form (PCF). We also propose iterative methods for scaling the convex formulation to larger systems. As an application, we study the problem of storage placement in power grids with renewable generation. Numerical simulation results for simple artificial examples as well as an example using the IEEE 14-bus test case with real wind generation data are presented to demonstrate the usage of OUQ analysis.

1 Introduction

Suppose we are given an optimal control problem of minimizing the operating cost of a system that depends on some random parameter $\theta \in \mathbb{R}^n$. One prerequisite for this is evaluating the operating cost under a certain control strategy. Conventional stochastic optimal control normally assumes that the probability distribution of θ is given as d, in which case this evaluation amounts to computing the expectation of the cost function $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $\mathbb{E}_{\theta\sim d}[f(\theta)]$ over d. For large-scale systems (i.e., θ is high-dimensional), however, computing this expectation becomes impractical for two reasons. Firstly, there may not be sufficient data for modeling d, unless d belongs to some special class of distributions that can be described by a few parameters (e.g., uniform, Gaussian). For example, for a generic discrete distribution d, the amount of data required for building a tabular representation of d grows exponentially with n and will quickly become intractable. Secondly, even if d could be obtained and represented, exact computation of the expectation would require high-dimensional numerical integration and still remain expensive.

In light of these difficulties, current available solutions have been largely based on two prevalent approaches. These approaches either ignore the stochasticity of θ and fall back to deterministic

*S.H., H.O., and R.M.M. are with the Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125. {hanshuo,owhadi}@caltech.edu, murray@cds.caltech.edu. U.T. is with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104. utopcu@seas.upenn.edu. M.T. is with the Courant Institute of Mathematical Sciences, New York University, New York, NY 10012. mtao@cims.nyu.edu.

analysis, or approximate d using tractable models and compute the expectation using sampling-based methods. Although both have been able to show some useful insight on system design, they are based on strong prior knowledge about the distribution and may not stay consistent with data.

The proposed approach in this paper seeks a *scalable* and *information-based* method that evaluates the expected cost under *imperfect knowledge* about distribution, either due to insufficient data or computational constraints. Unlike the two previous approaches, this approach considers a *set* of distributions rather than commits to a *specific* distribution. This is similar to uncertainty set used in robust control, with the difference that the set lies in the infinite-dimensional space of distributions. On the other hand, it uses information from available data to build a *restricted* set of distributions to avoid over-conservatism. For instance, suppose we are able to estimate from data the mean of d as $\hat{\mu} \in \mathbb{R}^n$, then we can impose a constraint on the mean, i.e.,

$$\mathbb{E}_{\theta \sim d}[\theta] = \hat{\mu} \tag{1}$$

to obtain a smaller set of distributions that still contains the true distribution. In fact, we can use any test function g defined on $\mathbb{R}^n$ to impose constraints in the form[1]

$$\mathbb{E}_{\theta \sim d}[g(\theta)] = \hat{g} \tag{2}$$

for some constant $\hat{g}$. The purpose of g is to include information about the distribution d known from data. Its form normally depends on the type of computation available on the data and/or the mechanism of data collection. A common choice is to specify the moments of d. The constraint (1), for example, corresponds to the case $g(\theta) = \theta$, which specifies the first moment. This kind of constraints does not rely on explicit modeling of d and is often not difficult to generate from historical data consisting of instantiations of θ in the past.

What distribution should one choose within this class of distributions defined by the constraint (2)? One viable choice is the worst-case distribution, which maximizes the operating cost under evaluation. Formally, this can be cast as an optimization problem over d,

$$\underset{d}{\text{maximize}} \quad \mathbb{E}_{\theta \sim d}\left[f(\theta)\right] \qquad \text{subject to} \quad \mathbb{E}_{\theta \sim d}[g(\theta)] = \hat{g}.$$

This choice has the merit of *healthy conservatism*: prepare for the worst case, but only among the cases consistent with the data. The problem is called *optimal uncertainty quantification* (OUQ). Unfortunately, it is an *infinite-dimensional* optimization problem and cannot be solved directly by any optimization packages up-to-date. Due to recent advances, this problem is now known to adopt a *finite reduction*, i.e., there exists a finite-dimensional problem that gives the same optimal value [1]. To some extent, considering the worst-case distribution can be computationally more amenable because it only requires solving an optimization problem (convex in some cases, as will be shown later) as opposed to (high-dimensional) numerical integration.

The first contribution of this paper is to scale OUQ to large systems of which the cost function under evaluation satisfies a particular form. For a generic OUQ problem, the reduced finite-dimensional problem is non-convex and may still be highly computationally demanding to obtain the global optimum. However, it has been shown that there is a special class of OUQ problems solvable using convex optimization [2, 3]. This paper extends this result and studies a more general

[1] More generally, the constraints can be inequalities (cf. [1]).

form called the *polytopic canonical form.* The exact method for solving problem in this form, however, has complexity that grows exponentially with the dimension of the uncertain parameter. In light of this, we propose an approximate iterative method that can be applied to large systems.

The second contribution of this paper is to study the problem of storage placement in power grids with renewable generation under the OUQ framework. Placing storage devices in the grids is considered a promising solution to mitigating the effect of random fluctuations in the renewables [4] and related problems were recently studied in the control community [5, 6]. In this context, it is important to evaluate the ramifications of a given storage placement plan [7]. This paper shows that the evaluation problem can be transformed into an OUQ problem with cost function in the polytopic canonical form. We also present numerical results using a standard power system test case and renewable generation data.

2 Convex optimal uncertain quantification

In this section, we start from an important class of OUQ problems, for which the solution can be obtain using convex optimization. Then we generalize its cost function to what we call the polytopic canonical form. Exact method for solving problems in this form can be prohibitively expensive for a large number of random variables. In order to partially alleviate this difficulty, we propose an iterative approximate method that only requires solving smaller problems at each iteration. The method is guaranteed to converge, and it often converges close to the true optimum for problems we have tested.

2.1 The polytopic canonical form (PCF)

In general, an OUQ problem is non-convex and therefore difficult to obtain its global optimum. However, there is a special case for which the solution can be obtained using convex optimization, provided that two conditions hold: (1) The function f is piecewise linear and convex; (2) the constraints only consist of first and second moments. Under these conditions, it has been shown that the OUQ problem can be solved by convex optimization, according to Theorem 1 due to [2].

