Online Convex Optimization for Demand Response

Antoine Lesage-Landry and Joshua A. Taylor
The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
University of Toronto
Toronto, Ontario, Canada, M5S 3G4
{alandry@ece.,josh.taylor@}utoronto.ca.

***Abstract*—Renewable integration requires additional flexibility to balance sudden changes in generation. In this work, we use online convex optimization to track setpoints with uncertain, flexible loads. We define a mean-regularizer to minimize the impact on the comfort of the loads. We extend our framework to the case where the aggregator does not have access to individual load measurements. Finally, we give numerical examples to demonstrate the performance our approach.**

I. Introduction

Renewable sources of energy are becoming a significant part of the power generation infrastructure [1], [2]. There is need to store energy for when natural phenomena are not adequate for generation. On second to minute time-scales, there is an increased need for power balancing and regulation [3], [4]. To add flexibility to the grid, we propose an approach that enables load aggregators to track a power setpoint using Demand Response (DR). The setpoint can be, for example, a regulation signal such as the area control error.

We use Online Convex Optimization (OCO) [5], [6] to adjust the power consumption of flexible loads. We consider both positive and negative adjustments for loads such as thermostatic loads and electric vehicles. We assume that the resulting change in the load's consumption is uncertain. We address this uncertainty using OCO.

The power grid is subject to several sources of uncertainty such as weather, unknown load models, and human behavior [7]. Due to this uncertainty, neither the loads' responses nor the signal to follow are perfectly known at the moment when load commands are given. OCO enables us to simultaneously dispatch loads for DR while learning their uncertain features.

Our OCO-based algorithm minimizes a setpoint tracking objective plus two regularizers. A sparsity regularizer is used to reduce the number of loads mobilized each round. A mean-regularizer reduces the long-term impact on the loads by ensuring that each receives its nominal average power over time. In the context of thermostatically controlled loads, for example, this corresponds to penalizing temperature drift.

To deal with an objective function consisting of a loss function and regularizers, we use the Composite Objective MIrror Descent (`COMID`) algorithm proposed in [8]. We give a limited or bandit feedback extension for our proposed OCO algorithm. Finally, we evaluate our work in numerical simulation, and find that the algorithm achieves effective setpoint tracking.

A. Related work

There have been a number of recent applications of online learning to demand response. Several papers have used multi-armed bandit based approaches [9], [10], [11], [12], [13], and more recently OCO has been applied.

OCO was originally proposed in [14]. Since then, it has found applications in several fields including demand response. Electric Vehicles (EVs) are heavy power consumers and are also a large source of flexibility. Based on these characteristics, Soltani et al. [15] and Ma et al. [16] proposed online convex optimization-based algorithms for load-shifting with EVs via pricing. OCO and real-time pricing were used in [17] to minimize the variance of the power demand and hence flatten the demand. They design a limited feedback extension to their model which they apply to the EV. Lastly, Ledva et al. [18] used an extension to online mirror descent that identifies the controllable portion of the aggregate load in real-time.

B. Contributions

Our approach enables aggregators to track power setpoints under resource uncertainty, for instance, when providing power balancing for renewables.

Reference [17] is the most closely related our approach. We differ in the following ways. We aim to track a changing setpoint by adjusting the loads' power consumption whereas, they look at demand flattening. We work with self-reported or measured load responses as opposed to load price sensitivity. We also propose a novel mean-regularizer and minimize the impact on the loads. Our contributions are:

- We formulate a generic online model for setpoint tracking using flexible loads (Section III);
- We propose a mean-regularizer to control and minimize impact of adjustment over the loads (Section III-B2);
- We give an improved Bandit-`COMID` (`BCOGD`) algorithm, theoretically bound its regret, and apply it to our model (Section IV);
- We give numerical examples illustrating the performance of the proposed approaches (Section V).

II. Background

In this section, we set our notation and state the existing results we build on.

We denote the rounds t and the time horizon T. Let N be the number of loads. We denote the sample mean after t

rounds as $\langle\cdot\rangle_t$ and we define $\langle\cdot\rangle_t$ over a vector as the vector of the means.

Define the regularization function $\mathcal{R} : \mathcal{K} \longrightarrow \mathbb{R}$, where $\mathcal{K} \subseteq \mathbb{R}^N$, and define be the Bregman divergence with respect to $\mathcal{R}$ as

$$B_{\mathcal{R}}(\mathbf{x}, \mathbf{y}) = \mathcal{R}(\mathbf{x}) + \mathcal{R}(\mathbf{y}) - \nabla\mathcal{R}(\mathbf{y})^{\mathrm{T}}(\mathbf{x} - \mathbf{y})$$

We assume that the Bregman divergence is α-strongly convex with respect to a norm $\|\cdot\|$. Finally, denote the dual norm $\|\cdot\|_*$ as

$$\|\mathbf{z}\|_* = \sup\left\{\mathbf{z}^{\mathrm{T}}\mathbf{x} \mid \|\mathbf{x}\| \leq 1\right\}$$

A. Online convex optimization

Online Convex Optimization (OCO) can be interpreted as a repeated game with a convex loss function and a convex and compact decision set [6]. The learner provides, to the best of his knowledge, a prediction to minimize a loss function set by the adversary. Suppose an online algorithm produces a sequence of decisions $\mathbf{x}_t$, $t = 1, ..., T$. The regret of this algorithm is

$$R_T = \sup_{\{F_1, F_2, ..., F_T\} \subset \mathcal{L}} \left\{ \sum_{t=1}^{T} F_t(\mathbf{x}_t) - \min_{\mathbf{x} \in \mathcal{K}} \sum_{t=1}^{T} F_t(\mathbf{x}) \right\}, \quad (1)$$

where $\mathcal{K}$ the set of decisions, F_t is the loss function at round t and $\mathcal{L}$ is a the set of loss functions. The loss function for the round t is only revealed after the learner chooses $\mathbf{x}_t$. This regret qualifies the difference between the cumulative loss of the learner and the loss of the optimal solution in hindsight [19]. We let $\mathbf{x}^*$ denote the optimal solution in hindsight (cf. (1), second term of right-hand side) and it represents the best fixed decision. Finally, an online learning algorithm is said to perform well if its regret is sublinear in its number of rounds [6], [5]. An sublinear regret ensures that, in average, the learner does as good as the fixed optimal solution in hindsight.

B. COMID

We apply the Composite Objective MIrror Descent (COMID) algorithm proposed by Duchi et al. [8]. The composite objective function refers to an objective made of a loss function and round-invariant regularizer terms. COMID is chosen for its strong performance in presence of regularizers. Its update is based on the online mirror descent algorithm but handles the composite objective function differently. Specifically, it uses only the gradient of the loss function and not of the regularizer. This latter term is then incorporated into in the projection step.

