This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2020.2997259, IEEE Control Systems Letters

Co-Optimization of VaR and CVaR for Data-Driven Stochastic Demand Response Auction

Matt Roveto, *Student Member, IEEE*, Robert Mieth, *Student Member, IEEE*, Yury Dvorkin, *Member, IEEE*

***Abstract*—The ability to make optimal decisions under uncertainty remains important across a variety of disciplines from portfolio management to power engineering. This generally implies applying some safety margins on uncertain parameters that may only be observable through a finite set of historical samples. Nevertheless, the optimized decisions must be resilient to all probable outcomes, while ideally providing some measure of severity of any potential violations in the less probable outcomes. It is known that the conditional value-at-risk (CVaR) can be used to quantify risk in an optimization task, though may also impose overly conservative margins. Therefore, this paper develops a means of co-optimizing the value-at-risk (VaR) level associated with the CVaR to guarantee resilience in probable cases while providing a measure of the average violation in less probable cases. To further combat uncertainty, the CVaR and VaR co-optimization is extended in a distributionally robust manner using the Wasserstein metric to establish an ambiguity set constructed from finite samples, which is guaranteed to contain the true distribution with a certain confidence.**

***Index Terms*—Conditional value-at-risk, Demand Response, Distributionally robust optimization, Optimization, Wasserstein metric**

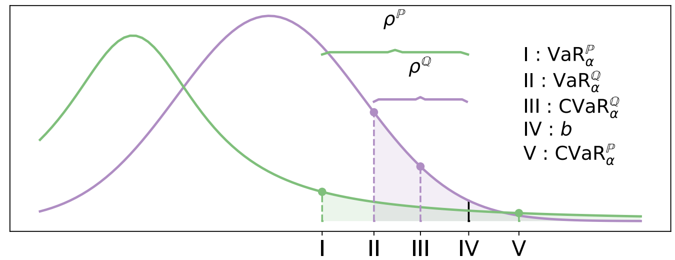

Fig. 1. Probability density of an uncertain loss function for two different probability distributions, in purple ($\mathbb{Q}$) and green ($\mathbb{P}$), and their respective VaR_α and CVaR_α. The shaded area is the probability of events beyond the VaR_α. The acceptability threshold is denoted by b and the distance between b and the VaR is ρ. While the VaR_α of both beliefs lies below b, only the $\mathrm{CVaR}^{\mathbb{Q}}_\alpha$ of the purple distribution is below b. As such, if the true distribution follows the green curve, the value of the uncertain variable may be much larger than expected because the CVaR of the green curve is greater than that of the purple, motivating the use of a distributionally robust approach incorporating the co-optimization of VaR and CVaR.

I. Introduction

CHANCE-CONSTRAINED (CC) [1] and Distributionally Robust Optimization (DRO) [2], [3] both offer a means of optimal decision making under uncertainty that considers risk associated with stochastic parameters and can provide guarantees on constraint satisfaction with a high confidence, while reducing solution conservativeness of deterministic methods with exogenous safety margins. By solving over an ambiguity set, DRO solutions may be less prone to errors arising from inaccurate parameter estimation and incorrect assumptions on the underlying distribution of the uncertain stochastic parameters than CC problems, but DRO requires constructing a proper ambiguity set over which the problem is solved. In turn, ambiguity sets can be formulated either in a moment-based manner, i.e. characterized by all distributions with a given empirical mean and variance [4], and in a metric-based manner, i.e. all distributions within a certain "distance" according to some metric such as the Kullback-Leibler divergence or Wasserstein metric [5]. Metric-based sets, however, may outperform moment-based sets by incorporating more information than provided by statistical moments, [5]. In this paper, we bridge the gap between modeling risk associated with stochastic parameters and metric-based DRO methods.

In most stochastic optimization methods decision-dependent, uncertain (random), outcomes are replaced with a deterministic equivalent that evaluates the underlying stochastic processes using a measure of risk. Most prominently, the expectation operator replaces the uncertain outcome of a decision (e.g. future cost) with the average outcome of infinite repetitions of the decision with respect to some underlying probability distribution. Alternatively, the value-at-risk (VaR) evaluates the uncertain outcome in terms of the probability that it will not exceed a predefined threshold. Leveraging this property, the VaR also allows enforcing probabilistic (chance) constraints of the form $\mathbb{P}(\boldsymbol{X} \le b) \ge \alpha$ by requiring $\mathrm{VaR}_\alpha(\boldsymbol{X}) \le b$, where here and following a **bolded** symbol denotes an uncertain quantity such that the expression $\mathbb{P}(\boldsymbol{X} \le b)$ denotes the probability that random variable $\boldsymbol{X} \in \mathbb{R}$ is less than a scalar value b.

Among large-scale, real-life applications, tractable closed form expressions exist for example in the context of optimal power flow (OPF) formulations, [1], and have become a popular tool in power systems due to the close relation of such chance constraints with existing reliability measures.

While certain distributions may have convex reformulations, in general the VaR has tractability issues. For example, if the underlying probability distribution must be inferred via sampling, attaining a global optimal solution is obstructed by nonconvexity and nondifferentiability, [6], [7]. Alternatively, enabled by the work of Rockafellar *et al.*, [7]–[9], the conditional value-at-risk (CVaR) may be calculated using convex optimization techniques. Notably, optimizing the CVaR implies an optimization of the VaR, i.e. as $\mathrm{VaR}_\alpha(\boldsymbol{X})$ determines an upper bound on $\boldsymbol{X}$ with probability α, $\mathrm{CVaR}_\alpha(\boldsymbol{X})$ is the expected value of $\boldsymbol{X}$ such that $\boldsymbol{X} \ge \mathrm{VaR}_\alpha$ [7], [9], as shown for two different distributions in Fig. 1. This property can be leveraged as a means of providing a convex approximate upper bound on chance constraints, e.g. as shown in Dall'Anese *et al.* [10] for a distributed chance-constrained OPF.