Theorem 1 (Delage and Ye)**.** *Let $\mathcal{K}$ be a finite index set. If the function $f : \mathbb{R}^n \to \mathbb{R}$ can be written as*

$$f(\theta) = \max_{k\in\mathcal{K}}\{a_k^T\theta + b_k\} \tag{3}$$

for some $\{a_k\}_{k\in\mathcal{K}} \subset \mathbb{R}^n$ and $\{b_k\}_{k\in\mathcal{K}} \subset \mathbb{R}$, then the optimization problem over d,

$$\begin{aligned} \underset{d}{\text{maximize}} \quad & \mathbb{E}_{\theta\sim d}\left[f(\theta)\right] \\ \text{subject to} \quad & \mathbb{E}_{\theta\sim d}[\theta] = \hat{\mu}, \quad \text{cov}_{\theta\sim d}[\theta] = \widehat{\Sigma}, \end{aligned} \tag{4}$$

achieves the same optimal value as the semidefinite program (SDP) over $Q \in \mathbb{S}^n_+$, $q \in \mathbb{R}^n$, and $r \in \mathbb{R}$,

$$\underset{Q,q,r}{\text{minimize}} \quad \text{tr}((\widehat{\Sigma} + \hat{\mu}\hat{\mu}^T)Q) + \hat{\mu}^T q + r \tag{5}$$

$$\text{subject to} \quad \begin{bmatrix} Q & (q-a_k)/2 \\ (q-a_k)^T/2 & r-b_k \end{bmatrix} \succeq 0, \quad k \in \mathcal{K}. \tag{6}$$

Given the set $\mathcal{K}$, and $\mathcal{C} = \{(a_k, b_k)\}_{k\in\mathcal{K}}$, we denote the optimization problem (4) and (5) as $\mathrm{COUQ}_{\mathcal{D}}(\mathcal{K},\mathcal{C})$ and $\mathrm{COUQ}(\mathcal{K},\mathcal{C})$, respectively. We also denote their (same) optimal value as $\mathrm{COUQ}^*(\mathcal{K},\mathcal{C})$. The dependence on $\hat{\mu}$ and $\widehat{\Sigma}$ is omitted. This notation also applies to any subset $\mathcal{A} \subset \mathcal{K}$, i.e., $\mathrm{COUQ}(\mathcal{A},\mathcal{C})$ denotes the optimization problem

$$\begin{array}{ll} \underset{Q,q,r}{\text{minimize}} & \mathrm{tr}((\widehat{\Sigma} + \hat{\mu}\hat{\mu}^T)Q) + \hat{\mu}^T q + r \\ \text{subject to} & \begin{bmatrix} Q & (q-a_k)/2 \\ (q-a_k)^T/2 & r-b_k \end{bmatrix} \succeq 0, \quad k \in \mathcal{A}. \end{array}$$

In this paper, we will consider a more general form of f, which we call the polytopic canonical form (PCF).

Definition 2 (Polytopic canonical form)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be in the *polytopic canonical form* (PCF) if it can be written as

$$f(\theta) = \max_{(a,b)\in\mathcal{P}} \{a^T\theta + b\}, \quad a \in \mathbb{R}^n, b \in \mathbb{R} \tag{7}$$

for some polytope $\mathcal{P}$ of dimension $(n+1)$.

Alternatively, f can be regarded as the optimal value of a family of linear programs (LP) parameterized by θ:

$$\underset{a,b}{\text{maximize}} \quad a^T\theta + b \qquad \text{subject to} \quad (a,b) \in \mathcal{P}. \tag{8}$$

The PCF (7) subsumes (3). For any f in the form (3) with $\mathcal{C} = \{(a_k, b_k)\}_{k\in\mathcal{K}}$, we can choose $\mathcal{P}$ to be the convex hull of $\mathcal{C}$. This implies $\mathcal{C} \subset \mathcal{P}$, and hence

$$f(\theta) = \max_{(a_k,b_k)\in\mathcal{C}} \{a_k^T\theta + b_k\} \le \max_{(a,b)\in\mathcal{P}} \{a^T\theta + b\}. \tag{9}$$

The last inequality is always tight, which can be shown by using a basic property of LP. Denote the vertices (extreme points) of $\mathcal{P}$ as $\mathcal{V}$. We have $\mathcal{V} \subseteq \mathcal{C}$, hence

$$\max_{(a_k,b_k)\in\mathcal{V}} \{a_k^T\theta + b_k\} \le \max_{(a_k,b_k)\in\mathcal{C}} \{a_k^T\theta + b_k\}. \tag{10}$$

From the optimality of the extreme points, we know that any optimum for the LP (8) can always be attained at some $(a_k, b_k) \in \mathcal{V}$, no matter what θ is chosen, i.e.,

$$\max_{(a,b)\in\mathcal{P}} \{a^T\theta + b\} = \max_{(a_k,b_k)\in\mathcal{V}} \{a_k^T\theta + b_k\}. \tag{11}$$

Therefore, from (9)–(11), the equality

$$\max_{(a,b)\in\mathcal{P}} \{a^T\theta + b\} = \max_{(a_k,b_k)\in\mathcal{C}} \{a_k^T\theta + b_k\}$$

must hold and $f(\theta) = \max_{(a,b)\in\mathcal{P}}\{a^T\theta + b\}$, i.e., any f in the form (3) can be rewritten in PCF. On the other hand, given any function f in PCF, we can also rewrite it in the form (3) by setting $\mathcal{C}$ as the vertices of $\mathcal{P}$. The benefit of using PCF is its flexibility. In PCF, $\mathcal{P}$ can be defined either by its vertices, in which case it reduces to the form (3), or by the intersection of half-spaces. The latter representation can sometimes be more compact, e.g. for the storage placement problem in Section 3.

2.2 Exact iterative method method for PCF

For any f in PCF, there is at least one practical issue in directly applying Theorem 1 by rewriting f in the form (3). Obtaining the vertices $\mathcal{V}$, usually through vertex enumeration algorithms, can be computationally demanding when the dimension of $\mathcal{P}$ is high or the number of its composing constraints is large. In general, the cardinality of $\mathcal{V}$, denoted as $|\mathcal{V}|$, grows exponentially with the dimension n. This becomes prohibitively expensive even for a moderate n and a moderate number of constraints. Even if $\mathcal{V}$ could be obtained, solving the SDP (5) would also be expensive when $|\mathcal{V}|$ (hence $|\mathcal{K}|$) is large.

