An Online Convex Optimization Approach to Real-Time Energy Pricing for Demand Response

Seung-Jun Kim and Geogios B. Giannakis

***Abstract*—Real-time price setting strategies are investigated for use by demand response programs in future power grids. The major challenge is that consumers have varying degrees of responsiveness to price adjustments at different time instants, which must be learned and accounted for by demand response initiatives. To this end, an online learning approach is developed here offering strong performance guarantees with minimal assumptions on the dynamics of load levels and consumer elasticity, even when consumers are adversarial and take actions strategically. The developed algorithms can determine electricity prices sequentially so as to elicit desirable usage behavior and flatten load curves while implicitly learning individual consumers' price elasticity, based on available feedback information. Two feedback structures are considered: a full information setup, where aggregate load levels as well as individual price elasticity parameters are directly available; and a partial information (bandit) case, where only the aggregate load levels are revealed. Fairness and sparsity constraints are also incorporated via appropriate regularizers. Numerical tests verify the effectiveness of the proposed approach.**

I. Introduction

The smart grid vision aims at revitalizing the aging electric power infrastructure by capitalizing on the state-of-the-art information and power system technologies, making it more efficient, reliable, open, and environment friendly. Ubiquitous sensors and processors, connected by two-way data communication networks, constitute major components in future power grids. Leveraging them allows intelligent grid monitoring, optimization and control, which are expected to bring forth significant benefits for all partakers in power generation, delivery, and consumption [1], [2], [3].

Demand response (DR) is an important smart grid task, in which power consumptions of end users are coordinated so as to elicit economically desirable power usage patterns [4]. Such a coordination can be effected either directly by allowing the load serving entities to exert direct control over consumer loads [5], or indirectly by adjusting electricity prices over time, inducing consumers to adapt their loads accordingly [6], [7], [8]. The pricing mechanism also emerges when devising distributed algorithms that seek optimal load schedules based on some global optimization formulations, in which case it is also assumed that the consumer responses to prices can be programmed in the smart meters at consumer premises [9]. A decentralized day-ahead load scheduling algorithm was proposed in [10] to shift elastic loads to fill the overnight dip in the demand. Simulated annealing was employed in [11] to optimize the prices by the retailer with a limited number of interactions with the customers. In any case, the idea is to raise the price during the periods of peak usage, and lower it at the valleys so that variations in the load curve are abated.

DR becomes effective when more loads are "elastic" in the sense that more consumers are willing to adjust the amount of power used and/or shift their time of use. Some loads are inherently elastic. For example, electric vehicle (EV) charging can wait until the electricity price is right, as long as the desired amount of energy can be accumulated by some specified time [12], [13]. On the other hand, elasticity is related to consumers' behavioral patterns and preferences as well. Some consumers may be more responsive to prices than others, providing more elasticity, while some may be relatively indifferent. Although DR with humans in the chain is welcome to the smart grid's open paradigm promoting consumer participation, it is quite challenging since the patterns may be difficult to model and learn, especially when they can vary drastically owing to changes in individual needs and preferences.

Assuming that the customers make rational decisions, one can adopt a game-theoretic framework to model their strategic behaviors [14], [15]. Energy consumption was scheduled in a game-theoretic setup to minimize the electricity cost and flatten the demand curve [16]; see also [17] for a game-theoretic time-of-use (ToU) pricing approach. A game between the utility, who sets the prices to maximize the revenue, and the EV charging customers, who optimize the charging schedules, was analyzed in a Stackelberg game framework [18]. An auction setup was used to determine the prices, where the energy stored in EV batteries was traded [19]. However, the consumer behaviors are not always rational and often hard to model accurately.

Attempts to capture consumer behavioral patterns in DR via statistical means have appeared recently. In [20], individual consumer responsiveness to price was estimated via linear regression, where the shift in the total load was fit to the price changes announced to individual consumers. Various time series models were adopted to forecast price-responsive loads, based on which a price controller was designed to drive loads to certain desired levels [21]. An incentive design algorithm for a load serving entity (LSE) was proposed in [22] by taking into account the randomness in consumer participation in response to posted incentives. Different consumer behaviors under price-based and incentive-based DR approaches were modeled

S.-J. Kim is with Computer Science and Electrical Engineering (CSEE) Department, University of Maryland, Baltimore County. G. B. Giannakis is with Dept. of Electrical and Computer Engineering, University of Minnesota. This work was supported in part by NSF grants 1423316, 1442686, and 1547347 and by the Institute of Renewable Energy and the Environment (IREE) at the University of Minnesota, under grant No. RL-0010-13. Part of this work was presented at the *IEEE PES Conference on Innovative Smart Grid Technologies*, Washington, D.C., Feb. 19–22, 2014.

in [23]. An algorithm to learn EV charging preferences was developed based on a conditional random field (CRF) model in [24], capturing spatial dependencies among consumers. A common theme in these works is that once a model that allows prediction of consumer behaviors is acquired, it can be exploited subsequently to appropriately set the prices so as to induce favorable load profiles.