The VaR and CVaR, however, may provide more than a bound on constraints, and offer an additional tool to

2475-1456 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05,2020 at 15:48:31 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2020.2997259, IEEE Control Systems Letters

handling uncertainty in the system. Naively enforcing that $\mathrm{CVaR}_\alpha(\boldsymbol{X}) < b$ may lead to an overly conservative solution since $\mathrm{VaR}_\alpha \leq \mathrm{CVaR}_\alpha$, [7], [9], as shown in Fig. 1. Thus, if one only requires constraints to hold with probability α, a feasible solution may exist in which $\mathrm{VaR}_\alpha(\boldsymbol{X}) = b$ that improves upon the solution in which $\mathrm{VaR}_\alpha(\boldsymbol{X}) < b$, i.e. one in which ρ, the gap between $\mathrm{VaR}_\alpha(\boldsymbol{X})$ and b, is equal to zero (see Fig. 1). This still enforces the desired chance constraints without accruing extra costs of added security. As such, we would like to control the VaR from being lower than b, while minimizing the CVaR, i.e. co-optimizing the VaR and CVaR with the superior mathematical properties of the CVaR, which will prevent excessive conservativeness of solutions, thereby increasing the utility of the decision maker.

This paper synthesizes the most desirable properties of the VaR and CVaR by presenting a model that endogenizes considerations of risk and cost by co-optimizing VaR/CVaR. The solution thus provides insights into how severely constraints may be violated via constrained minimization of the CVaR, while ensuring that the corresponding VaR does not produce an overly conservative solution, and still guarantees that the corresponding chance constraint holds with a desired probability. Relative to [7], we not only constrain the VaR, but do so in a distributionally robust manner that hedges against over-specifying the distribution by means of the Wasserstein metric which is guaranteed to contain the true distribution under certain conditions [11], [12]. Using results of [5], [13] the model scales well, and, due to the convexity of CVaR calculations, preserves the convexity of problems allowing for solutions with off-the-shelf solvers.

We demonstrate the usefulness of this model on a stochastic reverse auction to procure demand response (DR) in a power system. Here a DR service provider (aggregator) tries to select bids from DR participants to provide a target amount of DR capacity at minimal cost. This is a particularly challenging task as the actual response of DR participants is subject to uncertainty and may differ from the initial bid.

II. Data-Driven VaR/CVaR Co-Optimization

A. Preliminaries

Let us first define the α-Value-at-Risk (VaR_α) of a random $\boldsymbol{X}$ as the smallest possible x for which the probability that $\boldsymbol{X}$ is less than or equal to x is at least α, i.e.

$$\mathrm{VaR}_\alpha(\boldsymbol{X}) := \min\{x | \mathbb{P}(\boldsymbol{X} \leq x) \geq \alpha\}, \tag{1}$$

where $\mathbb{P}$ is the probability measure of $\boldsymbol{X}$. Notice then that by (1), when $\mathbb{P}$ is continuous and strictly increasing, the VaR_α is the unique value satisfying $\mathbb{P}(\boldsymbol{X} \leq x) = \alpha$, [9]. We then define the CVaR_α of $\boldsymbol{X}$ as the mean of the α-tail distribution of $\boldsymbol{X}$ as in [9], i.e.

$$\mathrm{CVaR}_\alpha \text{ of } \boldsymbol{X} := \mathbb{E}\big\{\boldsymbol{X} | \boldsymbol{X} > \mathrm{VaR}_\alpha\big\}. \tag{2}$$

Naturally then, we have that $\mathrm{CVaR}_\alpha \geq \mathrm{VaR}_\alpha$, and that $\mathbb{P}(\boldsymbol{X} \leq \mathrm{VaR}_\alpha) \geq \alpha$ from the definition of the VaR_α. Following [7], in order to practically calculate the CVaR_α and VaR_α, we use the following minimization rule:

$$\mathrm{CVaR}_\alpha(\boldsymbol{X}; z) = \min_{z \in \mathbb{R}}\{z + \frac{1}{1-\alpha}\mathbb{E}\big\{[\boldsymbol{X} - z]^+\big\}\} \tag{3}$$

where $[t]^+ := \max\{0, t\}$. We make the minimization over z explicit such that $\mathrm{CVaR}_\alpha(\boldsymbol{X}; z)$ implies a minimization over z. By [9, Theorem 10], the minimization in (3) is convex as a function of z, and further, (3) is equivalent to (2). Intuitively the CVaR_α is a weighted average of all values beyond the VaR_α, which in (3) is equal to the optimizer z^* as per [8, Theorem 1]. Further, (3) can be rewritten as

$$\mathrm{CVaR}_\alpha(\boldsymbol{X}; z) = \min_{z \in \mathbb{R}} \mathbb{E}\big\{\max_{j \leq J} l_j(\boldsymbol{X}, z)\big\}, \tag{4}$$

where $l_j(\boldsymbol{X}, z) := a_j \boldsymbol{X} + b_j z$ for $J = 2$, $a_1 = 0, b_1 = 1, a_2 = 1/(1-\alpha), b_2 = 1 - 1/(1-\alpha)$.

Consider now a more general case in which the random variable is also dependent on some other deterministic decision variable(s) $\vec{y} \in \mathbb{R}^P$ (where henceforth we adopt an overarrow ($\vec{\cdot}$) to denote vectors), i.e. $\boldsymbol{X}(\vec{y}) : \mathbb{R}^P \to \mathbb{R}$. We are no longer interested in just the calculation of CVaR_α, but would like to find such $\vec{y}$ that minimizes our CVaR_α. To do so, we note that function $\mathbb{E}\big\{\max_{j \leq J} l_j(\boldsymbol{X}(\vec{y}), z)\big\}$ is jointly convex over both z and any convex functions of $\boldsymbol{X}(\vec{y})$, such that by [9, Theorem 14], minimizing CVaR_α of $\boldsymbol{X}(\vec{y})$ with respect to $\vec{y} \in Y$ is equivalent to minimizing $\mathbb{E}\big\{\max_{j \leq J} l_j(\boldsymbol{X}(\vec{y}), z)\big\}$ over all $\big(\boldsymbol{X}(\vec{y}), z\big) \in Y \times \mathbb{R}$, i.e.

$$\min_{\vec{y}} \mathrm{CVaR}_\alpha \text{ of } \boldsymbol{X}(\vec{y}) = \min_{\vec{y}, z} \mathbb{E}\big\{\max_{j \leq J} l_j\big(\boldsymbol{X}(\vec{y}), z)\big)\big\}. \tag{5}$$

Provided the expectation over $\boldsymbol{X}(\vec{y})$ is convex, (5) is convex and easily solved by many off-the-shelf solvers.