To this end, we seek iterative methods that solve a smaller problem at each iteration. In general, if we choose an arbitrary subset $\mathcal{A} \subset \mathcal{K}$ and solve the problem COUQ$(\mathcal{A}, \mathcal{V})$, we are only guaranteed to obtain a lower bound COUQ$^*(\mathcal{A}, \mathcal{V}) \leq$ COUQ$^*(\mathcal{K}, \mathcal{V})$ since the constraints for $k \in \mathcal{K} \backslash \mathcal{A}$ have been ignored. The inequality is tight if and only if the optimal solution (Q^*, q^*, r^*) for COUQ$(\mathcal{A}, \mathcal{V})$ also satisfies the constraints for $k \in \mathcal{K} \backslash \mathcal{A}$, i.e.,

$$\begin{bmatrix} Q^* & (q^* - a_k)/2 \\ (q^* - a_k)^T/2 & r^* - b_k \end{bmatrix} \succeq 0, \quad \forall k \in \mathcal{K} \backslash \mathcal{A}. \tag{12}$$

Based on this fact, one can use the following procedure to obtain COUQ$^*(\mathcal{K}, \mathcal{V})$, without including all the constraints in $\mathcal{K}$ in the optimization problem at first:

1. Start with an initial index set $\mathcal{A} \subset \mathcal{K}$.
2. Obtain (Q^*, q^*, r^*) for the problem COUQ$(\mathcal{A}, \mathcal{V})$.
3. If (Q^*, q^*, r^*) satisfies (12), report (Q^*, q^*, r^*) as the solution to COUQ$(\mathcal{K}, \mathcal{V})$ and terminate. Otherwise, there must exist a set $\mathcal{B} \subset \mathcal{K} \backslash \mathcal{A}$ such that the condition (12) is violated for $k \in \mathcal{B}$. Set $\mathcal{A} := \mathcal{A} \cup \mathcal{B}$ and repeat steps 2–3.

2.3 Approximate iterative method for PCF

There are two issues with this procedure. One issue is that checking the condition (12) can be difficult, because the number of constraints to be checked is $|\mathcal{K}| - |\mathcal{A}|$, which is usually large (on the same order as $|\mathcal{K}|$ assuming $|\mathcal{A}|$ is small). The other issue is that, in the worst case, the index set $\mathcal{A}$ may continue to grow until $\mathcal{A} = \mathcal{K}$, in which the final problem to solve has the same complexity as the original problem.

Fortunately, when f can be expressed in PCF, we have a theorem that finds a violating constraint in step 3 without exhaustively checking all the constraints in $\mathcal{K} \backslash \mathcal{A}$. Moreover, Corollary 7 will show that, once such a constraint is found, it can replace an existing constraint in $\mathcal{A}$ without affecting convergence of the method. This prevents $\mathcal{A}$ from growing and avoids the possibility of solving a problem as large as $\mathcal{A} = \mathcal{K}$. This method of finding a violating constraint uses an important property of the solution to the problem COUQ$(\mathcal{A}, \mathcal{V})$. Theorem 3 (cf. [2]) shows that, when we obtain the optimal solution (Q^*, q^*, r^*) to COUQ$(\mathcal{A}, \mathcal{V})$, we automatically obtain the corresponding optimal probability distribution d^* for the problem COUQ$_{\mathcal{D}}(\mathcal{A}, \mathcal{V})$ from the Lagrange multipliers. This optimal distribution d^* can always be realized by a discrete distribution consisting of Dirac masses located at $\{\theta_k\}_{k \in \mathcal{A}} \subset \mathbb{R}^n$ with probability weights $\{p_k\}_{k \in \mathcal{A}} \subset \mathbb{R}_+$.

Theorem 3. *Suppose the Lagrange multiplier for the constraint* (6) *in the problem* $\mathrm{COUQ}(\mathcal{K}, \mathcal{V})$ *is*

$$\begin{bmatrix} \Gamma_k & \gamma_k \\ \gamma_k^T & p_k \end{bmatrix}$$

for each $k \in \mathcal{K}$*. Then for every* $p_k \neq 0$*, the optimal distribution* d^* *for the problem* $\mathrm{COUQ}_{\mathcal{D}}(\mathcal{K}, \mathcal{V})$ *contains a Dirac mass located at* $\theta_k = \gamma_k / p_k$ *with probability* p_k*.*

Remark 4. Theorem 3 implies that the maximum number of Dirac masses in d^* is $|\mathcal{K}|$. On the other hand, the finite reduction theorem (cf. [1]) shows that the number of Dirac masses required for realizing d^* is at most $N + 1$, where N is the number of independent scalar equalities in the constraint (2). In the case of problem (4), for example, the number $N = n + n(n + 1)/2$ (the factor $1/2$ is due to the symmetry of $\widehat{\Sigma}$). Combining the two results, we know the maximum number of required Dirac masses is $\min(|\mathcal{K}|, N + 1)$. In practice, depending on the problem, the actual number of nonzero Dirac masses can be even smaller than $\min(|\mathcal{K}|, N + 1)$.

This gives us another way to compute $\mathrm{COUQ}^*(\mathcal{A}, \mathcal{V})$, i.e.,

$$\mathrm{COUQ}^*(\mathcal{A}, \mathcal{V}) = \sum_{k \in \mathcal{A}} p_k (a_k^T \theta_k + b_k).$$

By using this alternative expression, Theorem 5 shows that $\{\theta_k\}_{k \in \mathcal{A}}$ corresponding to a suboptimal solution (Q^*, q^*, r^*) can be used for finding a violating constraint in $\mathcal{K} \backslash \mathcal{A}$.

Theorem 5. *For a given set* $\mathcal{A}$*, suppose* (Q^*, q^*, r^*) *is the optimal solution for* $\mathrm{COUQ}(\mathcal{A}, \mathcal{V})$ *and the set of Dirac masses of the optimal distribution is* $\{\theta_k\}_{k \in \mathcal{A}}$*. If for any* $u \in \mathcal{A}$*, there exists some* $v \in \mathcal{K}$ *such that*

$$a_v^T \theta_u + b_v > a_u^T \theta_u + b_u, \tag{13}$$

then the constraint

$$\begin{bmatrix} Q^* & (q^* - a_v)/2 \\ (q^* - a_v)^T/2 & r^* - b_v \end{bmatrix} \succeq 0 \tag{14}$$

is violated.

Proof. We prove the theorem by contradiction. Consider the optimization problem $\mathrm{COUQ}(\mathcal{A} \cup \{v\}, \mathcal{V})$. Suppose the condition (14) is not violated, then (Q^*, q^*, r^*) would also be the optimal solution for $\mathrm{COUQ}(\mathcal{A} \cup \{v\}, \mathcal{V})$, which implies that $\mathrm{COUQ}^*(\mathcal{A} \cup \{v\}, \mathcal{V})$ is

$$\sum_{k \in \mathcal{A}} p_k (a_k^T \theta_k + b_k), \tag{15}$$

when $f(\theta) = \max_{k \in \mathcal{A} \cup \{v\}} \{a_k^T \theta + b_k\}$. On the other hand, $\mathrm{COUQ}^*(\mathcal{A} \cup \{v\}, \mathcal{V})$ should be at least

$$p_u (a_v^T \theta_u + b_v) + \sum_{k \in \mathcal{A} \backslash \{u\}} p_k (a_k^T \theta_k + b_k), \tag{16}$$

which is attained under the same discrete distribution consisting of $\{(\theta_k, p_k)\}_{k \in \mathcal{A}}$. The quantity (16) will always be greater than (15), hence a contradiction. □

Remark 6. Condition (13) is only sufficient. Hence, it is not guaranteed to find all the violating constraints.