In this work, prices are set through learning the consumer responsiveness to prices implicitly and robustly. Our focus is on real-time pricing, where adjustments to the prices announced prior to a scheduling horizon are made during operation to cope with the discrepancy in the actual consumer behavior and the load levels from forecast and planning. Thus, the algorithm should be able to accommodate unmodeled variations underlying the consumer behavior and load levels. Furthermore, relevant measurements may be of limited availability due to privacy or infrastructural/technical constraints [25], [26], and may even contain effects from adversarial behaviors of rational consumers that take strategic actions.

Our approach consists of an online convex optimization (OCO) framework, which can provide performance guarantees with minimal modeling assumptions even against adversarial players [27], [28]. Depending on the scope of available observations necessary for learning and adaptation, two cases are considered. In the full information setting, the (true) elasticity of individual consumers is revealed after the price is set, allowing a gradient descent-type update for prices. In the partial information (bandit) case, only the total load is revealed corresponding to the price set, levying the additional task of estimating the gradient. Desirable patterns in the generated prices, such as fairness and sparsity, can be incorporated in the form of appropriate regularizers.

The OCO framework has attracted significant attention from the machine learning community thanks to its strong performance guarantees without invoking stochastic models, as well as efficient implementation suitable for processing large-scale datasets. In particular, leveraging recent advances in compressive sampling and low-dimensional statistical models, online learning algorithms for composite objectives, which explicitly account for sparsity-promoting regularizers during the learning phase, have been developed for the *full* information case [29], [30], [31]. The present work develops a related novel algorithm tailored for *quadratic* cost functions in the *partial* information case.

OCO can be viewed as a multi-round game between a player (learner) and an opponent (nature). Our conference precursor [32] dealt with the so-called *oblivious* adversary, where the adversarial actions are fixed before the beginning of the game, while the current work considers an *adaptive* adversary, where the adversarial action taken at any time may depend on the entire history of actions taken up to that point. Moreover, detailed proofs for all propositions and lemmas, as well as expanded numerical test results are included here.

The rest of the paper is organized as follows. The DR problem formulation is presented in Sec. II. In Sec. III, the OCO framework is summarized. Real-time pricing algorithms are developed in Sec. IV. Numerical tests are performed in Sec. V. Conclusions are provided in Sec. VI.

II. Problem Formulation

Consider a distribution grid serving K customers. To effectively control load, the utilities can adopt a ToU pricing mechanism, in which the electricity prices $\{\bar{p}_k^t\}$ for individual consumers $k = 1, 2, \ldots, K$ for different times t over a horizon $t = 1, 2, \ldots, T$ (e.g., hourly prices over a day) are announced before the beginning of the horizon. (Dependence on k is generally due to different pricing zones; see also Remark 1 at the end of Sec. II.) An important premise of the smart grid is that based on such information, intelligent scheduling of load becomes feasible, which yields a planned aggregate load profile [33]. However, day-ahead scheduling inherently involves many uncertainties that cannot be always predicted with sufficient accuracy. Indeed, day-ahead prices $\bar{p}_k^t$ are essentially forecast values, which the utility may or may not choose to disclose to the customers; see e.g., [34].