B. VaR/CVaR Co-Optimization in Stochastic Optimization

Consider a general stochastic optimization problem:

$$\min_{\vec{y}} \quad \mathbb{F}_0\{\boldsymbol{f}_0(\vec{y}, \vec{\boldsymbol{X}}_0(\vec{y}))\} \tag{6a}$$

$$\text{s.t.} \quad \mathbb{F}_i\{\boldsymbol{f}_i(\vec{y}, \vec{\boldsymbol{X}}_i(\vec{y}))\} \leq b_i \quad i = 1, \ldots, M \tag{6b}$$

for some mapping, $\mathbb{F}_i : \mathbb{R} \to \mathbb{R}, i = 0, 1, \ldots, M$, of an uncertain variable $\boldsymbol{f}_i \in \mathbb{R}$ to a real number, which may differ for the evaluation of the objective in (6a) and each constraint in (6b). While (6) is primarily concerned with the uncertain $\boldsymbol{f}_i$ which are in $\mathbb{R}$, these $\boldsymbol{f}_i$ may be functions of other random variables in $\mathbb{R}^N$. For example, in an optimal power flow (OPF) problem, $\boldsymbol{f}_0$ may be the uncertain total cost of generation, which is the sum of the uncertain costs of the individual N generators. Implicit here is the assumption that each individual $\boldsymbol{f}_i$ has arguments of at most $\mathbb{R}^{P+N}$ and maps to $\mathbb{R}$, and we omit $\boldsymbol{f}_i$'s dependence when it is unambiguous. The choice of $\mathbb{F}_i$ will affect the solution to (6), and will depend on the needs of the decision maker. Note here by affixing the subscript i on each uncertain $\vec{\boldsymbol{X}}_i(\vec{y})$ we allow for the possibility that the uncertainty in each term may arise from distinct random vectors not necessarily related to one another. To allow for more general results, $\vec{\boldsymbol{X}}_i(\vec{y})$ denotes the random vector present in the i^{th} expression. In an optimal power flow (OPF) problem, for example, $\mathbb{F}_0$ may be the expectancy operator (i.e. $\mathbb{F}_0 \equiv \mathbb{E}$) such that the objective is to minimize the expected cost, whereas $\mathbb{F}_i$ is commonly some VaR_α or chance constraint that aims to maintain uncertain physical states $\boldsymbol{f}_i$ within their feasible limits with high probability.

Choosing $\mathbb{F}_i$ as a chance constraint or the VaR_α may be difficult to work with in practical solutions as only a relatively

2475-1456 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05,2020 at 15:48:31 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2020.2997259, IEEE Control Systems Letters

few number of distributions offer exact deterministic reformulations, and the formulation for most distributions destroys convexity that may otherwise be available, [7]. Furthermore, merely imposing a bound on the VaR_α only provides a probabilistic guarantee on $\boldsymbol{f}_i$ not exceeding a certain value, with no measure of the expected loss if $\boldsymbol{f}_i$ exceeds that bound. Although choosing $\mathbb{F}_i = \mathrm{CVaR}_\alpha$ provides both a probabilistic bound through the associated VaR_α *as well as* a sense of the risk of potential constraint violations, the CVaR on its own has some limitations.

If the CVaR_α of $\boldsymbol{f}_i(\vec{y}, \vec{\boldsymbol{X}}_i(\vec{y})))$ is $\leq b_i$, then because $\mathrm{CVaR}_\alpha \geq \mathrm{VaR}_\alpha$ this implies that $\mathbb{P}\big(\boldsymbol{f}_i(\vec{y}, \boldsymbol{X}_i(\vec{y})) \leq b_i\big) \geq \alpha$ also holds as per (1). Consequently, if we only want (6b) to hold with probability α (i.e. a chance constraint on (6b)), then replacing this with a constraint such that CVaR_α of $\boldsymbol{f}_i(\vec{y}, \boldsymbol{X}_i(\vec{y})))$ is $\leq b_i$ will actually *over-ensure* the original constraint holds. Hence, there may be a solution for some $\beta \leq \alpha$ at which ensuring the CVaR_β of $\boldsymbol{f}_i(\vec{y}, \boldsymbol{X}_i(\vec{y})))$ is $\leq b_i$ leads to $\mathrm{VaR}_\alpha = b_i$ such that we can relax the CVaR constraint and still ensure that the chance constraint holds, which may reduce the solution conservativeness.

Finding β such that $\beta < \alpha$ is not trivial, since it is not clear how α and β are related, which this paper addresses. Ultimately, we do not want the resulting VaR that comes from our calculation of the CVaR to be too conservative on our chance constraints, because it will affect our objective value. As such, we want to not only minimize this CVaR at some level (i.e. control the CVaR), but since the z defined in (3) in that minimization is the VaR at that level, we also want to keep z *as close as possible* to the critical value b_i while ensuring we will not exceed it (i.e. co-optimize the VaR). In order to accomplish this, while minimizing the CVaR of the constraints in (6b), we also constrain the VaR of each by penalizing deviations from limit b_i without exceeding it. This will control the VaR while simultaneously minimizing over the CVaR, i.e. co-optimizing both quantities.