We re-writte [8] the loss function $F_t(\mathbf{x})$ as,

$$F_t(\mathbf{x}) = f_t(\mathbf{x}) + r(\mathbf{x}),\ \forall\ t \quad (2)$$

where $f_t(\mathbf{x})$ is the round-dependent loss function and $r(\mathbf{x})$ the regularizer. The COMID update is given by,

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x} \in \mathcal{K}} \eta \nabla_{\mathbf{x}} f_t(\mathbf{x}_t)^{\mathrm{T}}\mathbf{x} + B_{\mathcal{R}}(\mathbf{x}, \mathbf{x}_t) + \eta r(\mathbf{x}) \quad (3)$$

where η is a numerical parameter. The following lemma establishes COMID's sublinear regret. It is essentially Corollary 4 in [8] specialized to a choice of η that ensures large enough step sizes for our application.

Lemma 1 (Regret bound for COMID)**.** *Let $\{f_t(\boldsymbol{\mu}_t)\}_{t=1,...,T}$ be a sequence of L-Lipschitz functions with $\|\nabla f_t(\boldsymbol{\mu}_t)\|_* \leq G_*$ and $r(\boldsymbol{\mu}_1) = 0$. Set*

$$\eta = \chi\sqrt{\frac{2\alpha B_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1)}{G_*^2 T}} \quad (4)$$

where $\chi \geq 1$ is a tuning parameter. Then, the COMID update leads to the upper bound

$$R_T(\texttt{COMID}) \leq \sqrt{\frac{2TB_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1)G_*^2\chi^2}{\alpha}} \quad (5)$$

Proof. Define $\eta = \chi\eta_{COMID}$ where χ is a tuning parameter and η_{COMID} is defined as in Corollary 4 of [8]. Corollary 3 of [8] gives the following upper bound on COMID's regret:

$$R_T(\texttt{COMID}) \leq \frac{B_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1)}{\eta} + \frac{T\eta G_*^2}{2\alpha} + r(\boldsymbol{\mu}_1) \quad (6)$$

By assumption, $r(\boldsymbol{\mu}_1) = 0$, then we set η as in (4). Substituting (4) in (6) leads to,

$$R_T(\texttt{COMID}) \leq \frac{\sqrt{B_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1)G_*^2 T}}{\sqrt{2\alpha}\chi} + \frac{\chi\sqrt{B_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1)G_*^2 T}}{\sqrt{2\alpha}} \quad (7)$$

Then, since $\chi \geq 1$, we have,

$$R_T(\texttt{COMID}) \leq \sqrt{\frac{2TB_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1)G_*^2\chi^2}{\alpha}}, \quad (8)$$

and we find back [8]'s bound using $\chi = 1$. □

III. Online convex optimization formulation

We formulate an OCO algorithm for setpoint tracking with flexible loads. We first define a loss function and two regularizers.

A. Setpoint tracking loss function

The goal is to track the setpoint s_t with the total flexible consumption of all the loads. Let $\boldsymbol{\mu}_t$ be in the N-dimension compact and convex set $\mathcal{K}$. We let $\mathcal{K} = [-1, 1]^N$. At each time, the aggregator sends the signal $\boldsymbol{\mu}_t$ to the loads. The loads then consume in total $\mathbf{c}_t^T\boldsymbol{\mu}_t$, where $\mathbf{c}_t \in \mathbb{R}^N$ scales the signal $\boldsymbol{\mu}_t$ to the power consumption. Each entry of $\mathbf{c}_t$ represents a load's response to an adjustment signal. The load cannot predict its exact response due to several of source of uncertainty like consumer behavior. $\mathbf{c}_t$ is modeled as unknown and its real value can only be observed after round t. Hence, the aggregator has to come up with a decision based on previous time steps using OCO for the coming round. The loss is

$$\ell_t(\boldsymbol{\mu}_t) = \left(s_t + \mathbf{c}_t^{\mathrm{T}}\boldsymbol{\mu}_t\right)^2, \quad (9)$$

which is the square setpoint tracking error. The aggregator observes $\mathbf{c_t}$ after suffering the loss which corresponds to assessing each loads response.

```
Parameters: Given $T$, $\lambda$, $\rho$ and $\chi$.
Initialization: Set $\boldsymbol{\mu}_1 = \mathbf{0}$ and set $\eta$ as in (14).
for $t = 1, 2, \dots, T$ do
    Deploy adjustment according to $\boldsymbol{\mu}_t$.
    Compute the gradient,
        $\mathbf{w}_t = 2\mathbf{c}_t(s_t + \mathbf{c}_t^{\mathrm{T}}\boldsymbol{\mu}_t) + \frac{2\rho}{t}\frac{(t-1)\langle\boldsymbol{\mu}\rangle_{t-1}+\boldsymbol{\mu}_t}{t}.$
    Update load dispatch,
        $\boldsymbol{\mu}_{t+1} = \arg\min_{\boldsymbol{\mu}\in[-1,1]^N} \eta\mathbf{w}_t^{\mathrm{T}}\boldsymbol{\mu} + \frac{1}{2}\left\|\boldsymbol{\mu}_t-\boldsymbol{\mu}\right\|_2^2 + \eta\lambda\left\|\boldsymbol{\mu}\right\|_1.$
end for
```

Fig. 1. COGD for setpoint tracking algorithm

B. Regularizers

1) Sparsity regularizer: We use an ℓ^1-norm to promote sparsity when computing the signal $\boldsymbol{\mu}_t$ sent to the load. The motivation is to mobilize a minimum number of loads in each time period.

2) Mean regularizer: We use a mean-regularizer to promote a zero-mean adjustment to the loads' power consumption. The regularizer is defined as

$$\|\langle\boldsymbol{\mu}\rangle_t\|_2^2 = \left\|\frac{1}{t}\sum_{s=1}^{t}\boldsymbol{\mu}_s\right\|_2^2 = \left\|\frac{(t-1)\langle\boldsymbol{\mu}\rangle_{t-1}+\boldsymbol{\mu}_t}{t}\right\|_2^2.$$

Because this regularizer is round-dependent, it must be incorporated into the loss function f_t in the COMID update.

3) Total loss function: The total loss function for the setpoint tracking problem at time t is

$$F_t(\boldsymbol{\mu}_t) = \left(s_t - \mathbf{c}_t^{\mathrm{T}}\boldsymbol{\mu}_t\right)^2 + \lambda\left\|\boldsymbol{\mu}_t\right\|_1 + \rho\left\|\langle\boldsymbol{\mu}\rangle_t\right\|_2^2, \tag{10}$$

where λ and ρ are numerical parameters.