If f is in PCF, finding such (a_v, b_v) for θ_u only requires solving the LP

$$\underset{a,b}{\text{maximize}} \quad a^T\theta_u + b \qquad \text{subject to} \quad (a, b) \in \mathcal{P}.$$

If the optimal solution (a^*, b^*) for this LP satisfies

$$(a^*)^T\theta_u + b^* > a_u^T\theta_u + b_u,$$

then we have successfully found $(a_v, b_v) = (a^*, b^*)$. Otherwise, no such (a_v, b_v) exists. Another useful by-product of this new way of finding a violating constraint is that the constraint corresponding to u can be removed from $\mathcal{A}$ in the next iteration while still ensuring that $\mathrm{COUQ}^*(\mathcal{A}, \mathcal{V})$ is increasing.

Corollary 7. *For $\mathcal{A}$, $\mathcal{V}$, u and v defined in Theorem 5, let $\mathcal{A}'(u, v) = (\mathcal{A}\backslash\{u\}) \cup \{v\}$. Then*

$$\mathrm{COUQ}^*(\mathcal{A}'(u, v), \mathcal{V}) > \mathrm{COUQ}^*(\mathcal{A}, \mathcal{V}).$$

Proof. For $\{\theta_k\}_{k\in\mathcal{A}}$ in the proof of Theorem 5,

$$\mathrm{COUQ}^*(\mathcal{A}'(u, v), \mathcal{V}) \geq p_u(a_v^T\theta_u + b_v) + \sum_{k\in\mathcal{A}\backslash\{u\}} p_k(a_k^T\theta_k + b_k).$$

The proof of Theorem 5 has shown that the right hand side is strictly greater than

$$\sum_{k\in\mathcal{A}} p_k(a_k^T\theta_k + b_k) = \mathrm{COUQ}^*(\mathcal{A}, \mathcal{V}),$$

which completes the proof. □

Due to Corollary 7, we can use a modified iterative method than the one proposed at the beginning of this section. In particular, Step 3 can be changed to:

3') Obtain $\{\theta_k\}_{k\in\mathcal{A}}$ and check if for any $u \in \mathcal{A}$, there exists $v \in \mathcal{K}$ such that (a_v, b_v) satisfies (13). If not, then report (Q^*, q^*, r^*) as the optimal solution to the problem $\mathrm{COUQ}(\mathcal{K}, \mathcal{V})$ and terminate. Otherwise, for every (u, v) satisfying (13), set $\mathcal{A} := \mathcal{A}'(u, v)$ and repeat steps 2 and 3'.

This approximate method is guaranteed to converge. At each iteration, the new index set $\mathcal{A}$ will give a non-decreasing optimal value for the corresponding optimization problem. Therefore, this sequence of optimal values is monotone and, at the same time, must be bounded by $\mathrm{COUQ}^*(\mathcal{K}, \mathcal{V})$. By the monotone convergence theorem, this sequence, consisting of real numbers, must have a limit, i.e., the method converges. This method is not, in general, guaranteed to converge to the true optimum since there may still be violating constraints when the algorithm exits (see Remark 6). However, the result will always be a lower bound of the true optimal value, since some constraints in $\mathcal{K}$ have been removed from the minimization problem $\mathrm{COUQ}(\mathcal{K}, \mathcal{V})$. Therefore, we can run the same optimization problem multiple times with different initial assignments of $\mathcal{A}$ and choose the highest among all the results to get an improved approximation.

Choosing the size of $\mathcal{A}$ can be potentially important for this method to work properly, since $|\mathcal{A}|$ remains constant over iterations. If its size is too small, $\mathcal{A}$ may not be capable of including all the Dirac masses necessary for realizing the optimal distribution. One possible choice of $|\mathcal{A}|$ is $N + 1$, which is an upper bound of necessary Dirac masses, although this can be conservative for a particular problem (see Remark 4). It remains an open question whether knowing such conservatism *a priori* can help speed up the optimization procedure.

We now use a simple example to test this approximate method on small problems. In these examples, we arbitrarily generate $\hat{\mu} \in \mathbb{R}^n$, $\widehat{\Sigma} \in \mathbb{S}^n_+$, and choose $\mathcal{P}$ as the $(n+1)$-dimensional hypercube

$$\{(a, b) : \mathbf{0} \preceq a \preceq \mathbf{1},\, 0 \leq b \leq 1\},$$

where $\mathbf{1}$ and $\mathbf{0}$ denote vectors in $\mathbb{R}^n$ containing all ones and all zeros, respectively. For each n, we compare the relative error between the exact solution from (5) and the approximate solution. When computing the approximate solution, we choose $|\mathcal{A}| = N + 1$ based on previous discussions. in Fig. 1 shows the results for n from 1 to 16. The choice of n is limited by the computational time of the exact method (for $n = 16$, it takes about 18.6 hours on an Intel Xeon 3.00 GHz workstation). To obtain statistics about the approximate method, we perform 100 trials for each n, and compute the 10% and 90% quantile of the errors. It can been seen that most of the errors are within 5%. Establishing a bound on the rate of growth of the error as n increases is subject to our current work.

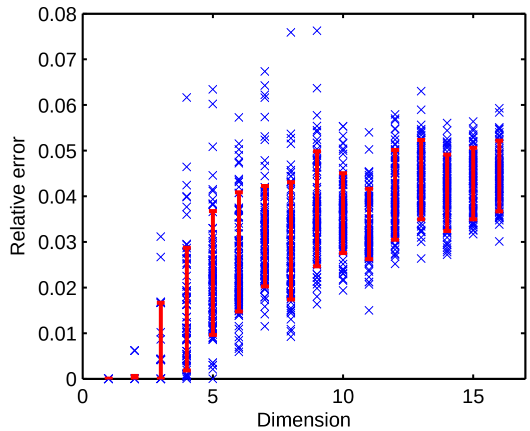


Figure 1: Relative errors of the approximate method. Blue crosses: relative errors. Red bars: 10% and 90% quantiles of the relative errors.

3 Storage placement problem

In this section, we introduce the storage placement problem in power grids as one application of OUQ. We show that this evaluation problem can be converted into PCF using Lagrange duality from optimization theory and solved within the framework of convex OUQ.