Thus we can choose measures $\mathbb{F}_0$ and $\mathbb{F}_i$ in (6) to incorporate the CVaR of each term as in the optimization problem of [7, Eq. (5.14)] and modify it to contain an additional constraint on z_i of each chance constraint as follows:

$$\min_{\vec{y}, \vec{\rho} \geq 0} \quad \sum_{i=0}^{M} \mathrm{CVaR}_{\alpha_i}(\boldsymbol{f}_i(\vec{y}, \boldsymbol{X}_i(\vec{y}); z) + \sum_{i=1}^{M} \eta_i \rho_i \tag{7a}$$

$$\text{s.t.} \quad z_i + \rho_i = b_i \qquad \forall i, \tag{7b}$$

where $\vec{\rho} \geq 0$ holds for each component of $\vec{\rho} := \{\rho_1, \rho_2, \ldots, \rho_M\} \in \mathbb{R}^M$, a vector of positive slack variables penalized by parameters η_i to ensure that the resulting VaR of each constraint is *as close as possible* to the critical value b_i without being greater than that. Whereas [7, Eq. (5.14)] minimizes the CVaR_α of an uncertain quantity while simultaneously minimizing the VaR, the addition of (7b) in (7) and the penalty term in (7a) differentiates this work from that of [7, Eq. (5.14)], and allows for constrained minimization of the CVaR by constraining the VaR to be as close to some specified value as possible. Doing so can reduce the conservativeness of the solution, and offers more options for tuning risk preferences. Furthermore, due to the relationship between the VaR and CVaR and the formulation of the problem, we have the following theorem:

Theorem 1. *The optimal ρ_i^* in* (7) *is the exact amount by which the upper limit b_i exceeds the optimal VaR_{α_i}.*

Proof. This comes from direct inspection of (7b) and that $\rho_i \geq 0$, noting that these constitute a reformulation of the inequality $z_i \leq b_i$ with positive slack variable ρ_i, and that z_i is the VaR_{α_i} of each $\boldsymbol{f}_i$ from (4). □

Furthermore, since Theorem 1 allows for exactly calculating the difference between VaR of an uncertain value and the upper bound of that value, and the VaR is by definition the probability that the uncertain value is less than the VaR with probability α, then for $\boldsymbol{f}_i$ with continuous, strictly increasing probability measures, which occurs for any $\boldsymbol{f}_i$ that has no probability masses like the normal or exponentially distributed random variable, we show:

Corollary 1. *For a problem of the form of* (7), *the optimal ρ_i^* can be used to find the exact amount of added security for constraint $\boldsymbol{f}_i$ beyond the α_i tolerance level at the optimum, i.e. the amount by which $\mathbb{P}(\boldsymbol{f}_i^* \leq b_i) > \alpha_i$.*

Proof. In the minimization of CVaR_α, as defined in [7], z_i is the resulting VaR_α, which from (1) implies $\mathbb{P}(\boldsymbol{f}_i \leq z_i) \geq \alpha_i$. For $\boldsymbol{f}_i$ with continuous, strictly increasing probability measures, the VaR_α is the unique value, z_i for which, [9]:

$$\mathbb{P}(\boldsymbol{f}_i \leq z_i) = \alpha_i. \tag{8}$$

Note that $\mathbb{P}(\boldsymbol{f}_i \leq b_i) \equiv \mathbb{P}(\boldsymbol{f}_i \leq z_i + \rho_i) = \mathbb{P}(\boldsymbol{f}_i \leq z_i) + \mathbb{P}(z_i \leq \boldsymbol{f}_i \leq z_i + \rho_i)$. Substituting in (8), we have that

$$\mathbb{P}(\boldsymbol{f}_i \leq b_i) = \alpha_i + \mathbb{P}(z_i \leq \boldsymbol{f}_i \leq z_i + \rho_i), \tag{9}$$

i.e. solving (7) ensures $\boldsymbol{f}_i \leq b_i$ with probability $\alpha_i + \mathbb{P}(z_i \leq \boldsymbol{f}_i \leq z_i + \rho_i)$, and $\mathbb{P}(z_i \leq \boldsymbol{f}_i \leq z_i + \rho_i)$ represents the exact amount of added security in the system. □

We remark that, if the underlying distribution of $\boldsymbol{f}_i$ is known exactly, then one may correspondingly exactly calculate how much added security is in the system. In cases where $\mathbb{P}$ is not known for $\boldsymbol{f}_i$, this quantity provides an estimate of added security, thereby offering a metric that compares the conservativeness of different solutions.

C. Wasserstein Metric in Stochastic Optimization

Calculating (7a) requires solving expectations with respect to a certain probability distribution of each of the uncertain $\boldsymbol{f}_i$, see (4). If the form of the distribution is known, then there may exist exact reformulations of the expectations (e.g for a normally distributed variable), however generally we do *not* know the true distribution and can only infer it from a finite set of historical samples. Consequently, solutions of (7) that assume a specific distribution based on the available observation may exhibit poor out-of-sample performance. With this in mind, we seek to employ a distributionally robust solution to (7) that accounts for all distributions that could have generated the available observations to further immunize against uncertainty. Specifically, for a set of such candidate distributions (i.e.

2475-1456 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05,2020 at 15:48:31 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2020.2997259, IEEE Control Systems Letters

ambiguity set) $\mathcal{P}$, we minimize over the worst case distribution such that

$$\min_{\vec{y},\vec{\rho}\geq 0} \sup_{\mathbb{Q}\in\mathcal{P}} \sum_{i=0}^{M} \mathrm{CVaR}_{\alpha_i}^{\mathbb{Q}}(\boldsymbol{f}_i(\vec{y},\boldsymbol{X}_i(\vec{y}));z_i) + \sum_{i=1}^{M} \eta_i\rho_i \quad (10a)$$