In our proposed scheme, the utility will use $\{\bar{p}_k^t\}$ to estimate consumer elasticity, which will capture the slope of load variation with respect to the price variation, around the forecast prices/loads. In addition, the uncertainty associated with $\bar{p}_k^t$ will be handled via real-time pricing (RTP). Once the time horizon of interest emerges, RTP will allow the utilities to make price adjustments in order to reflect real-time supply and demand imbalances. Let p_k^t denote the price adjustment for consumer k announced at the beginning of time slot t such that the actual price becomes $\bar{p}_k^t + p_k^t$. Since individual consumers have different constitution of elastic loads (such as compressed or rescheduled demands) and distinct preferences, the load-to-price sensitivities vary. Making a first-order approximation that the load change is linear in price variation, the increase (postulated by the utility) in consumer k's load demand at the end of slot t can be modeled as

$$d_k^t = -\theta_k^t p_k^t \tag{1}$$

where $-\theta_k^t$ is the slope, capturing customer k's price elasticity. Note that this is essentially a "small-signal" model, widely used in microeconomics.[1,2] This results in an aggregate load given by

$$l_a^t := l^t + \sum_{k=1}^{K} d_k^t \tag{2}$$

[1]It is true that an automated energy scheduler would in general take into account forecasts and other information available to optimize the load schedule, which would render the load variation nonlinear and correlated across time. It is possible to incorporate these effects in designing the pricing strategies [6], [9]. The main motivation for our approach however, is that it is not easy to capture those effects accurately, especially if the scheduling algorithm involves human preferences and decisions. Thus, we advocate a robust universal approach, designed to work against a wide range of scheduling decisions, including correlated, nonlinear, and even strategic ones, still providing some type of performance guarantees; see Propositions 1 and 2. Note also that severe nonlinearity may be handled through a basis expansion model by postulating $d_k^t = \sum_{i=1}^{N_b} \theta_k^t(i) b_i(p_k^t)$, which is still linear in the basis expansion coefficients $\{\theta_k^t(i)\}_{i=1}^{N_b}$, while nonlinearity is captured through the basis functions $\{b_i(\cdot)\}_{i=1}^{N_b}$. The tradeoff here is between model complexity and performance, which is beyond the scope of this work.

[2]It should be emphasized that the proposed model is transparent to whether the load changes materialize through the load compression or temporal rescheduling. For example, if the EV charging demand of the current time slot can be rescheduled to a later time slot, elasticity of the current time slot increases while that of the later time slot would decrease.

where l^t is the load that would have been realized if there had been no price adjustment. Upon defining $\boldsymbol{\theta}^t := [\theta_1^t, \ldots, \theta_K^t]^\mathsf{T}$ and $\mathbf{p}^t$ likewise, l_a^t can be equivalently expressed as $l^t - {\boldsymbol{\theta}^t}^\mathsf{T} \mathbf{p}^t$, where $\cdot^\mathsf{T}$ denotes transposition. Note that θ_k^t (or d_k^t) might not be directly available to the utility due to privacy and/or technical reasons. Instead, it may be that only l_a^t can be observed at the end of time slot t.

An important goal of the utility is to minimize load variations over time by judicious RTP [12], [13]. This can be achieved by minimizing over $\{\mathbf{p}^t\}$ the variance of the aggregate load around average load levels $\{m^t\}$, given by

$$\frac{1}{2}\sum_{t=1}^{T}\left(l^t - {\boldsymbol{\theta}^t}^\mathsf{T}\mathbf{p}^t - m^t\right)^2 \tag{3}$$

where $m^t := \frac{1}{t}\sum_{\tau=1}^{t} l_a^\tau = \frac{1}{t}[\sum_{\tau=1}^{t-1} l_a^\tau + l_a^t]$ is recursively available through the running average as

$$m^t = \frac{t-1}{t} m^{t-1} + \frac{1}{t} l_a^t\ . \tag{4}$$

One issue with adjusting individual customer prices is *fairness*. It is undesirable to have large adjustments only in few customers' prices since this may conflict with the social expectation that the basic energy needs should be served without discrimination. To ensure that price variation among customers remains small, one can adopt a penalty term discouraging large deviation, such as

$$\sum_{t=1}^{T} ||\mathbf{p}^t||_2^2 := \sum_{t=1}^{T}\sum_{k=1}^{K}(p_k^t)^2 \tag{5}$$

which is essentially the variance of $\{p_k^t\}$ when they are assumed to be zero-mean. Penalty terms enforcing the prices to be close to some reference values can be readily incorporated in a similar fashion [21].

On the other hand, the set of elastic consumers that react sensitively to price adjustments may constitute only a small portion of the entire population [20]. For many consumers and time instants, θ_k^t will remain zero, and thus choosing nonzero p_k^t will have no effect in inducing desired load changes. Also, the LSEs may prefer to select few customers whose prices are adjusted, rather than making small adjustments for all customers. Therefore, it is natural to incorporate such prior knowledge or constraints by promoting *sparsity* in $\mathbf{p}^t$ per slot t, which can be achieved by incorporating the penalty term $\sum_{t=1}^{T} ||\mathbf{p}^t||_1$, where $||\mathbf{p}^t||_1 := \sum_{k=1}^{K} |p_k^t|$ [35].