3.1 A simple power grid model with energy storage

We model a power grid as a discrete-time dynamical system on a finite graph $(\mathcal{N}, \mathcal{E})$ with time indices $\mathcal{T} = \{1, 2, \cdots, T\}$. The vertices $\mathcal{N}$ are also called *buses*. For simplicity, we use the shorthand notation x to denote the vectorization of $\{x_i(t)\}_{i\in\mathcal{N}, t\in\mathcal{T}}$. At any time t, we refer to $g_i(t)$, $d_i(t)$, and $r_i(t)$ as power generation from renewables, user consumption, and charge rate of storage devices at bus i. As a convention, if the storage devices are being charged, then $r_i(t) > 0$. Under this convention, the total local net power consumption becomes

$$d_i(t) - g_i(t) + r_i(t).$$

Due to physical constraints, the storage level at bus i must stay between 0 and the maximum capacity E_i, i.e.,

$$0 \leq \sum_{\tau=0}^{t} r_i(\tau) \leq E_i, \quad \forall i \in \mathcal{N},\ t \in \mathcal{T} \cup \{0\}.$$

In an abuse of notation, we use $r_i(0)$ to denote the initial level of storage at bus i. Aside from local generation and consumptions, power can also flow between adjacent buses. For any neighboring buses i and j (i.e., $(i,j) \in \mathcal{E}$), the power flow from i to j is given by

$$B_{ij}(\alpha_i(t) - \alpha_j(t)),$$

where B_{ij} is the susceptance of the transmission line between i and j, and $\alpha_i(t)$ is the voltage angle of bus i. Here we use a DC power flow model for simplicity (cf. [8] for its applicability). A transmission line can only support a limited amount power flow $Q_{ij} \geq 0$, which imposes

$$|B_{ij}[\alpha_i(t) - \alpha_j(t)]| \leq Q_{ij}, \quad \forall (i,j) \in \mathcal{E},\ t \in \mathcal{T}.$$

In summary, the total net power consumption at bus i is

$$P_i(t) = \delta_i(t) + r_i(t) + \sum_{(i,j)\in\mathcal{E}} B_{ij}[\alpha_i(t) - \alpha_j(t)],$$

where $\delta_i(t) \triangleq d_i(t) - g_i(t)$. If $P_i(t) \leq 0$, the consumption is covered by all the sources, including local sources and and power flow from adjacent buses. However, if $P_i(t) > 0$, the unmet portion must be matched by additional power sources, usually from the so-called *spinning reserves* in the form of conventional generation.

For simplicity, we assume that the operating cost only depends on the amount of power drawn from spinning reserves. All the other factors, including renewable usage, charging/discharging, and power transmission are assumed to incur no cost. This simplification can potentially be crude. For example, storage devices such as chemical batteries often have a finite number of charging cycles, so charging/discharging cannot be treated as entirely free. These potential refinements will be left for future work. Under this assumption, at time t, the cost for bus i can be modeled as a hinge cost

$$J_i(t) = [P_i(t)]^+ \triangleq \max\{P_i(t), 0\},$$

and the operating cost for the entire grid over time is

$$J = \sum_{i\in\mathcal{N}} \sum_{t=1}^{T} J_i(t).$$

Suppose $\delta_i(t)$ is known, for a given placement of storage $\{E_i\}_{i\in\mathcal{N}}$, one can choose how to operate the storage devices and transmit power over the network to minimize the operating cost by solving the problem

$$
\begin{array}{ll}
\underset{r,\alpha}{\text{minimize}} & J(\delta, r, \alpha) \\
\text{subject to} & |B_{ij}[\alpha_i(t) - \alpha_j(t)]| \le Q_{ij}, \quad \forall (i,j) \in \mathcal{E},\ t \in \mathcal{T}, \\
& 0 \le \sum_{\tau=0}^{t} r_i(\tau) \le E_i, \quad i \in \mathcal{N},\ t \in \mathcal{T} \cup \{0\}, \\
& \sum_{\tau=1}^{T} r_i(\tau) \ge 0, \quad i \in \mathcal{N}.
\end{array}
\tag{17}
$$

The last constraint is added in order to prevent one from minimizing the operating cost by setting a large initial level of charge (which in practice will incur cost). This optimization problem is always feasible, since $r = 0$ and $\alpha = 0$ will satisfy all the constraints.

3.2 Worst-case analysis

We would like to compute the worst-case operating cost under a given placement of storage $\{E_i\}$. We treat $\delta_i(t)$ as the uncertainties for capturing the stochasticity in both renewable generation and user demand. There are two candidate formulations due to the extra freedom in optimizing the power flow by choosing r and α.

- $\max_{\delta\sim d} \mathbb{E}_\delta[\min_{r,\alpha} J(\delta, r, \alpha)]$: This is the "clairvoyant" worst-case analysis. It assumes that power flow optimization will have full knowledge about the actual instantiation of δ.
- $\min_{r,\alpha}[\max_{\delta\sim d} \mathbb{E}_\delta[J(\delta, r, \alpha)]]$: This is the "conservative" worst-case analysis. It assumes a fixed plan for power flow, independent of the actual instantiation of δ.

In this paper, we choose the first formulation because the time horizon under consideration will be 24 hours, and one normally has good knowledge about δ within this horizon (into the future) from forecast, which has been a common practice for many system operators. The second formulation seems too conservative by abandoning any real-time control on the power flow. Formally, the OUQ problem becomes

$$
\begin{array}{ll}
\underset{d}{\text{maximize}} & \mathbb{E}_{\delta\sim d}\left[G(\delta)\right] \\
\text{subject to} & \mathbb{E}_{\delta\sim d}[\delta] = \hat{\mu}, \quad \mathrm{cov}_{\delta\sim d}[\delta] \preceq \widehat{\Sigma},
\end{array}
$$

where $G(\delta)$ is the optimal value of the optimization problem (17) for a given δ.

3.3 Conversion into PCF

Unfortunately, the function G is not in PCF. However, it is possible to convert G into PCF using Lagrange duality. By introducing slack variables, the optimization problem (17) can be rewritten as an LP, i.e.,

$$
\begin{aligned}
\underset{r,\alpha,J_i(t)}{\text{minimize}} \quad & \sum_{i\in\mathcal{N}}\sum_{t=1}^{T} J_i(t) \\
\text{subject to} \quad & B_{ij}[\alpha_i(t)-\alpha_j(t)] \le Q_{ij}, \quad \forall (i,j)\in\mathcal{E},\ t\in\mathcal{T}, \\
& B_{ij}[\alpha_i(t)-\alpha_j(t)] \ge -Q_{ij} \quad \forall (i,j)\in\mathcal{E},\ t\in\mathcal{T}, \\
& 0 \le \sum_{\tau=0}^{t} r_i(\tau) \le E_i, \quad i\in\mathcal{N},\ t\in\mathcal{T}\cup\{0\}, \\
& \sum_{\tau=1}^{T} r_i(\tau) \ge 0, \quad i\in\mathcal{N}, \\
& J_i(t) \ge 0, \\
& J_i(t) \ge \delta_i(t) + r_i(t) + \sum_{j\in N(i)} B_{ij}[\theta_i(t)-\theta_j(t)],
\end{aligned}
$$