$$\text{s.t.} \quad z_i + \rho_i = b_i \quad \forall i, \quad (10b)$$

where $\mathrm{CVaR}_{\alpha_i}^{\mathbb{Q}}(\boldsymbol{X};z)$ denotes the CVaR_{α_i} of random $\boldsymbol{X}$ with respect to probability measure $\mathbb{Q}$ of $\boldsymbol{X}$. Without proper construction of the ambiguity set, however, there is the possibility that the ambiguity set does not contain the true distribution, and inadequately accounts for uncertainty, or that the ambiguity set is too large, containing distributions that are not representative of the true underlying distribution, and thus yield an overly conservative solution. Ideally we would like an ambiguity set that is just big enough to ensure that the true distribution is contained within, but not too big as to also contain many excess distributions. To accomplish this, recent works have shown that a data-driven paradigm of constructing the ambiguity set using the Wasserstein metric leads to distributionally robust solutions that perform better than single distribution problems. Furthermore, these solutions guarantee that the chance constraints with respect to the underlying true probability distribution can be robustly guaranteed based on a limited number of historical data points [5], [13], [14]. For the set, $\mathcal{P}$, the Wasserstein metric for any two distributions $\mathbb{Q}_1$, $\mathbb{Q}_2 \in \mathcal{P}$ can be defined as, [5]:

$$W(\mathbb{Q}_1,\mathbb{Q}_2) := \inf_{\Pi} \int_{\Omega^2} \|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\| \, \Pi(d\boldsymbol{\omega}_1, d\boldsymbol{\omega}_2), \quad (11)$$

where Π is a joint distribution of $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ with marginal distributions $\mathbb{Q}_1$ and $\mathbb{Q}_2$, support Ω, and $\|\cdot\|$ can be any norm.

For a collection of historical observations of a random variable, denoted by a hat ($\hat{}$), we can construct an optimal ambiguity set with the Wasserstein metric. Given the K-dimensional historical sample set $\{\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_K\}$, the best estimate of the true distribution is the empirical distribution $\hat{\mathbb{P}}_K(t) = \frac{1}{K}\sum_{i=1}^{K} \mathbf{1}_{\{\hat{\omega}_i \leq t\}}$, where $\mathbf{1}_{\{\hat{\omega}_i \leq t\}}$ is the indicator function. Furthermore, [11], [12] have shown that the unknown data-generating distribution belongs to the Wasserstein ambiguity set centered around $\hat{\mathbb{P}}_K$ with confidence $1-\beta$ if the corresponding Wasserstein radius grows sublinearly as a function of $\log(1/\beta)/K$.

Thus for empirical distribution $\hat{\mathbb{P}}_K$ and true distribution $\mathbb{P}$ we have $W(\hat{\mathbb{P}}_K, \mathbb{P}) \leq \varepsilon(K)$ for some sample-dependent monotone function $\varepsilon(\cdot)$ that decreases to zero as its argument tends to infinity. Accordingly, for a historical data set with K samples, the true distribution $\mathbb{P}$ lies within the Wasserstein ball of radius $\epsilon(K)$ centered at empirical distribution $\hat{\mathbb{P}}_K$ such that data-driven ambiguity set $\hat{\mathcal{P}}_K$ is given by

$$\hat{\mathcal{P}}_K = \left\{\mathbb{P} \in \mathcal{P}(\Omega) | W(\mathbb{P}, \hat{\mathbb{P}}_K) \leq \varepsilon(K)\right\} \quad (12)$$

and represents the reliable information about the true distribution $\mathbb{P}$ observed from the K historical samples. To solve (10), we follow [5] and set $\mathcal{P} = \hat{\mathcal{P}}_K$ as given by (12).

With CVaR_α defined in (4) as the expectancy of the pointwise maximum of loss function $l(\boldsymbol{f}(\vec{y},\boldsymbol{X}_i(\vec{y})),z)$ we extend [5, Theorem 6.1 and 4.2] in the context of our problem as:

Theorem 2. *Let each $\boldsymbol{f}_i$ in* (10a) *be the inner product $\langle \vec{y}, \vec{\boldsymbol{X}}_i(\vec{y})\rangle$ of $\vec{y} \in \mathbb{R}^N$ and $\vec{\boldsymbol{X}}_i(\vec{y}) \in \mathbb{R}^N$, i.e. $\boldsymbol{f}_i : \mathbb{R}^{2N} \to \mathbb{R}$ and $\boldsymbol{f}_i$ is convex in both $\vec{y}$ and $\vec{\boldsymbol{X}}_i(\vec{y})$. Accordingly, the loss functions l_j are then affine functions of the form $l_j(\boldsymbol{f}_i(\vec{y},\vec{\boldsymbol{X}}_i),z_i) = a_j\langle\vec{y},\vec{\boldsymbol{X}}_i\rangle + b_j z_i$, and $\max_{j\leq J}(a_j\langle\vec{y},\vec{\boldsymbol{X}}_i\rangle + b_j z_i)$ is convex over $\vec{\boldsymbol{X}}_i(\vec{y})$, where $a_1 = 0, b_1 = 1, a_2 = 1/(1-\alpha_i), b_2 = 1 - 1/(1-\alpha_i)$ as in Section II-A for the CVaR_α. Further, let the support Ω_i of random vectors $\vec{\boldsymbol{X}}_i(\vec{y})$ be polytopes, i.e. $\Omega_i = \{\vec{\boldsymbol{X}}_i(\vec{y}) \in \mathbb{R}^N : C_i\vec{\boldsymbol{X}}_i(\vec{y}) \leq \vec{d}_i\}$, where C_i is a matrix and $\vec{d}_i$ a vector of appropriate dimensions $\forall i$ such that the Ω_i are closed and convex. Let $||\cdot||$ be any norm on $\mathbb{R}^N$, dual norm $||\cdot||_* := \sup_{||\xi||\leq 1}\langle\cdot,\xi\rangle$, and inner product of two vectors, $a,b \in \mathbb{R}^N$ $\langle a,b\rangle := a^T b$. For ambiguity set $\hat{\mathcal{P}}_K$ defined via the Wasserstein metric as in* (12) *with radius $\varepsilon(K) \geq 0$, the problem in* (10) *must be smaller or equal to*

$$\min_{\substack{\vec{y},\vec{z},\vec{\rho}\geq 0,\\ \lambda, s_{ik}, \vec{\gamma}_{ikj}}} \quad \lambda\varepsilon(K) + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=0}^{M} s_{ik} + \sum_{i=1}^{M}\eta_i\rho_i \quad (13a)$$