C. COMID formulation

We solve this problem using COMID. The corresponding quantities in the COMID update (2) are

$$f_t(\boldsymbol{\mu}_t) = \left(s_t + \mathbf{c}_t^{\mathrm{T}}\boldsymbol{\mu}_t\right)^2 + \rho\left\|\langle\boldsymbol{\mu}\rangle_t\right\|_2^2 \tag{11}$$

$$r(\boldsymbol{\mu}_t) = \lambda\left\|\boldsymbol{\mu}_t\right\|_1 \tag{12}$$

We let $\mathcal{R}(\cdot) = \frac{1}{2}\left\|\cdot\right\|_2^2$, in which case the Bergman divergence reduces to $B_{\mathcal{R}}(\mathbf{x},\mathbf{y}) = \frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|_2^2$. For the rest of this work, all norms are assumed to be Euclidean norm unless otherwise stated. The COMID update (3) then reduces to the Online Gradient Descent (OGD) with special consideration for regularizer. We refer to this special case of COMID as the Composite Objective Gradient Descent (COGD). The update is

$$\boldsymbol{\mu}_{t+1} = \arg\min_{\boldsymbol{\mu}\in[-1,1]^N}\left\{\frac{1}{2}\left\|\boldsymbol{\mu}_t-\boldsymbol{\mu}\right\|_2^2 + \eta\lambda\left\|\boldsymbol{\mu}\right\|_1 \right. \tag{13}$$
$$\left. + \eta\left[+2\mathbf{c}_t(s_t + \mathbf{c}_t^{\mathrm{T}}\boldsymbol{\mu}_t) + \frac{2\rho}{t}\frac{(t-1)\langle\boldsymbol{\mu}\rangle_{t-1}+\boldsymbol{\mu}_t}{t}\right]^{\mathrm{T}}\boldsymbol{\mu}\right\},$$

and is solved numerically using cvx [20], [21]. The following proposition ensures that the setpoint tracking algorithm given in Fig. 1 has a sublinear regret.

Proposition 1. *Let $\boldsymbol{\mu}_1 = \mathbf{0}$ and*

$$\eta = \chi\sqrt{\frac{4N}{G^2T}}. \tag{14}$$

the regret of the COGD for setpoint tracking algorithm (cf. Fig. 1) is bounded by,

$$R_T(\texttt{COGD}) \le 4\chi\sqrt{TKB}, \tag{15}$$

where $f_t(\boldsymbol{\mu}_t) \le B$ for all t and $K = \max_{t=1,2,\dots,T}\{\rho^2, \left\|\mathbf{c}_t\right\|^2\}$.

Proof. The result follows from Lemma 1. First, the Bregman divergence with respect to the Euclidean norm is 1-strongly convex and hence $\alpha = 1$. We upper bound the Bregman divergence by the diameter D of the compact set $\mathcal{K}$ with respect to the ℓ^2-norm.

$$\begin{aligned} B_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1) &= \frac{1}{2}\left\|\boldsymbol{\mu}^* - \boldsymbol{\mu}_1\right\|^2, \\ &\le \frac{1}{2}D^2, \end{aligned}$$

where $D = \operatorname{diam}\mathcal{K} = \sup\{\,\|\mathbf{x}-\mathbf{y}\|\,|\,\mathbf{x},\mathbf{y}\in\mathcal{K}\}$. Hence, with $\mathcal{K} = [-1,1]^N$, we have $D = 2\sqrt{N}$. The gradient of the loss function is upper-bounded as

$$\begin{aligned} \left\|\nabla f_t(\boldsymbol{\mu})\right\|^2 &= \left\|\nabla\ell(\boldsymbol{\mu}) + \nabla\rho\left\|\langle\boldsymbol{\mu}\rangle_t\right\|^2\right\|^2, \\ &\le \left\|\nabla\ell(\boldsymbol{\mu})\right\|^2 + \left\|\nabla\rho\left\|\langle\boldsymbol{\mu}\rangle_t\right\|^2\right\|^2, \\ &\le 4\left\|\mathbf{c}_t\right\|^2\ell(\boldsymbol{\mu}) + 4\left(\frac{\rho}{t}\right)^2\left\|\langle\boldsymbol{\mu}\rangle_t\right\|^2, \\ &\le 4\left\|\mathbf{c}_t\right\|^2\ell(\boldsymbol{\mu}) + 4\rho^2\left\|\langle\boldsymbol{\mu}\rangle_t\right\|^2, \\ &\le 4\max_{t=1,2,\dots,T}\left\{\left\|\mathbf{c}_t\right\|_2^2, \rho\right\}\left(\ell(\boldsymbol{\mu}) + \rho\left\|\langle\boldsymbol{\mu}\rangle_t\right\|^2\right), \\ &= 4\max_{t=1,2,\dots,T}\left\{\left\|\mathbf{c}_t\right\|^2, \rho\right\}f(\boldsymbol{\mu}), \end{aligned}$$

which yields to the regret given in (15). □

Remark 1. *The constant K in bound* (15) *reflects the priority given of the objective function. If, $\max_{t=1,2,\dots,T}\left\|\mathbf{c}_t\right\|^2 > \rho$, then the bound is dominated by the setpoint tracking objective. If $\rho > \max_{t=1,2,\dots,T}\left\|\mathbf{c}_t\right\|^2$, then priority is given to load comfort.*

IV. Bandit feedback

We now extend our approach to a limited feedback setting where the aggregator does not have access to all information. Practically, the aggregator may not have access to the load parameters, $\mathbf{c}_t$, due to privacy issues, lacking communication infrastructure or inability of a load to assess its response. This means that the aggregator cannot compute the gradient required for the load adjustment in COMID. The aggregator instead observes the total load adjustment, $\mathbf{c}_t^{\mathrm{T}}\boldsymbol{\mu}_t$. We refer to this as bandit feedback.

We use the approach of [22] and [6] to define an observable random variable $\mathbf{g}_t$ such that $\mathbb{E}[\mathbf{g}_t] \approx \nabla f_t$ for any convex loss function using a single evaluation at each round (cf. Lemma 6.4 [6]). By specializing to quadratic functions, we obtain the slightly stronger result $\mathbb{E}[\mathbf{g}_t] = \nabla f_t(\boldsymbol{\mu}_t)$.