whose Lagrange dual problem

$$
\begin{aligned}
\underset{\lambda,\nu}{\text{maximize}} \quad & \sum_{i\in\mathcal{N}}\sum_{t=1}^{T}\lambda_i^{(1)}(t)\delta_i(t) - \sum_{i\in\mathcal{N}}\sum_{t=0}^{T}\lambda_i^{(2)}(t)E_i - \frac{\theta_{\max}}{2}\sum_{(i,j)\in\mathcal{E}}\sum_{t=1}^{T}(\lambda_{ij}^{(3)}(t)+\lambda_{ji}^{(3)}(t)) \\
& \quad - \frac{1}{2}\sum_{(i,j)\in\mathcal{E}}\sum_{t=1}^{T} Q_{ij}(\lambda_{ij}^{(4)}(t)+\lambda_{ji}^{(4)}(t)) && (18) \\
\text{subject to} \quad & 0 \le \lambda_i^{(1)}(t) \le 1, \quad \lambda_i^{(2)}(t) \ge 0, \quad i\in\mathcal{N},\ t\in\mathcal{T} && (19) \\
& \lambda_{ij}^{(3)}(t), \lambda_{ij}^{(4)}(t) \ge 0, \quad (i,j)\in\mathcal{E},\ t\in\mathcal{T}, && (20) \\
& \lambda_i^{(2)}(t) \ge \lambda_i^{(1)}(t+1) - \lambda_i^{(1)}(t) \quad i\in\mathcal{N},\ t\in\mathcal{T}\backslash\{T\}, && (21) \\
& \lambda_i^{(2)}(T) \ge -\lambda_i^{(1)}(T) - \nu, \quad i\in\mathcal{N}, && (22) \\
& \lambda_i^{(2)}(0) \ge \lambda_i^{(1)}(1) + \nu, \quad i\in\mathcal{N}, && (23) \\
& \sum_{(i,j)\in\mathcal{E}}\Big[B_{ij}(\lambda_i^{(1)}(t) - \lambda_j^{(1)}(t) - \lambda_{ij}^{(4)}(t) + \lambda_{ji}^{(4)}(t)) \\
& \quad - \lambda_{ij}^{(3)}(t) + \lambda_{ji}^{(3)}(t)\Big] = 0, \quad i\in\mathcal{N},\ t\in\mathcal{T}. && (24)
\end{aligned}
$$

is also an LP. It can be seen that the dual LP (18) has the form (8) for $a = \lambda^{(1)}$,

$$
\begin{aligned}
b = & -\sum_{i\in\mathcal{N}}\sum_{t=0}^{T}\lambda_i^{(2)}(t)E_i - \frac{\theta_{\max}}{2}\sum_{(i,j)\in\mathcal{E}}\sum_{t=1}^{T}(\lambda_{ij}^{(3)}(t)+\lambda_{ji}^{(3)}(t)) \\
& - \frac{1}{2}\sum_{(i,j)\in\mathcal{E}}\sum_{t=1}^{T} Q_{ij}(\lambda_{ij}^{(4)}(t)+\lambda_{ji}^{(4)}(t)),
\end{aligned}
$$

and the polytope $\mathcal{P}$ defined by constraints (19)-(24). Since the primal problem is an LP and always feasible, strong duality holds, which implies that the dual LP gives the same optimal value as the primal LP. In other words, G can be redefined by the dual LP and hence can be rewritten in PCF.

4 Numerical results and discussions

In this section, we present numerical simulation results for the storage placement problem under three scenarios. For the first two scenarios, we use simple network configurations, in particular, 1-bus and 2-bus networks with synthetic renewable generation data. The purpose of these examples is to show some insight into the differences between deterministic analysis and the OUQ analysis. For the third scenario, we use the IEEE 14-bus test case as a more practical configuration and data from real renewable generation. Through this example, we aim to demonstrate that the approximate method is capable of analyzing a practical system.

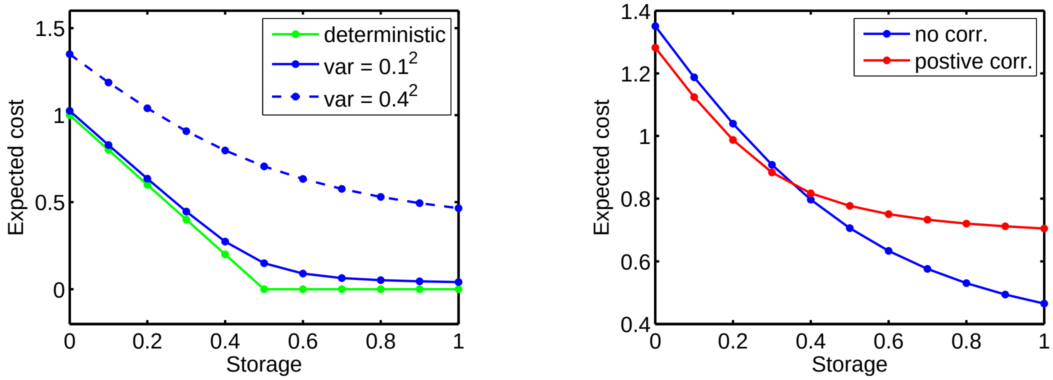
4.1 1-bus network

First we consider a network consisting of one isolated bus, i.e., $|\mathcal{N}| = 1$. This setting has the benefit of isolating any influence by power transmission. We will fix $\hat{\mu}$ and focus on the effect of $\widehat{\Sigma}$. The number of time slices is chosen as 5 so that the exact method can be used. Fig. 2a compares the results from (1) deterministic analysis, which assumes that δ follows $\hat{\mu}$ deterministically, (2) OUQ analysis with $\widehat{\Sigma} = (0.1)^2 I$ (I is the identity matrix), and (3) OUQ analysis with $\widehat{\Sigma} = (0.4)^2 I$. All the curves follow the law of diminishing returns, i.e., adding storage will become less helpful in reducing the operating cost if some storage has already been in place. The differences are in the slope of the curves. For the deterministic analysis, there is a hard threshold after which adding storage will have zero reduction on the cost, whereas the same hard threshold does not appear for the OUQ analysis. This trend is not difficult to understand for the deterministic case: the operating cost cannot be made lower than the cumulative net demand over the entire time horizon, since adding storage does not contribute to power generation. For the results from the OUQ analysis, lower variance will cause a steeper slope. This can be understood by treating the case with lower variance as closer to the deterministic case, which has the steepest slope among all the curves.