$$\begin{aligned}
\textit{s.t.} \quad & z_i + \rho_i = b_i && \forall i && (13b)\\
& b_j z_i + \langle\vec{\gamma}_{ikj}, \vec{d}_i - C_i\hat{\vec{\omega}}_i^k\rangle \\
& \qquad + a_j\langle\vec{y},\hat{\vec{\omega}}_i^k\rangle \leq s_{ik} && \forall i,k,j && (13c)\\
& ||C_i^T\vec{\gamma}_{ikj} - a_j\vec{y}||_* \leq \lambda && \forall i,k,j && (13d)\\
& \vec{\gamma_{ikj}} \geq 0 && \forall i,k,j, && (13e)
\end{aligned}$$

where $\hat{\vec{\omega}}_i^k$ is the k^{th} historical observation of the i^{th} random vector $\vec{\boldsymbol{X}}_i(\vec{y})$, and $\lambda, s_{ik}, \vec{\gamma}_{ikj}$ are auxiliary and epigraphical decision variables.

Proof. Consider the CVaR term in objective (10a), where we rewrite it using the formulation in (5). Namely, we have $\min_{\vec{y},\vec{z}}\sup_{\mathbb{Q}\in\hat{\mathcal{P}}_K}\mathbb{E}\{\sum_{i=0}^{M}\max_{j\leq J} l_j(\langle\vec{y},\vec{\boldsymbol{X}}_i(\vec{y})\rangle, z_i)\}$. Since $\langle\vec{y},\vec{\boldsymbol{X}}_i(\vec{y})\rangle$ is convex, then l_j is convex in $\vec{\boldsymbol{X}}_i(\vec{y})$, and this term is equivalent to $\min_{\vec{y},\vec{z},\lambda,s_{ik},\vec{t}_{ikj},\vec{v}_{ikj}} \lambda\varepsilon(K) + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=0}^{M} s_{ik}$ subject to constraints (14c) and (14d) for auxiliary and epigraphical decision variables $\lambda, s_{ik}, \vec{t}_{ikj}, \vec{v}_{ikj}$ by [5, Theorem 6.1 and 4.2].

The second term in (10a), $\sum_{i=1}^{M}\eta_i\rho_i$, is convex by construction. Thus adding this to the convex reformulation above preserves the overall convexity of the problem. Further, constraint (7b) is affine, merely restricting the feasible set of values, and similarly for the convex constraint that $\rho_i \geq 0$, which restricts the feasible set to a half space.

Define the conjugate of a function as $[g]^*(\cdot) := \sup_{\vec{\xi}\in\mathbb{R}^N}\langle\cdot,\vec{\xi}\rangle - g(\cdot)$ and the support function of closed, convex Ω_i as $\sigma_{\Omega_i}(\cdot) := \sup_{\vec{\omega}\in\Omega_i}\langle\cdot,\vec{\omega}\rangle$. Then by using [5, Theorem 6.1 and 4.2][1], (10) is equivalent to

$$\min_{\substack{\vec{y},\vec{z},\vec{\rho}\geq 0,\lambda,\\ s_{ik},\vec{t}_{ikj},\vec{v}_{ikj}}} \quad \lambda\varepsilon(K) + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=0}^{M} s_{ik} + \sum_{i=1}^{M}\eta_i\rho_i \quad (14a)$$

[1]For brevity, we direct the reader to the proofs of [5, Theorem 6.1 and 4.2] for the full details and further insights, and note that (14) extends [5, Theorem 4.2] for multiple random vectors $\vec{\boldsymbol{X}}_i$ by including an additional objective term and constraint. As noted, the inclusion of these does not invalidate the reformulation in [5]. Thus we have that (10) with (12) is equivalent to (14).

2475-1456 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05,2020 at 15:48:31 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2020.2997259, IEEE Control Systems Letters

$$
\begin{aligned}
\text{s.t.}\quad & z_i + \rho_i = b_i && \forall i && \text{(14b)}\\
& [-l_j]^*(\vec{t}_{ikj} - \vec{v}_{ikj}) + \sigma_{\Omega_i}(\vec{v}_{ikj}) && && \\
& \qquad - \langle \vec{t}_{ikj}, \hat{\vec{\omega}}_{ik} \rangle \le s_{ik} && \forall i,k,j && \text{(14c)}\\
& ||\vec{t}_{ikj}||_* \le \lambda && \forall i,k,j. && \text{(14d)}
\end{aligned}
$$

Since the feasible set of (10) and (14) are equivalently restricted, the reformulation from [5] remains valid.

Finally, using [5, Corollary 5.1] and details therein we reformulate constraints (14c) and (14d) into (13c) – (13e), which is valid for the given assumptions such that $\boldsymbol{f}_i = \langle \vec{y}, \vec{\boldsymbol{X}}_i(\vec{y}) \rangle$ and that suppports Ω_i are polytopes. As before, the full details of this part of the proof are beyond the scope of this paper, and we direct the reader to [5, Corollary 5.1] for further insights into the reformulation of (14c) and (14d) into (13c) – (13e). □

Naturally then, $\varepsilon(K)$ is crucial to the quality and conservativeness of the result of (13). As shown in [13], following the derivations in [3], [15], the radius can be expressed as

$$
\varepsilon(K) = C\sqrt{\frac{1}{K}\log\left(\frac{1}{1-\beta}\right)} \qquad (15)
$$

for some confidence level β such that we may control the radius to be optimal size given the available data via the value of C. Further shown in [13], we may safely estimate C as

$$
C \approx 2\inf_{\xi>0}\left(\frac{1}{2\xi}\left(1+\ln\left(\frac{1}{K}\sum_{i=1}^{K} e^{\xi\left\|\hat{\vec{\omega}}_i-\hat{\vec{\mu}}\right\|_1^2}\right)\right)\right)^{1/2}, \qquad (16)
$$

where $\hat{\vec{\mu}}$ is the sample mean of the data, and the minimization over ξ can be performed by the bisection search method.