Upon introducing non-negative weights λ and μ for the sparsity and fairness terms, which are used to control the severity of the constraints, the overall cost per time slot is

$$c^t(\mathbf{p}^t) := \frac{1}{2}\left(l^t - {\boldsymbol{\theta}^t}^\mathsf{T}\mathbf{p}^t - m^t\right)^2 + \lambda||\mathbf{p}^t||_1 + \frac{\mu}{2}||\mathbf{p}^t||_2^2. \tag{6}$$

The goal becomes to attain small total cost $\sum_{t=1}^{T} c^t(\mathbf{p}^t)$ by setting $\{\mathbf{p}^t\}$ appropriately, even when l^t and $\boldsymbol{\theta}^t$ are unknown at the time of determining $\mathbf{p}^t$.

If exact values of l^t and $\boldsymbol{\theta}^t$ for all $t \in \{1, 2, \ldots, T\}$ were available to the utility before the time horizon of interest, optimal $\{\mathbf{p}^t\}$ that minimize $\sum_t c^t(\mathbf{p}^t)$ would be readily obtained by solving a quadratic program (QP). In reality, however, only forecasts of those quantities are available, which are subject to uncertainties. Therefore, it is prudent to adjust prices in an online fashion by observing consumers' reactions to the prices presented to them and learning load elasticity on the fly.

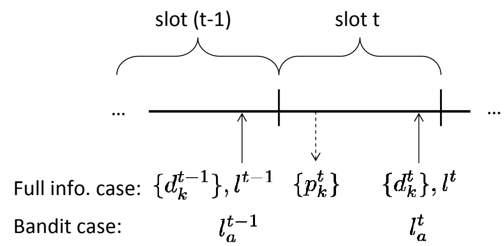

Fig. 1. Timing and types of feedback.

Thus, the problem of interest is as follows. At the beginning of each time slot t, based on the past prices $\mathbf{p}^\tau$, $\tau = 1, 2, \ldots, t-1$, and the observations accumulated up to the time point, set price adjustments $\{p_k^t\}_{k=1}^K$ so that at the end of the horizon, the total cost $\sum_{t=1}^{T} c^t(\mathbf{p}^t)$ is small.

Two types of observations are considered in this context as shown in Fig. 1. In the *full* information case, it is assumed that the utilities can acquire the values of l^t and $\boldsymbol{\theta}^t$ (possibly through observing $\{d_k^t\}_{k=1}^K$) explicitly at the end of each time slot t. If $\{d_k^t\}_{k=1}^K$ are available, one way to estimate $\{\theta_k^t\}_{k=1}^K$ is to use $\hat{\theta}_k^t = -d_k^t/(p_k^t + \varepsilon)$ per (1), where small $\varepsilon > 0$ prevents division by zero. However, gathering full information may require high-bandwidth communication between the utility and the individual customers, and have privacy implications as well. Therefore, in some cases, it is reasonable to consider the *partial* information or *bandit* case, where only the aggregate load l_a^t can be observed at the end of each time slot t.

Remark 1. In our modeling, the prices p_k^t can potentially be different for different "consumers." The index k may correspond to individual consumers, buildings, regions, and even smart grid-capable intelligent electronic devices (IEDs), depending on the pricing granularity and interpretation. In the case of individual consumers, the price differentiation may be realized through various forms of incentives, and could correspond to the actual rate charged, if the consumer signs up with such an arrangement. For IEDs, the prices may just be control signals with economic interpretations. On the other hand, any regulatory and technical requirements on pricing can be readily accommodated through additional constraints/regularizers. The bottom line is that our modeling approach is general and flexible, being compatible with diverse interpretation/incentivization, thus readily accommodating practical deployment constraints.

Remark 2. Regarding communication requirements, the proposed pricing scheme exploits a broadcast communication network from the LSE to the customers to announce the price adjustments p_k^t in the beginning of each time slot t. At the end of each time slot, the resulting load changes must be fed back to the LSE. In the full information case, load changes of individual customers are communicated, requiring a two-way communication network between the LSE and the customers. In the partial information case, only the aggregate load change needs to be reported, which in addition to privacy, it markedly reduces the required communication overhead; that is, the backward communication is needed only from the aggregation point to the LSE.