Lemma 2 (Point-wise gradient estimator of quadratic functions)**.** *Let $h(\boldsymbol{\mu}) = \boldsymbol{\mu}^{\mathrm{T}} Q \boldsymbol{\mu} + \mathbf{p}^{\mathrm{T}}\boldsymbol{\mu} + r$, where $\boldsymbol{\mu}, \mathbf{p} \in \mathbb{R}^N$, $Q \in \mathbb{R}^{N\times N}$ and $r \in \mathbb{R}$. Define the point-wise gradient estimator*

$$\mathbf{g} = \frac{N}{\delta} h(\boldsymbol{\mu} + \delta \mathbf{v})\mathbf{v} \tag{16}$$

where $\delta > 0$ and $\mathbf{v}$ is a random variable sampled from

$$\mathbb{S}_1 = \left\{ \mathbf{v} \in \mathbb{R}^N \,\middle|\, \|\mathbf{v}\|_2 = 1 \right\},$$

the surface of a N-dimension hypersphere of radius 1. Then,

$$\mathbb{E}_{\mathbf{v}\sim\mathbb{S}_1}[\mathbf{g}] = \nabla h(\boldsymbol{\mu}). \tag{17}$$

The proof of Lemma 2 is given in Section A of the appendix. Next, let,

$$\mathbf{g}_t = \frac{N}{\delta} f_t(\boldsymbol{\mu}_t + \delta \mathbf{v}_t)\mathbf{v}_t \tag{18}$$

where f_t is defined as in (11), $\delta > 0$ and $\mathbf{v}_t$ is sampled uniformly from $\mathbb{S}_1$. By Lemma 2,

$$\mathbb{E}_{\mathbf{v}\sim\mathbb{S}_1}[\mathbf{g}_t] = \nabla f_t(\boldsymbol{\mu}_t). \tag{19}$$

We can now use (18) as the gradient in the COMID update. To ensure that the gradient is evaluated inside $\mathcal{K}$, define

$$\begin{aligned}\mathcal{K}^{1-\delta} &\equiv \left\{ \boldsymbol{\mu} \,\middle|\, \frac{\boldsymbol{\mu}}{(1-\delta)} \in \mathcal{K} \right\}, \\ &= [\delta - 1, 1-\delta]^N .\end{aligned} \tag{20}$$

The update for the bandit version of COGD (BCOGD) is given by

$$\boldsymbol{\mu}_{t+1} = \underset{\boldsymbol{\mu}\in\mathcal{K}^{1-\delta}}{\arg\min}\, \eta \mathbf{g}_t^{\mathrm{T}}\boldsymbol{\mu} + \frac{1}{2}\|\boldsymbol{\mu}_t - \boldsymbol{\mu}\|_2^2 + \eta r(\boldsymbol{\mu}), \tag{21}$$

The algorithm is presented in Fig. 2.

We now bound the regret of BCOGD. We base the proof of the next result on Theorem 6.6 of [6] to take into account the point-wise gradient estimator in addition in the COGD algorithm. A similar proof was proposed by [17] which uses [22] to deal with the gradient. However, this approach imposes a lower bound on the time horizon T. In [17], the bandit algorithm is limited to cases where the time horizon satisfies

$$T \geq \frac{1}{\rho_{in}^4}\left(\frac{\rho_{out} BN}{L + \frac{2C}{\rho_{in}}}\right)^2 .$$

Not meeting this condition is equivalent to setting δ in (20) greater than one, which leads to an empty feasible set when updating the prediction. This time horizon condition also makes η, the gradient descent step, potentially very small, leading to insufficient change between each round. Our proposed BCOGD is not subject to this limitation and can be used for any time horizon. In our improved BCOGD, this issue is eliminated because the time horizon need not be large and the tuning parameter, χ, allows more control over the step size.

Theorem 1 (Regret bound for BCOGD)**.** *Let $\{F_t(\mu_t)\}_t$ be a sequence of L-Lipschitz, B-bounded functions and $r(\boldsymbol{\mu}_1) = 0$. Using using the point-wise gradient estimator $\mathbf{g}_t$ and setting*

$$\eta = \frac{D\chi}{BNT^{\frac{3}{4}}} \tag{22}$$

$$\delta = \frac{1}{T^{\frac{1}{4}}} \tag{23}$$

where $D = \operatorname{diam}\mathcal{K}$ and $\chi \geq 1$, the BCOGD regret is upper bounded by,

$$\mathbb{E}[R_T(\texttt{BOGD})] \leq (DBN\chi + 2DL + 2L)T^{\frac{3}{4}}. \tag{24}$$

The proof of Theorem 1 is given in Section B of the appendix. Theorem 1's $O\left(T^{\frac{3}{4}}\right)$ bound is looser than the COGD bound in Lemma 1 but still sublinear. This bound is similar to other bandit online convex optimization bound [6], [17], [22].

We evaluate the bound in Theorem 1 by setting $\boldsymbol{\mu}_1 = \mathbf{0}$ and η and δ as in (22) and (23). The regret of BCOGD for setpoint tracking (cf. Fig. 2) is

$$\mathbb{E}\left[R_T(\texttt{BOGD})\right] \leq \left(2N^{\frac{3}{2}}B\chi + 4\sqrt{N}L\right)T^{\frac{3}{4}}. \tag{25}$$

The bound follows from Corollary 1 with $D = 2\sqrt{N}$. Using Lemma 2, the third term in the parenthesis of the bound drops since F_t is a smooth function. Thus, step (29) of the proof in the appendix is unnecessary.

By adding a tuning parameter, χ, to η, we can ensure a proper step size for the update while keeping a sub-linear upper-bounded regret. The consequence are better round-to-round performances and, however, a looser regret bound.

V. Numerical results

We run numerical simulations for $T = 600$ rounds which correspond to 10 hours of one minute time periods. We let $N = 100$ loads and we set the adjustment parameter of each individual load to be

$$c_t(i) = c_0(i) + w_t(i),$$

where $c_0(i)$ is the average response of each load and $w_t(i)$ is a zero-mean noise. We sample $c_0(i) \sim \mathrm{U}[1,5]$ and set $w_t \sim \mathrm{N}_{[-1,1]}(0, \frac{1}{2})$, a truncated normal distribution, for all t. Finally, we let $s_t = 50\sin(0.1t)$, the setpoint to track with the load aggregation.