The cost-storage curve is not only affected by the variance (diagonal entries of $\widehat{\Sigma}$), but also by the (time) correlation (off-diagonal entries of $\widehat{\Sigma}$). Fig. 2b compares the results of no correlation and positive correlation, where $\widehat{\Sigma}$ is generated from a Laplace covariance function (also known as covariance kernel): $\widehat{\Sigma}_{ij} = \exp(|i - j|/\tau)$ for some constant τ. It can be seen that the presence of positive correlation leads to a slower decrease in the cost. This is expected, since the cost is dominated by the "bad event" during which the net demand at all time instances becomes higher than normal simultaneously, and this is more likely to happen with positive time correlation.

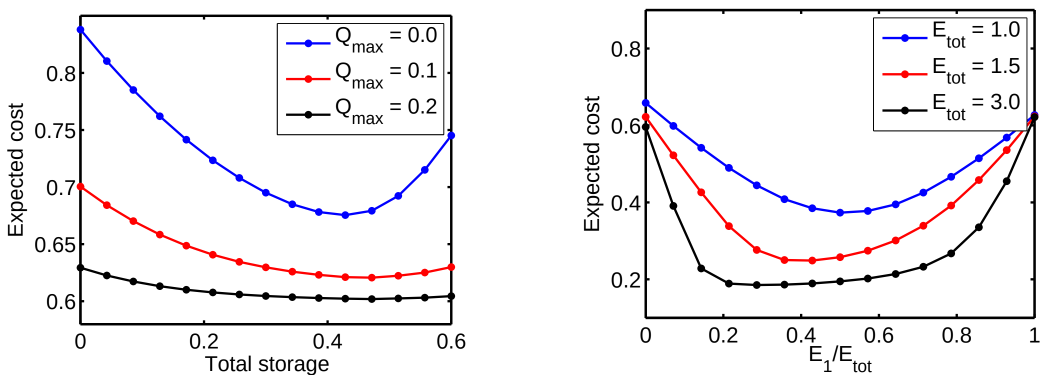
4.2 2-bus network

The purpose of the 2-bus example is to examine the effect of power flow, which can potentially make the operating cost less sensitive to the locations of storage. In the extreme case, if an infinite amount of power is allowed to flow across a fully connected network, then any storage placement will give the same operating cost. For a 2-bus network, there can be only one transmission path, and we study how the maximum power flow $Q_{\max}$ of this path affects the operating cost. The two buses are set to be identical, except for their covariance matrix: $\widehat{\Sigma}_1 = (0.1)^2 I$ and $\widehat{\Sigma}_2 = (0.4)^2 I$.



(a) (b)

Figure 2: Results for 1-bus network. (a) Effect of variance. (b) Effect of (positive) time correlation.



(a) (b)

Figure 3: Results for 2-bus network. (a) Effect of transmission capacity $Q_{\max}$. (b) Effect of total storage E_{tot}.

Fig. 3a compares the results for three power flow limits: $Q_{\max} = 0$ (the two buses are isolated), 0.1, and 0.2. In the simulation, the total storage E_{tot} is fixed, and the operating cost is plotted against E_1, the storage assigned to bus 1. As expected, as $Q_{\max}$ becomes larger, the distribution of storage between the two buses becomes less important.

We also study the effect of total storage E_{tot} on the distribution between the two buses. Fig. 2b shows the operating cost as a function of E_1/E_{tot}, the relative portion of storage for bus 1. As E_{tot} increases, assigning more portion to bus 2 becomes more beneficial. This can be understood from the diminishing return curves in Fig. 2a. Recall that bus 1, whose local demand has a lower variance, enters the diminishing return regime more quickly than bus 2. Therefore, when there has already been enough storage for bus 1, i.e., E_{tot} is large enough, it starts to become more helpful to assign more storage to bus 2, which has not yet entered the diminishing return regime.

4.3 IEEE 14-bus network with renewable generation

In this more practical example, we choose the IEEE 14-bus test case as the network model. The IEEE 14-bus system can be viewed as an abstraction of a portion of the Midwestern US transmission grid. It consists of 5 generator buses and 9 load-only buses. Daily load and generation profiles are created using the data set from [6]. For simplicity, we treat user demand as deterministic and assume that uncertainty only comes from generation, since uncertainty in generation often dominates that in user demand. The time horizon is set to be 24 hours and divided into 8 time slices, which gives 40 random variables in total. This choice of time resolution is limited by the size of the SDP that our machine (Intel Core2 Duo 2.33 GHz, 4 GB RAM, 32-bit) is capable of handling using a general-purpose solver (SeDuMi). We are currently working to overcome this limitation by exploiting the structure of the problem. Statistics, including $\hat{\mu}$ and the diagonal entries of $\widehat{\Sigma}$, are computed from historical records. Since we do not have enough data to compute the full covariance matrix, part of $\widehat{\Sigma}$ is computed from historical data from the Alberta Electric System Operator (AESO), another source of wind generation data. Specifically, we compute the matrix of correlation coefficients from AESO and scale it accordingly to obtain $\widehat{\Sigma}$ for our example. Fig. 4a visualizes the matrix of correlation coefficients. It can be seen that nearby time slots are positively correlated, and the correlation decays as the time difference grows.

Given $\hat{\mu}$, we can solve for the optimal storage placement strategy in the deterministic case. This particular placement is then evaluated using the OUQ analysis. Due to the size of the problem, the approximate method in Section 2.2 is used. Similar to the 1-bus and 2-bus examples, correlation affects the result in the 14-bus example as well. Fig. 4b shows the results for (1) deterministic analysis, (2) OUQ without correlation (diagonal entries of $\widehat{\Sigma}$), and (3) OUQ with correlation. It can be seen that deterministic analysis gives the most optimistic prediction. For the OUQ results, incorporating positive correlation tends to give a more conservative prediction. This can be explained using similar reasoning as that used for Fig. 2b. We also compare the results with the worst-case interval analysis, which considers the worst single deterministic event by ignoring all the moment constraints. For this example, the worst case corresponds to constant zero renewable generation, since generation must stay nonnegative. Both OUQ results are considerably less conservative than the worst-case interval analysis, which gives a constant cost of 48.13 (not shown in Fig. 4b).

5 Related work

The earliest origin of OUQ, or similar problems under different names, can be traced back to the work on generalization of Chebyshev-type inequalities by Isii [9], Marshall and Olkin [10] in the 1950s and 1960s. Interested readers can refer to Owhadi et al [1] for a more recent development on OUQ and finite reduction of the optimization problem. Convex formulation of OUQ has been recently studied by a group of researchers. An incomplete list of these includes Bertsimas and Popescu [3], Vandenberghe et al. [11], Delage and Ye [2], and Topcu et al. [12]. This formulation has recently attracted attention due to the latest advancements in numerical methods for convex optimization problems, in particular, SDPs. The closest work to this paper is the one by Delage and Ye [2], except that they have assumed the problem is always tractable using the exact method.