III. Online Convex Optimization Approach

A. Online Convex Optimization Models

For online price setting, an OCO approach is pursued. The framework considers a multi-round game between a learner and an adversary [27]. In our setup, the utility plays the role of a learner, while consumers are the adversaries. In round t, the learner plays an action $\mathbf{p}^t \in \mathcal{P}$, where $\mathcal{P}$ is closed and convex. Using feedback information $\mathcal{F}^t$ from the adversary (that is observations l^t and $\{d_k^t\}$ or just l_a^t), the convex cost function $c^t : \mathcal{P} \to \mathbb{R}$ is specified; and the learner thus finds out the cost incurred, namely $c^t(\mathbf{p}^t)$. The goal of the learner is to minimize the so-termed *regret* $R_c(T)$ over T rounds, given by

$$R_c(T) := \sum_{t=1}^{T} c^t(\mathbf{p}^t) - \min_{\mathbf{p}\in\mathcal{P}} \sum_{t=1}^{T} c^t(\mathbf{p}) \tag{7}$$

which represents how well the learner performed compared to a single best action that can be chosen with the advantage of knowing all c^t, $t = 1, 2, \ldots, T$, as in hindsight.

Again, depending on the richness of feedback $\{\mathcal{F}^t\}$, either the full information or the bandit case emerges. The full information corresponds to revealing the entire function c^t in each round t. The bandit case refers to revealing only the value of c^t evaluated at a certain query point $\mathbf{p}^t$, that is, $c^t(\mathbf{p}^t)$. Under appropriate conditions, online algorithms can be constructed for both cases, to yield $\{\mathbf{p}^t\}$ with regret upper-bounds that grow sublinearly in T [27]. Thus, as T increases, the algorithms perform at least as good as the fixed action chosen in hindsight (in the sense that $R_c(T)/T$ tends non-positive).

B. Dealing with Composite Objectives

Consider the general form of per-time cost functions c^t that are *composite* and consist of two parts as

$$c^t(\mathbf{p}) = \phi^t(\mathbf{p}) + r(\mathbf{p}) \tag{8}$$

where $\phi^t : \mathcal{P} \to \mathbb{R}$ is a convex function related to the actual datum associated with the t-th round, while $r : \mathcal{P} \to \mathbb{R}$ is a convex regularization function, which encodes application-specific prior knowledge, such as sparsity.

Many efficient OCO algorithms entail first-order iterations that use the (sub)gradient of c^t to produce the next iterate $\mathbf{p}^{t+1}$ [36]. However, using the (sub)gradient of r might not successfully induce in $\mathbf{p}^{t+1}$ the properties intended by adopting the regularizer r. For instance, using a subgradient of an ℓ_1-norm-based regularizer does not necessarily yield sparse $\mathbf{p}^{t+1}$ for intermediate t, although a sublinear regret bound may still be achieved, yielding sparse iterates asymptotically.

A remedy is to incorporate the regularizer explicitly without resorting to (sub)gradients. In the full information case, a number of algorithms implementing this idea have been reported [29], [30]. In Sec. IV-B, a novel OCO algorithm that respects sparsity for quadratic ϕ^t is derived for the bandit case.

IV. Real-Time Price Setting

A. Full Information Case

For the full information case, where l^t and $\{d_k^t\}_{k=1}^K$ are available, the so-termed composite objective mirror descent (COMID) algorithm of [29] will be adapted. To this end, define first

$$\phi^t(\mathbf{p}) := \frac{1}{2}\left(l^t - {\boldsymbol{\theta}^t}^\top \mathbf{p} - m^t\right)^2 \tag{9}$$

$$r(\mathbf{p}) := \lambda||\mathbf{p}||_1 + \frac{\mu}{2}||\mathbf{p}||_2^2 \; . \tag{10}$$

Also, introduce the Bregman divergence $D_\psi(\mathbf{p}, \mathbf{p}')$ associated with a strongly convex function $\psi(\mathbf{p})$ as

$$D_\psi(\mathbf{p}, \mathbf{p}') := \psi(\mathbf{p}) - \psi(\mathbf{p}') - \langle \nabla\psi(\mathbf{p}'), \mathbf{p} - \mathbf{p}' \rangle \tag{11}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. Then, the COMID update sets $\mathbf{p}^{t+1}$ at time $t+1$ as (η is a weighting factor)

$$\mathbf{p}^{t+1} = \arg\min_{\mathbf{p}\in\mathcal{P}} \; \eta\langle \nabla\phi^t(\mathbf{p}^t), \mathbf{p}\rangle + D_\psi(\mathbf{p}, \mathbf{p}^t) + \eta r(\mathbf{p}). \tag{12}$$