Fig. 3 compares the total setpoint tracking loss of: (i) no online algorithm given by the blue line, (ii) the bandit feedback algorithm with regularizers given by the solid red line, (iii) the bandit feedback algorithm without regularizer given by the dotted red line (iv) the full information algorithm given by solid green line and (v) the full information without regularization represented by dotted green line. Note that, because regularizers are intended to promote sparsity and

```
**Parameters:** Given $T$, $\rho$, $\lambda$ and $\chi$.
**Initialization:** Set $\boldsymbol{\mu}_1 = \mathbf{0}$ and set $\eta$ and $\delta$ according to (22) and (23) respectively.
**for** $t = 1, 2, \ldots, T$ **do**
    Sample $\mathbf{v}_t \sim \mathbb{S}_1$.
    Deploy adjustment according to $\boldsymbol{\mu}_t + \delta \mathbf{v}_t$.
    Suffer loss $\ell_t(\boldsymbol{\mu}_t + \delta \mathbf{v}_t)$.
    Compute the point-wise gradient,
        $$\mathbf{g}_t = \frac{N}{\delta} f_t(\boldsymbol{\mu}_t + \delta \mathbf{v}_t)\mathbf{v}_t.$$
    Update load dispatch,
        $$\boldsymbol{\mu}_{t+1} = \underset{\boldsymbol{\mu} \in [\delta-1, 1-\delta]^N}{\arg\min} \left\{ \eta \mathbf{g}_t^{\mathrm{T}} \boldsymbol{\mu} + \frac{1}{2} \left\| \boldsymbol{\mu}_t - \boldsymbol{\mu} \right\|_2^2 + \eta\lambda \left\| \boldsymbol{\mu} \right\|_1 \right\}.$$
**end for**
```

Fig. 2. BCOGD for setpoint tracking with limited feedback algorithm

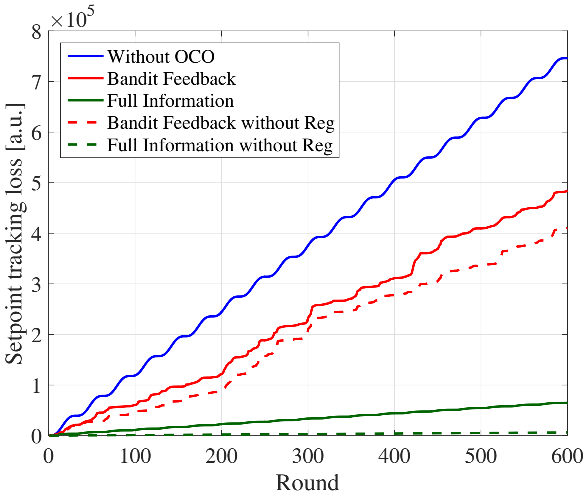


Fig. 3. Loss comparison between the full information, the bandit feedback settings and no OCO algorithm

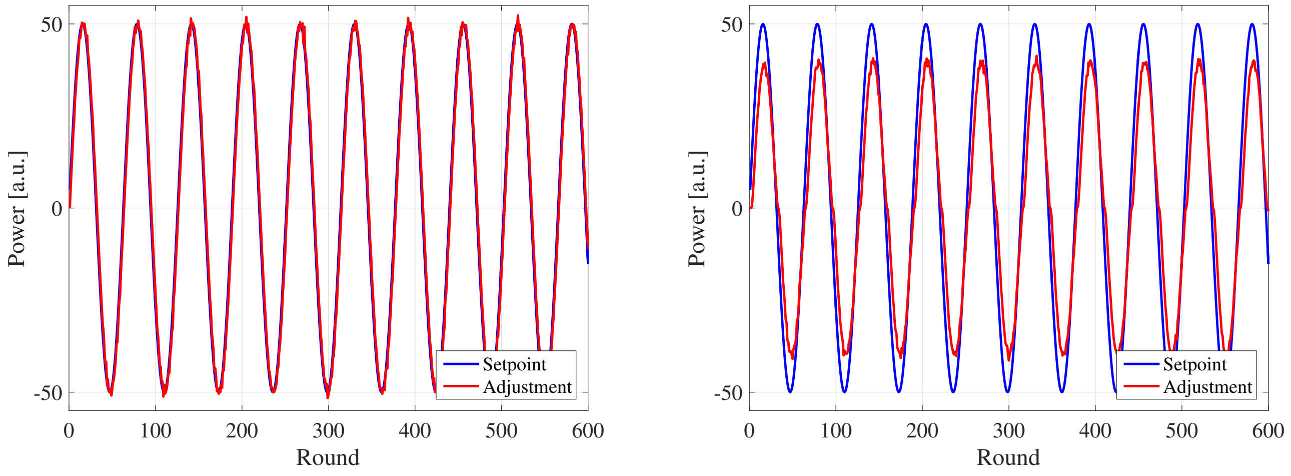
minimum impact rather than improve tracking, it slightly increases the loss.

In the full information setting, the aggregator observes individual load's adjustment when the current round ends and can, therefore, compute a gradient. We set the two regularization parameters $\rho = 750$ and $\lambda = 80$ and the tuning parameter to $\chi = 25$. In Fig. 3, the solid green curve shows great improvement which is translated by a decrease of 91.29% in the total incurred setpoint tracking loss using OCO with regularization. Fig. 4 presents the setpoint tracking ability of the proposed approach when $\mathbf{c}_t$ is uncertain with and without regularization. Fig. 4a, presents the performance of COGD for setpoint tracking without regularization. The setpoint tracking loss are decreased by 99.16%. Then, Fig. 4b shows how the algorithm can track the setpoint in the presence of regularization. We note, when comparing, that the aggregator will never commit loads to their maximum/minimum response because of the regularizer terms in place and thus will lead to higher loss near local optima. The regularizer parameters have hence to be chosen to balance the trade-off between a low loss, a high sparsity, and a minimum impact. Using $\rho = 750$ and $\lambda = 80$, we improve, in average by round, the sparsity by 49.00% and impact by 78.54% with respect to simulation ran without regularizers. These figures are obtained by computing the ℓ_2-norm of the mean of $\boldsymbol{\mu}$ up to the current round and the ℓ_1-norm of each signal and comparing them to the case with no regularization. Note that the sparsity regularizer also reduces $\left\| \langle \boldsymbol{\mu} \rangle_t \right\|^2$ since it promotes lower $\boldsymbol{\mu}$ values. The mean-regularizer then allows a larger decrease on the impact while having only a limited effect on the setpoint tracking loss. Fig. 5 illustrate the sparsity and impact improvement due to the regularizers. In Fig. 5a each dot corresponds to a non-zero signal sent to a load with a tolerance of 10^{-8} for selected loads. Without regularizer, all signals are non-zero for all rounds except for the first one due to the initialization. Fig. 5b compares the evolution of $\left\| \langle \boldsymbol{\mu} \rangle_t \right\|^2$ with and without regularization as a function of the rounds. It shows that the impact is minimized and no drift in the mean is observed using regularization.