Before proceeding, we highlight that Theorem 2 allows one to co-optimize the CVaR and VaR of an uncertain value in a distributionally robust manner for an interesting class of problems important to a wide range of fields, such as a portfolio optimization or knapsack problem. Along with Theorem 1 and Corollary 1, Theorem 2 offers one both a means of co-optimizing the CVaR and VaR, as well as a metric to compare conservativeness of solutions via the size of ρ, the gap between the VaR of a value and its upper limit.

III. Stochastic Reverse Auction

We turn now to a specific application of the above in which an aggregator operates a reverse auction for demand response (DR) as depicted in Fig. 2. The aggregator looks to receive an amount of DR D_t determined by the system operator, as in [16], and receives a penalty proportional to the amount by which they fail to meet D_t. The aggregator procures DR consumption from their base level before the call for DR r_n as well as a price π_n for each unit of reduced energy.

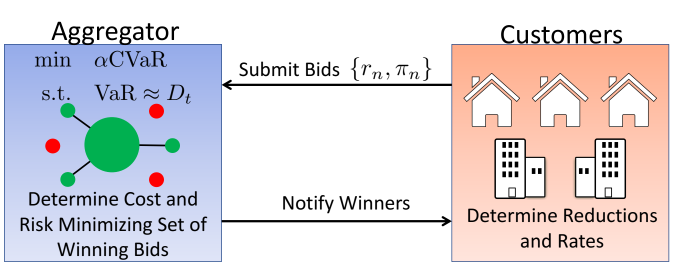

Fig. 2. A schematic representation of the reverse auction.

Furthermore, while customers submit bids of how much they are willing to reduce their consumption, they may not need to reduce their consumption to this full amount, and are paid based on a realized reduction. Because of either inaccuracy in the initial forecast of pre-DR call consumption levels or inability to actually deliver the scheduled amount of load consumption (voluntary or otherwise), the actual amount of delivered demand response of each customer is itself uncertain, rendering this a stochastic optimization problem. Appropriately, we will model the actual reduction of each customer as the submitted bid r_n plus an additional term to represent their unknown deviation from the bid $\boldsymbol{\Delta}_n$ which can be positive (over reduction), or negative (under reduction). We assume that the aggregator will also have access to some historical record of each customer's bidding and *actual* past consumption levels for the k^{th} DR call $\check{x}_n^k$, i.e. the true amount of energy consumed at the time of past DR calls, as well as to the amount each customer has been *scheduled* to consume $\hat{x}_n^k$. Accordingly, the aggregator will be able to calculate a performance measure for each single DR call as $\delta_n^k := \hat{x}_n^k - \check{x}_n^k$ for each customer i, which can be viewed as a *realization* of the random variable $\boldsymbol{\Delta}_n$.

Hence, the loss function of the aggregator is the sum of all realized reduction levels $\boldsymbol{\xi}_n := r_n + \boldsymbol{\Delta}_n$ multiplied by price π_n from all *accepted* bids. For ease of notation we define $\vec{\boldsymbol{\omega}}_0 := \{\boldsymbol{\xi}_1\pi_1, \ldots, \boldsymbol{\xi}_N\pi_N\}$ to be the N-dimensional vector of uncertain customer reductions multiplied by the price at which they wish to be compensated such that $\boldsymbol{\omega}_{0n} = \boldsymbol{\xi}_n\pi_n$, the uncertain reduction of customer n multiplied by customer n's desired compensation, Similarly, we define $\vec{\boldsymbol{\omega}}_1 := \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N\}$ to be the N-dimensional vector of uncertain customer reductions. The aggregator therefore tries to optimally select the bids in order to minimize this loss function. However, the aggregator must also ensure that the sum of all actual reduction levels (i.e. the DR deliverable to the system operator) is greater than their target goal, D_t, i.e. for the vector of binary decisions $\vec{u} := \{u_1, u_2, \ldots, u_N\}$

$$
\begin{aligned}
\min_{\vec{u}} \quad & \langle \vec{u}, \vec{\boldsymbol{\omega}}_0 \rangle && \text{(17a)}\\
\text{s. t.} \quad & -\langle \vec{u}, \vec{\boldsymbol{\omega}}_1 \rangle \le -D_t && \text{(17b)}\\
& u_n \in \{0,1\}. && \text{(17c)}
\end{aligned}
$$

We now recast (17) in the context of Section II, i.e. to ensure (17b) holds with a given probability while minimizing the harmful tail costs (CVaR_α) of (17a) and (17b) in a data-driven distributionally robust manner. We define $\hat{\vec{\omega}}_0^k$ as the the k^{th} observation of the uncertain vector $\vec{\boldsymbol{\omega}}_0$, and similarly for $\hat{\omega}_1^k$. Now, we reformulate the problem from (17) to co-optimize the CVaR and VaR in a distributionally robust manner using the ambiguity set defined via the Wasserstein metric and K historical observations of each customer using our results from Sections II-B and II-C as follows:

$$
\min_{\vec{u},\vec{z},\rho\lambda,s_{ik},\vec{\gamma}_{ikj}} \quad \lambda\varepsilon(K) + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=0}^{1} s_{ik} + \eta\rho \qquad (18a)
$$

2475-1456 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05,2020 at 15:48:31 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2020.2997259, IEEE Control Systems Letters

$$
\begin{aligned}
\text{s.t.} \quad & (13c) – (13e) && (18b)\\
& z_1 + \rho = -D_t && (18c)\\
& u_n \in \{0, 1\}. && (18d)
\end{aligned}
$$

Similar to Section II-C, all α_i are the tolerance to risk, i.e. the probability with which one desires a certain constraint to hold. Additionally all z_i are the VaR, controlled for the constraint given by (17b) via (18c), η is some number characterizing the decision maker's preference of added security in the system, and ρ is a positive slack variable that allows for control. Thus solving (18) allows us to find the least cost set of winners of the DR reverse auction that ensures the target DR is met, while simultaneously controlling the VaR and CVaR in a data-driven distributionally robust manner.