Choosing $\psi(\mathbf{p}) = \frac{1}{2}||\mathbf{p}||_2^2$ for simplicity of the resulting optimization, and plugging (9)–(10) into (12) yields

$$\mathbf{p}^{t+1} = \arg\min_{\mathbf{p}\in\mathcal{P}} \Big[-\eta(l^t - {\boldsymbol{\theta}^t}^\top \mathbf{p}^t - m^t){\boldsymbol{\theta}^t}^\top \mathbf{p} + \frac{1}{2}||\mathbf{p} - \mathbf{p}^t||_2^2 + \eta\left(\lambda||\mathbf{p}||_1 + \frac{\mu}{2}||\mathbf{p}||_2^2\right)\Big] . \tag{13}$$

If $\mathcal{P}$ is given by a Cartesian product of intervals, i.e., $\mathcal{P} := \prod_{k=1}^K [\underline{P}_k, \overline{P}_k]$, the update in (13) can be written in closed form per consumer k as

$$p_k^{t+1} = \left[\frac{\eta}{\eta\mu + 1}\text{soft\_th}_\lambda\left(\frac{p_k^t}{\eta} + (l^t - {\boldsymbol{\theta}^t}^\top \mathbf{p}^t - m^t)\theta_k^t\right)\right]_{\underline{P}_k}^{\overline{P}_k} \tag{14}$$

where $[\cdot]_a^b := \min\{\max\{\cdot, a\}, b\}$, and $\text{soft\_th}_\lambda(\cdot)$ is a soft-thresholding function defined as

$$\text{soft\_th}_\lambda(x) := \text{sgn}(x)\max\{0, |x| - \lambda\}. \tag{15}$$

It can be seen that the computational complexity of the per-step update is extremely low. The constants needed, namely, η, λ and μ are preselected, e.g., via cross-validation.

To assess performance of the pricing scheme in (14), the following regret bound can be derived based on the results in [29].

Proposition 1: *Suppose* $||\boldsymbol{\theta}^t||_2 \leq \theta_{\max}$ *for all* t *and* $\mathbf{p}^1 = \mathbf{0}$. *Let* $\mathbf{p}^* := \arg\min_{\mathbf{p}\in\mathcal{P}} \sum_{t=1}^T c^t(\mathbf{p})$ *denote the offline "hindsight" optimum. Then, with* $\eta \propto 1/\sqrt{T}$, *the update in* (13) *yields a sublinear regret (cf.* (7)*)*

$$R_c(T) = O(\theta_{\max}^2 \sqrt{T}||\mathbf{p}^*||_2^2). \tag{16}$$

Proof: The proof is a straightforward application of Corollary 5 in [29], which essentially states that if $\|\nabla\phi^t(\mathbf{p})\|_2^2 \leq \kappa\phi^t(\mathbf{p})$, $r(\mathbf{p}) \geq 0$, and $r(\mathbf{p}^1) = 0$, then setting $\eta \propto 1/\sqrt{T}$ yields $R_c(T) = O(\kappa\sqrt{T}D_\psi(\mathbf{p}^*, \mathbf{p}^1))$. Now, for ϕ^t and r defined as in (9)–(10), it can be easily seen that $\kappa = 2\theta_{\max}^2$ and $\mathbf{p}^1 = \mathbf{0}$ satisfy the conditions, leading to the desired result. ■

B. Partial Information (Bandit) Case

Since the values of $\boldsymbol{\theta}^t$ and l^t are not explicitly observed in the bandit case, $\nabla\phi^t$ necessary in (12) is not available. Since $\{\phi^t\}$ change over time in general, no more than one function evaluation can be performed per time t. Thus, finite difference approximations of gradients, which require multiple function evaluations at different nearby points, cannot be employed naturally for online optimization.