For the limited feedback extension, the aggregator only observes the total load adjustment. Using the bandit OCO algorithm for DR with $\chi = 2300$, $\rho = 100$ and $\lambda = 2500$, we obtained a total setpoint tracking loss reduction of 34.91% which represents 38.24% of the performance of the full information setting. In addition, the regularizers improved the sparsity and impact with respect to the bandit without regularization of 18.23% and 38.57% respectively in average by round. Hence, these numbers show the adequate performance of the OCO algorithm and this, even when the assumptions on the feedback are weakened. Lastly, the setpoint tracking is illustrated on Fig. 6. As shown in this figure, the dispatched adjustment curve is noisy due to the estimated gradient but still, properly follows the signal.

VI. Conclusion

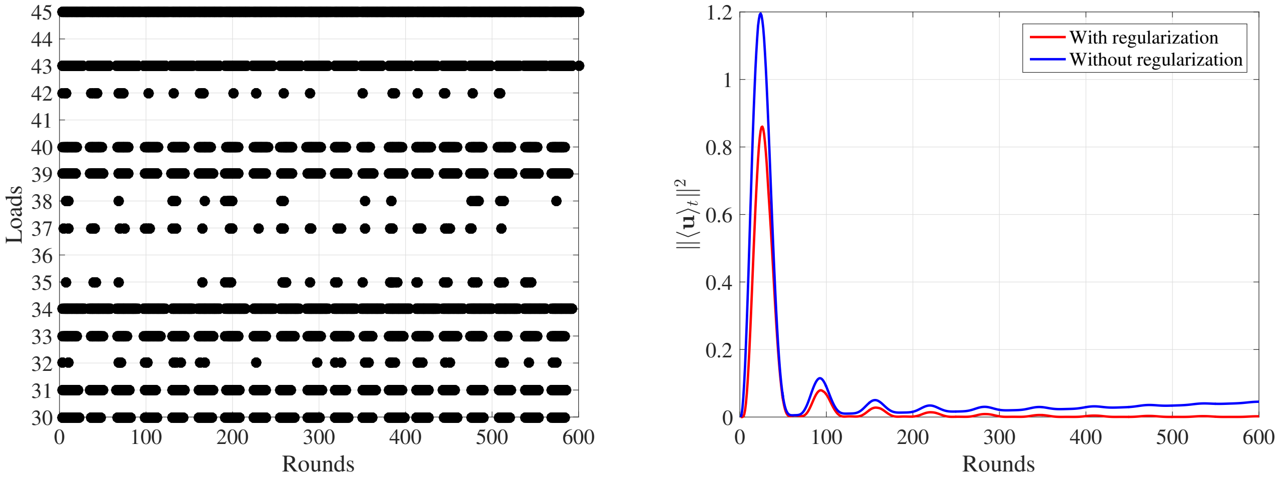
We have proposed an OCO-based algorithm for setpoint tracking algorithm using flexible loads. We used a sparsity regularizer to minimize the number of loads dispatched each round and a mean-regularizer to minimize the long-run impact of the DR program on each load's comfort. We provided an improved bandit version of the COMID algorithm, which enabled the aggregator to proceed without individual knowledge of the load. We demonstrated the performance of our approach numerically for both the full information and limited feedback setting. These simulations showed that using regularizers, we can minimize the number of dispatched loads and the impact on them while still effectively tracking a time-varying setpoint.



(a) Without regularization

(b) With regularization

Fig. 4. Setpoint tracking in full information setting



(a) Rounds with non-zero valued signal sent for selected loads (tolerance 10^{-8})

(b) Mean-regularizer performance comparison

Fig. 5. Regularizers performance in the full information setting

Appendix

A. Proof of Lemma 2

We first compute the expectation of the point-wise gradient estimator over the random variable $\mathbf{v} \sim \mathbb{S}_1$.

$$\begin{aligned}\mathbb{E}_{\mathbf{v}\sim\mathbb{S}_1}[\mathbf{g}] &= \frac{N}{\delta}\,\mathbb{E}\left[\big((\boldsymbol{\mu}+\delta\mathbf{v})^{\mathrm{T}}Q(\boldsymbol{\mu}+\delta\mathbf{v})\right.\\ &\quad\left. + \,\mathbf{p}^{\mathrm{T}}(\boldsymbol{\mu}+\delta\mathbf{v})+r\big)\,\mathbf{v}\right]\\ &= \frac{N}{\delta}\,\mathbb{E}\left[\big(\boldsymbol{\mu}^{\mathrm{T}}Q\boldsymbol{\mu}+\delta\mathbf{v}^{\mathrm{T}}Q\boldsymbol{\mu}+\delta\boldsymbol{\mu}^{\mathrm{T}}Q\mathbf{v}\right.\\ &\quad\left. +\delta^2\mathbf{v}^{\mathrm{T}}Q\mathbf{v}+\ \mathbf{p}^{\mathrm{T}}\boldsymbol{\mu}+\delta\mathbf{p}^{\mathrm{T}}\mathbf{v}+r\big)\,\mathbf{v}\right]\\ &= \frac{N}{\delta}\,\mathbb{E}\left[\big(\boldsymbol{\mu}^{\mathrm{T}}Q\boldsymbol{\mu}+\mathbf{p}^{\mathrm{T}}\boldsymbol{\mu}+r\big)\,\mathbf{v}\right]\\ &\quad+\frac{N}{\delta}\,\mathbb{E}\left[\delta^2\mathbf{v}^{\mathrm{T}}Q\mathbf{v}\ \mathbf{v}\right]\\ &\quad+\frac{N}{\delta}\,\mathbb{E}\left[\big(\delta\mathbf{v}^{\mathrm{T}}Q\boldsymbol{\mu}+\delta\boldsymbol{\mu}^{\mathrm{T}}Q\mathbf{v}+\delta\mathbf{p}^{\mathrm{T}}\mathbf{v}\big)\,\mathbf{v}\right]\\ &= \frac{N}{\delta}\,\big(\boldsymbol{\mu}^{\mathrm{T}}Q\boldsymbol{\mu}+\mathbf{p}^{\mathrm{T}}\boldsymbol{\mu}+r\big)\,\mathbb{E}\left[\mathbf{v}\right]\\ &\quad+N\delta\,\mathbb{E}\left[\mathbf{v}^{\mathrm{T}}Q\mathbf{v}\ \mathbf{v}\right]\\ &\quad+N\,\mathbb{E}\left[\mathbf{v}\mathbf{v}^{\mathrm{T}}\right]\big(Q\boldsymbol{\mu}+Q^{\mathrm{T}}\boldsymbol{\mu}+\mathbf{p}\big) \qquad (26)\end{aligned}$$

We must to compute the expected values with respect to $\mathbf{v} \sim \mathbb{S}_1$.