An extension of OUQ is robust optimization against a partially known probability distribution, sometimes referred to as *distributionally robust stochastic optimization*. This type of robust optimization problem was first proposed in the 1950s by Scarf [13] in the context of inventory optimization. The OUQ problem can be viewed as the inner-loop optimization for this, and is closely

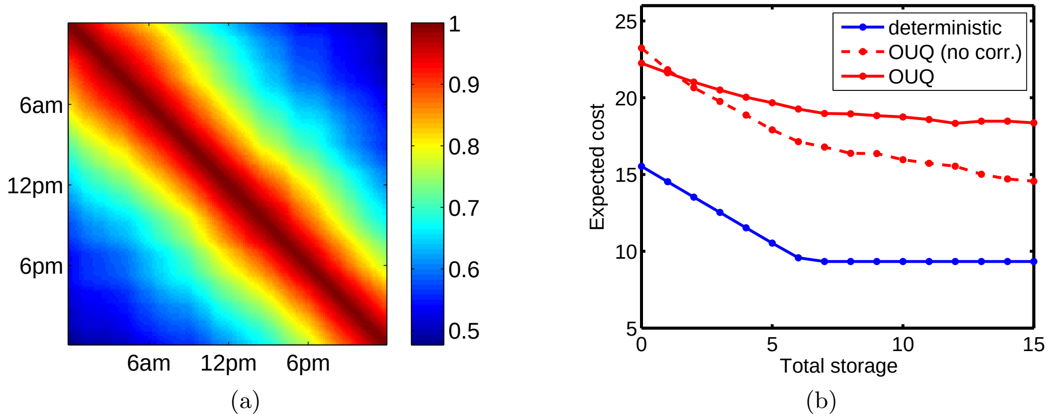


Figure 4: Results for the IEEE 14-bus case with real wind generation. (a) Time correlation in wind power generation. (Source: AESO). (b) Cost evaluation using different models of uncertainty.

related to recent work (e.g., [2, 14]) in this area. Other formulations include uncertainties in the transition probability in Markov decision processes [15, 16]. We consider the robust optimization problem an interesting direction for our future work.

For literature on the distribution of storage in smart grids, many have been based on deterministic analysis [17, 18]. Probabilistic analysis has also been adopted, in particular, the risk-limiting approach [6, 19]. The risk-limiting approach examines the probability that the power grids fall into a certain unsafe operating regime. This criterion has a similar flavor to the worst-case analysis used in this paper. However, current work often assumes a perfect probabilistic model (e.g., multivariate Gaussian) instead of considering a class of distributions consistent with historical data.

6 Conclusions

We have presented the framework of OUQ as a method for evaluating the worst-case performance of stochastic systems without accurate knowledge about the underlying probabilistic model (distribution). OUQ has the advantage of simultaneously considering a class of probability distributions that are consistent with observed data. We generalize previous results and propose the polytopic canonical form for the cost function under which that the OUQ problem can be solved using convex optimization. To scale the formulation to larger systems, we also present iterative methods to alleviate the issue of exponential growth of the constraints. As an application, we study the problem of storage placement in power grids with renewable generation. Numerical simulation results for simple artificial examples as well as an example using the IEEE 14-bus test case with wind generation data are presented. In particular, the OUQ approach is able to incorporate time correlation into the analysis, which has significant influence on the result but can be difficult to include using other approaches.

Acknowledgments. This work is supported by the National Science Foundation (NSF) grant CNS-0931746.

References

[1] H. Owhadi, C. Scovel, T. Sullivan, M. McKerns, and M. Ortiz, "Optimal uncertainty quantification," *SIAM Review*, 2012, (to appear) (preprint).

[2] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[3] D. Bertsimas and I. Popescu, "Optimal inequalities in probability theory: A convex optimization approach," *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 780–804, 2005.

[4] B. Roberts and C. Sandberg, "The role of energy storage in development of smart grids," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1139–1144, 2011.

[5] D. Gayme and U. Topcu, "Optimal power flow with distributed energy storage dynamics," in *American Control Conference*, 2011.

[6] A. E. Sjödin, D. Gayme, and U. Topcu, "Risk-mitigated optimal power flow with high wind penetration," in *American Control Conference*, 2012.

[7] A. Ipakchi and F. Albuyeh, "Grid of the future," *Power and Energy Magazine, IEEE*, vol. 7, no. 2, pp. 52–62, 2009.

[8] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans, "Usefulness of DC power flow for active power flow analysis," in *Power Engineering Society General Meeting*, 2005.

[9] K. Isii, "On a method for generalizations of Tchebycheff's inequality," *Annals of the Institute of Statistical Mathematics*, vol. 10, no. 2, pp. 65–88, 1959.

[10] A. Marshall and I. Olkin, "Multivariate Chebyshev inequalities," *The Annals of Mathematical Statistics*, pp. 1001–1014, 1960.

[11] L. Vandenberghe, S. Boyd, and K. Comanor, "Generalized Chebyshev bounds via semidefinite programming," *SIAM Review*, vol. 49, no. 1, p. 52, 2007.

[12] U. Topcu, G. Calafiore, and L. E. Ghaoui, "Multistage investments with recourse: A single-asset case with transaction costs," in *Proc. Conference on Decision and Control*, 2008.

[13] H. Scarf, K. Arrow, and S. Karlin, "A min-max solution of an inventory problem," *Studies in the mathematical theory of inventory and production*, vol. 10, pp. 201–209, 1958.

[14] A. Shapiro and A. Kleywegt, "Minimax analysis of stochastic problems," *Optimization Methods and Software*, vol. 17, no. 3, pp. 523–542, 2002.

[15] A. Nilim and L. Ghaoui, "Robust control of markov decision processes with uncertain transition matrices," *Operations Research*, pp. 780–798, 2005.

[16] H. Xu and S. Mannor, "Distributionally robust markov decision processes," *Mathematics of Operations Research*, 2012.

[17] M. Korpaas, A. Holen, and R. Hildrum, “Operation and sizing of energy storage for wind power plants in a market system,” *International Journal of Electrical Power & Energy Systems*, vol. 25, no. 8, pp. 599–606, 2003.

[18] K. Heussen, S. Koch, A. Ulbig, and G. Andersson, “Energy storage in power system operation: The power nodes modeling framework,” in *Innovative Smart Grid Technologies Conference Europe*, 2010.

[19] P. Varaiya, F. Wu, and J. Bialek, “Smart operation of smart grid: Risk-limiting dispatch,” *Proceedings of the IEEE*, vol. 99, no. 1, pp. 40–57, 2011.