An observation instrumental in this context is that an unbiased estimate of a gradient can still be constructed based on a single evaluation of ϕ^t. For instance, an estimate of the gradient at a point $\mathbf{p}$ of a smoothed version of c^t was obtained as $(K/\delta)c^t(\mathbf{p}+\delta\mathbf{u})\mathbf{u}$ in [37], where $\delta\mathbf{u}\in\mathbb{R}^K$ is a random perturbation vector chosen uniformly from a sphere of radius δ. Then, it was shown that employing this estimate in an online gradient descent algorithm could provide a sublinear regret bound. However, this technique cannot be applied here directly since it does not bring forth the desired sparsity in $\mathbf{p}^t$, as the algorithm does not account for composite objectives, and the perturbation vector used is not sparse.

Our approach is to recognize the composite objectives by developing a COMID-like algorithm in the bandit setting, which does not require a subgradient of r. Also, only a single entry of $\mathbf{p}$ per time is perturbed so that sparsity can be preserved [38]. It can be shown that an unbiased estimate of $\nabla\phi^t$ is still obtained for *quadratic* ϕ^t, which is the case in our problem setup.

Specifically, consider in general a linear-quadratic function $\phi^t(\mathbf{p}) = \mathbf{p}^\top\mathbf{A}^t\mathbf{p}+{\mathbf{b}^t}^\top\mathbf{p}$. To obtain a perturbation of $\mathbf{p}$, which is the point to evaluate the gradient of ϕ^t at, randomly generate integers $k^t\in\{1,2,\ldots,K\}$ and $\epsilon^t\in\{-1,1\}$ from uniform distributions. Then, upon introducing $\delta>0$ and $\mathbf{e}_k$, which represents the k-th canonical vector in $\mathbb{R}^K$, the perturbed vector $\tilde{\mathbf{p}}^t$ is given by

$$\tilde{\mathbf{p}}^t := \mathbf{p}^t + \delta\epsilon^t\mathbf{e}_{k^t}. \tag{17}$$

An estimate of the gradient is then formed from $\phi^t(\tilde{\mathbf{p}}^t)$ as

$$\mathbf{g}^t := \frac{K\epsilon^t}{\delta}\phi^t(\tilde{\mathbf{p}}^t)\mathbf{e}_{k^t}. \tag{18}$$

It is easy to verify that $\mathbf{g}^t$ is an unbiased estimate of $\nabla\phi^t(\mathbf{p}^t)$.

Lemma 1: $\mathbb{E}[\mathbf{g}^t|\mathbf{p}^t] = \nabla\phi^t(\mathbf{p}^t) = (\mathbf{A}^t+{\mathbf{A}^t}^\top)\mathbf{p}^t+\mathbf{b}^t$

Proof: Substituting (17) into (18), it holds that

$$\begin{aligned}&\mathbb{E}[\mathbf{g}^t|\mathbf{p}^t,k^t]\\ =&\frac{K{\mathbf{e}_k}^t}{2\delta}\Big[(\mathbf{p}^t+\delta\mathbf{e}_{k^t})^\top\mathbf{A}^t(\mathbf{p}^t+\delta\mathbf{e}_{k^t})+{\mathbf{b}^t}^\top(\mathbf{p}^t+\delta\mathbf{e}_{k^t})\\ &-(\mathbf{p}^t-\delta\mathbf{e}_{k^t})^\top\mathbf{A}^t(\mathbf{p}^t-\delta\mathbf{e}_{k^t})-{\mathbf{b}^t}^\top(\mathbf{p}^t-\delta\mathbf{e}_{k^t})\Big]\\ =&K\mathbf{e}_{k^t}\mathbf{e}_{k^t}^\top\left[(\mathbf{A}^t+{\mathbf{A}^t}^\top)\mathbf{p}^t+\mathbf{b}^t\right].\end{aligned} \tag{19}$$

Since $\mathbb{E}[\mathbf{e}_{k^t}\mathbf{e}_{k^t}^\top]=\frac{1}{K}\mathbf{I}$, one obtains

$$\mathbb{E}[\mathbf{g}^t|\mathbf{p}^t] = \mathbb{E}[\mathbb{E}[\mathbf{g}^t|\mathbf{p}^t,k^t]|\mathbf{p}^t] = (\mathbf{A}^t+{\mathbf{A}^t}^\top)\mathbf{p}^t+\mathbf{b}^t \tag{20}$$

as desired. ■

TABLE I

PROPOSED ALGORITHM FOR THE BANDIT CASE.