We first observe that by symmetry of the uniform distribution over $\mathbb{S}_1$, $\mathbb{E}_{\mathbf{v}}[\mathbf{v}]$ and $\mathbb{E}_{\mathbf{v}}[\mathbf{v}^{\mathrm{T}}Q\mathbf{v}\ \mathbf{v}]$ are equal to $\mathbf{0}$.

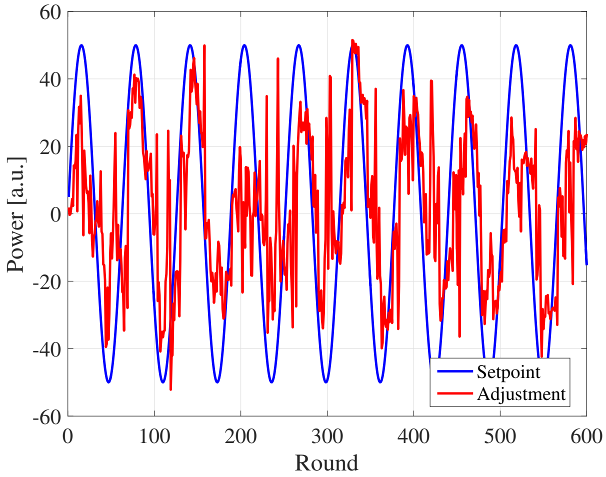


Fig. 6. Setpoint tracking in the bandit setting

For $\mathbb{E}_{\mathbf{v}}[\mathbf{v}\mathbf{v}^{\mathrm{T}}]$, let $\mathbf{v} \sim \mathbb{S}_1$ and

$$\begin{aligned}\mathbf{v}_a =&(v_1, v_2, \ldots, v_i, \ldots, v_N)^{\mathrm{T}}\\ \mathbf{v}_b =&(v_1, v_2, \ldots, -v_i, \ldots, v_N)^{\mathrm{T}}\end{aligned}$$

where v_i is the i^{th} component of $\mathbf{v}$. Then, both $\mathbf{v}_a$ and $\mathbf{v}_b$ are uniformly distributed on $\mathbb{S}_1$. Thus, they must have the same correlation $\mathbb{E}[\mathbf{v}\mathbf{v}^{\mathrm{T}}] = \mathbb{E}[\mathbf{v}_a\mathbf{v}_a^{\mathrm{T}}] = \mathbb{E}[\mathbf{v}_b\mathbf{v}_b^{\mathrm{T}}]$ which implies that for $i \neq j$,

$$\begin{aligned}\mathbb{E}[v_i v_j] &= \mathbb{E}[(-v_i)v_j],\\ &= -\mathbb{E}[v_i v_j],\\ &= 0,\end{aligned}$$

and for $i = j$,

$$\mathbb{E}[v_i v_i] = \mathbb{E}[(v_i)^2].$$

Hence, $\mathbb{E}[\mathbf{v}\mathbf{v}^{\mathrm{T}}]$ is a diagonal matrix. Finally, using Theorem 2.1 (c) of [23], we have,

$$|v_i|^p \sim \mathrm{Beta}\left(\frac{1}{p}, \frac{N-1}{p}\right)$$

Setting $p = 2$,

$$\mathbb{E}[(v_i)^2] = \frac{1}{N}$$

for $i = 1, 2, \ldots, N$. It follows that

$$\mathbb{E}[\mathbf{v}\mathbf{v}^{\mathrm{T}}] = \frac{1}{N}\mathbf{I}$$

Substituting this into the expectation, we obtain the desired result,

$$\mathbb{E}_{\mathbf{v}}[\mathbf{g}] = (Q + Q^{\mathrm{T}})\boldsymbol{\mu} + \mathbf{p},$$

which is the gradient of $f(\boldsymbol{\mu})$.

B. Proof of Theorem 1

We apply the proof technique used in Theorem 6.6 of [6] and substitute the COGD for the standard online gradient descent regret. We observe the following three relations.

First, let the algorithm optimum be $\hat{\boldsymbol{\mu}}^*$, the projection of $\boldsymbol{\mu}^* \in \mathcal{K}$ onto $\mathcal{K}^{1-\delta}$. $\mathcal{K}^{1-\delta}$ is a subset of $\mathcal{K}$ where all directions are reduced by a factor of δ. Thus, along every axis of the subset, the reduction is less or equal to δD, with equality in the direction that leads to the diameter. The point $\hat{\boldsymbol{\mu}}^*$ is the closest point (with respect to the chosen norm) in $\mathcal{K}^{1-\delta}$ to $\boldsymbol{\mu}^*$ by the definition of the projection and lies on the boundary of $\mathcal{K}^{1-\delta}$. Hence,

$$\|\hat{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|_2 \leq \delta D$$

This relation upper-bounds the difference between the optimal point and the closest one can expect in the feasible set. By the Lipschitz assumption,,

$$|F_t(\hat{\boldsymbol{\mu}}^*) - F_t(\boldsymbol{\mu}^*)| \leq \delta L D$$

Second, using the above reasoning, we find that

$$|F_t(\boldsymbol{\mu}_t + \delta\mathbf{v}_t) - F_t(\boldsymbol{\mu}_t)| \leq \delta L D,$$

for any $\boldsymbol{\mu}_t \in K^{1-\delta}$ and $\mathbf{v}_t \sim \mathbb{S}_1$ This bounds the difference between the BCOGD update in (21) and the actual decision.

Third, the definition of the point-wise gradient $\mathbf{g}_t$ is valid for any δ-smoothed function by Lemma 6.4 of [6]. The reader is referred to [6] for the definition of δ-smoothed functions. Let $\hat{F}_t$ be the δ-smoothed version of F_t. Then, since F_t is L-Lipschitz, Lemma 2.6 of [6] ensures that

$$\left|\hat{F}_t(\boldsymbol{\mu}_t) - F_t(\boldsymbol{\mu}_t)\right| \leq \delta L$$