IV. Numerical Results

This case study considers a stochastic reverse auction in Fig. 2 as modeled in (18) with 10 customers using the Pecan Street data [17] from the New York area. Each customer is allowed to bid up to γ of their average energy consumption, for γ ranging from 10% to 30%, at the unitary price $\pi_n = 1$. The penalty factor η is set as a fraction η^π of π and the total amount of DR to be collected (D_t) is set to 50% of the collected bids. To solve (18), we generated 100 random samples in Ω_i with variance of samples p of the bid, with p varying from 10% to 30%. In other words, customers with a smaller bid had a smaller variance, and customers with a larger bid had a larger variance, which allows for studying a trade-off between low-risk, low-return customers (i.e. small bid, small variance) and higher-risk, higher-return customers (i.e. those with a larger bid but also a larger variance). All simulations were conducted using Julia/JuMP and the Gurobi solver [18], [19].

Table I compares the performance of (18) for different penalty values. We use the case with no penalty ($\eta^\pi = 0$), i.e. CVaR is minimized with no co-optimization on VaR, to establish the base case. As η^π increases, so does the relative decrease (in %, relative to the base case with $\eta_\pi = 0$) in ρ for all combinations of γ and σ. In other words, penalty η reduces a gap between the VaR and a desired target goal D_t. While Table I shows that co-optimization enabling constraint (18c) reduces the gap between the VaR and target goal, the size of the gap depends on the variance and a larger variance yields a lower gap reduction.

TABLE I
Relative Differences of ρ for Various η, γ, and σ (in %, relative to the base case with $\eta_\pi = 0$)

	$\gamma = 10\%$			$\gamma = 20\%$			$\gamma = 30\%$		
σ \\ η^π	0.0	0.5	1.0	0.0	0.5	1.0	0.0	0.5	1.0
10%	0.0	99.8	100	0.0	99.4	100	0.0	29.7	100
20%	0.0	99.9	100	0.0	98.2	100	0.0	59.9	100
30%	0.0	86.5	100	0.0	93.7	100	0.0	73.0	100

V. Conclusion

This paper has presented a method to calculate the CVaR of a constraint from a finite number of samples, while simultaneously controlling the VaR level to be as close as possible without exceeding the upper bound of a constraint. This co-optimization method has also been extended to accommodate the underlying uncertainty in a distributionally robust manner. Building on the existing result in [5], we proved that the CVaR and VaR co-optimization over an ambiguous uncertainty set can be implemented in a computationally tractable manner. The proposed CVaR and VaR co-optimization has been applied to a stochastic reverse auction, in which an aggregator seeks to procure uncertain amounts of DR from a pool of customers at the lowest cost, similar to a portfolio optimization problem. Our results demonstrate that the proposed control scheme effectively reduces the gap between the VaR and the upper bound of a constraint. Future efforts will investigate the effects of this co-optimization model on revenue adequacy and cost recovery properties of the reverse auction with distribution power flow constraints.

References

[1] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-Constrained Optimal Power Flow: Risk-Aware Network Control under Uncertainty," *Siam Review*, vol. 56, no. 3, pp. 461–495, 2014.
[2] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009, vol. 28.
[3] C. Zhao and Y. Guan, "Data-driven risk-averse stochastic optimization with Wasserstein metric," *Op. Res. Let.*, vol. 46, no. 2, pp. 262–7, 2018.
[4] 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.
[5] P. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and reformulations," *Math. Prog.*, vol. 171, no. 1-2, pp. 115–166, 2018.
[6] C. Lim, H. Sherali, and S. Uryasev, "Portfolio optimization by minimizing conditional value-at-risk via nondifferentiable optimization," *Comp. Opt. and App.*, vol. 46, no. 3, pp. 391–415, 2010.
[7] R. T. Rockafellar, "Coherent Approaches to Risk in Optimization Under Uncertainty," in *OR Tools and App.: Gl. of Fut. Tech.*, 2007, pp. 38–61.
[8] R. T. Rockafellar *et al.*, "Optimization of Conditional Value-at-Risk," *Journal of Risk*, vol. 2, pp. 21–42, 2000.
[9] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *J. Bank. & Fin.*, vol. 26, no. 7, pp. 1443–1471, 2002.
[10] E. Dall'Anese, K. Baker, and T. Summers, "Chance-Constrained AC Optimal Power Flow for Distribution Systems With Renewables," *IEEE Tran. Pwr. Syst.*, vol. 32, no. 5, pp. 3427–3438, 2017.
[11] F. Bolley *et al.*, "Quantitative concentration inequalities for empirical measures on non-compact spaces," *Prob. Th.*, vol. 137, 2007.
[12] N. Fournier and A. Guillin, "On the rate of convergence in Wasserstein distance of the empirical measure," *Probability Theory and Related Fields*, vol. 162, no. 3-4, 2015.
[13] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally Robust Chance-Constrained Approximate AC-OPF with Wasserstein Metric," *IEEE Tran. Pwr. Syst.*, vol. 33, no. 5, pp. 4924–4936, 2018.
[14] R. Gao and A. J. Kleywegt, "Distributionally Robust Stochastic Optimization with Wasserstein Distance," *arXiv preprint arXiv:1604.02199*, 2016.
[15] C. Zhao and Y. Guan, "Data-Driven Risk-Averse Stochastic Program with ζ-Structure Metrics," *Available on Optimization Online*, 2015.
[16] K. Khezeli and E. Bitar, "Risk-Sensitive Learning and Pricing for Demand Response," *IEEE Tran. Sm. Gr.*, vol. 9, no. 6, 2017.
[17] Pecan Street, "Dataport: the World's Largest Energy Data Resource," *Pecan Street Inc*, 2015. [Online]. Available: "http://www.https://www.pecanstreet.org/dataport/"
[18] I. Dunning, J. Huchette, and M. Lubin, "JuMP: A Modeling Language for Mathematical Optimization," *SIAM Review*, vol. 59, no. 2, 2017.
[19] Gurobi Optimization, LLC, "Gurobi optimizer reference manual," 2020. [Online]. Available: http://www.gurobi.com

2475-1456 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05,2020 at 15:48:31 UTC from IEEE Xplore. Restrictions apply.