Using, these three relations, we can rewrite the bandit-COGD regret as,

$$\begin{aligned}\mathbb{E}\left[R_T(\mathrm{BCOGD})\right] &= \mathbb{E}\left[\sum_{t=1}^{T} F_t(\boldsymbol{\mu}_t + \delta\mathbf{v}_t) - F_t(\boldsymbol{\mu}^*)\right]\\ &= \sum_{t=1}^{T} \mathbb{E}\left[F_t(\boldsymbol{\mu}_t + \delta\mathbf{v}_t)\right] - F_t(\boldsymbol{\mu}^*)\\ &\leq \sum_{t=1}^{T} \mathbb{E}\left[F_t(\boldsymbol{\mu}_t)\right] - F_t(\boldsymbol{\mu}^*) + \delta D L \qquad (27)\\ &\leq \sum_{t=1}^{T} \mathbb{E}\left[F_t(\boldsymbol{\mu}_t)\right] - F_t(\hat{\boldsymbol{\mu}}^*) + 2\delta D L \qquad (28)\\ &\leq \sum_{t=1}^{T} \mathbb{E}\left[\hat{F}_t(\boldsymbol{\mu}_t)\right] - \hat{F}_t(\hat{\boldsymbol{\mu}}^*)\\ &\quad + 2\delta D L + 2\delta L \qquad (29)\\ &= \mathbb{E}\left[R_T(\mathrm{COGD}; \mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_T)\right]\\ &\quad + \sum_{t=1}^{T} 2\delta D L + 2\delta L\\ &\leq \frac{B_{\mathcal{R}}(\hat{\boldsymbol{\mu}}^*, \boldsymbol{\mu}_1)}{\eta} + \frac{T\eta G^2}{2\alpha} + r(\boldsymbol{\mu}_1)\\ &\quad + 2\delta D L T + 2\delta L T, \qquad (30)\end{aligned}$$

where we use the three previously stated relations to get (27)-(29). Lastly, (6) of Lemma 1 leads to (30) where $\mathbb{E}[\mathbf{g}_t] = \nabla f_t(\boldsymbol{\mu}_t)$ and $\|\mathbf{g}_t\| \leq G$ for all t. Assuming $r(\boldsymbol{\mu}_1) = 0$ and recalling that $B_{\mathcal{R}}(\boldsymbol{\mu}^*, \boldsymbol{\mu}_1) \leq D^2/2$ when using $\mathcal{R}(\cdot) = \frac{1}{2}\left\|\cdot\right\|^2$ and $\alpha = 1$, we have

$$\mathbb{E}\left[R_T(\texttt{BCOGD})\right] \leq \frac{D^2}{2\eta} + \frac{T\eta G^2}{2} + 2\delta DLT + 2\delta LT.$$

Using the definition of the point-wise gradient we have $\|\mathbf{g}_t\| = \left\|\frac{N}{\delta} f_t(\boldsymbol{\mu}_t + \delta\mathbf{v}_t)\mathbf{v}_t\right\| \leq \frac{N}{\delta} B = G$. Hence,

$$\mathbb{E}\left[R_T(\texttt{BCOGD})\right] \leq \frac{D^2}{2\eta} + \frac{N^2 T\eta B^2}{2\delta^2} + 2\delta DLT + 2\delta LT.$$

Setting η and δ according to (23) and (23), we obtain,

$$\begin{aligned}\mathbb{E}\left[R_T(\texttt{BCOGD})\right] \leq & \frac{D^2 BNT^{\frac{3}{4}}}{2D\chi} + \frac{D\chi NBT^{\frac{3}{4}}}{2} \\ & + 2DLT^{\frac{3}{4}} + 2LT^{\frac{3}{4}}, \\ & \leq (DBN\chi + 2DL + 2L)T^{\frac{3}{4}}\end{aligned}$$

Acknowledgment

This work was funded by the Fonds de recherche du Québec – Nature et technologies and the Natural Sciences and Engineering Research Council of Canada.

References

[1] D. S. Callaway and I. A. Hiskens, “Achieving controllability of electric loads,” *Proceedings of the IEEE*, vol. 99, no. 1, pp. 184–199, 2011.

[2] P. Palensky and D. Dietrich, “Demand side management: Demand response, intelligent energy systems, and smart loads,” *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, 2011.

[3] J. A. Taylor, S. V. Dhople, and D. S. Callaway, “Power systems without fuel,” *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 1322–1336, 2016.

[4] P. Siano, “Demand response and smart gridsa survey,” *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[5] S. Shalev-Shwartz, “Online learning and online convex optimization,” *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[6] E. Hazan *et al.*, “Introduction to online convex optimization,” *Foundations and Trends® in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.

[7] J. Taylor and J. Mathieu, *Uncertainty in Demand Response – Identification, Estimation, and Learning*, ser. Tutorials in Operations Research. INFORMS, 2015, ch. 5, pp. 56–70.

[8] J. C. Duchi, S. Shalev-Shwartz, Y. Singer, and A. Tewari, “Composite objective mirror descent.” in *COLT*, 2010, pp. 14–26.

[9] J. A. Taylor and J. L. Mathieu, “Index policies for demand response,” *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1287–1295, 2014.

[10] Q. Wang, M. Liu, and J. L. Mathieu, “Adaptive demand response: Online learning of restless and controlled bandits,” in *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, 2014, pp. 752–757.

[11] D. Kalathil and R. Rajagopal, “Online learning for demand response,” in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 218–222.

[12] S. Bandyopadhyay, P. Kumar, and V. Arya, “Planning curtailment of renewable generation in power grids,” in *Twenty-Sixth International Conference on Automated Planning and Scheduling*, 2016.

[13] A. Lesage-Landry and J. A. Taylor, “Learning to shift thermostatically controlled loads,” in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[14] M. Zinkevich, “Online convex programming and generalized infinitesimal gradient ascent,” in *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*, 2003, pp. 928–936.

[15] N. Y. Soltani, S.-J. Kim, and G. B. Giannakis, “Real-time load elasticity tracking and pricing for electric vehicle charging,” *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1303–1313, 2015.

[16] W.-J. Ma, V. Gupta, and U. Topcu, “Distributed charging control of electric vehicles using online learning,” *IEEE Transactions on Automatic Control*, 2016.

[17] S.-J. Kim and G. Giannakis, “An online convex optimization approach to real-time energy pricing for demand response,” *IEEE Transactions on Smart Grid*, 2016.

[18] G. S. Ledva, L. Balzano, and J. L. Mathieu, “Inferring the behavior of distributed energy resources with online learning,” in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 187–194.

[19] S. Bubeck, “Introduction to online optimization,” *Lecture Notes*, pp. 1–86, 2011.

[20] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” http://cvxr.com/cvx, Mar. 2014.

[21] ——, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/∼boyd/graph_dcp.html.

[22] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, “Online convex optimization in the bandit setting: gradient descent without a gradient,” in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2005, pp. 385–394.

[23] D. Song and A. Gupta, “L_p-norm uniform distribution,” *Proceedings of the American Mathematical Society*, vol. 125, no. 2, pp. 595–601, 1997.