1: Set $\mathbf{p}^1=\mathbf{0}$
2: At each $t=1,2,\ldots,T$
3: Select $k^t\in\{1,2,\ldots,K\}$ and $\epsilon^t\in\{-1,1\}$ randomly
4: Set $\tilde{\mathbf{p}}^t=\mathbf{p}^t+\delta\epsilon^t\mathbf{e}_{k^t}$
5: Play $\tilde{\mathbf{p}}^t$ and observe cost $\phi^t(\tilde{\mathbf{p}}^t)$
6: $\mathbf{p}^{t+1}=\arg\min_{\mathbf{p}\in(1-\alpha)\mathcal{P}}\Big[\eta\langle\frac{K\epsilon^t}{\delta}\phi^t(\tilde{\mathbf{p}}^t)\mathbf{e}_{k^t},\mathbf{p}\rangle$ $+\frac{1}{2}\lvert\lvert\mathbf{p}-\mathbf{p}^t\rvert\rvert_2^2+\eta\left(\lambda\lvert\lvert\mathbf{p}\rvert\rvert_1+\frac{\mu}{2}\lvert\lvert\mathbf{p}\rvert\rvert_2^2\right)\Big]$
7: Next t

Now, define

$$(1-\alpha)\mathcal{P} := \{(1-\alpha)\mathbf{p} : \mathbf{p}\in\mathcal{P}\},\quad 0\le\alpha<1 \tag{21}$$

$$\rho\mathbb{B} := \{\mathbf{p}\in\mathbb{R}^K : ||\mathbf{p}||_2\le\rho\}. \tag{22}$$

Then, the algorithm in Table I is proposed. It can be seen that the algorithm uses $\mathbf{g}^t$ instead of $\nabla\phi^t(\mathbf{p}^t) = -(l^t-{\boldsymbol{\theta}^t}^\top - m^t\mathbf{p}^t)\boldsymbol{\theta}^t$, and replaces $\mathcal{P}$ with $(1-\alpha)\mathcal{P}$ in the update in (13). The latter is necessary to make sure that $\tilde{\mathbf{p}}^t$ stays in $\mathcal{P}$. Due to the COMID-like update in line 6, $\mathbf{p}^t$ is sparse with the level of sparsity controlled by λ, and consequently $\tilde{\mathbf{p}}^t$ is also sparse (since it has at most one more nonzero entry than $\mathbf{p}^t$).

The regret bound for this algorithm is assessed in the following proposition.

Proposition 2: *Let $\mathcal{P}\subset\mathbb{R}^K$ be a compact and convex set with $\rho_{\text{in}}\mathbb{B}\subset\mathcal{P}\subset\rho_{\text{out}}\mathbb{B}$. Suppose that $\phi^t:\mathcal{P}\to\mathbb{R}$ is quadratic and convex, $r:\mathcal{P}\to\mathbb{R}$ convex, and $|\phi^t(\mathbf{p})|, |\phi^t(\mathbf{p})+r(\mathbf{p})|\le C$ for all t and $\mathbf{p}\in\mathcal{P}$. Function ϕ^t may be chosen adaptively such that ϕ^t is dependent on previous query points $\tilde{\mathbf{p}}^1,\ldots,\tilde{\mathbf{p}}^t$. Since $\mathcal{P}$ is compact, c^t is Lipschitz continuous with Lipschitz constant L. Then, the algorithm in Table I with*

$$\delta = \sqrt{\frac{\rho_{\text{out}}CK}{L+2C/\rho_{\text{in}}}}T^{-\frac{1}{4}} \tag{23}$$

$$\alpha = \delta/\rho_{\text{in}} \tag{24}$$

$$\eta = \sqrt{\frac{{\rho_{\text{out}}}^3}{CK(L+2C/\rho_{\text{in}})}}\;T^{-\frac{3}{4}} \tag{25}$$

produces iterates $\{\tilde{\mathbf{p}}^t\}$ that satisfy

$$\mathbb{E}\left[\sum_{t=1}^T\left(\phi^t(\tilde{\mathbf{p}}^t)+r(\tilde{\mathbf{p}}^t)\right)\right]-\min_{\mathbf{p}\in\mathcal{P}}\mathbb{E}\left[\sum_{t=1}^T\left(\phi^t(\mathbf{p})+r(\mathbf{p})\right)\right] = O\left(\rho_{\text{out}}CK\left(L+2C/\rho_{\text{in}}\right)T^{\frac{3}{4}}\right). \tag{26}$$

Proof: See Appendix.

Note that conditions $|\phi^t(\mathbf{p})|, |\phi^t(\mathbf{p})+r(\mathbf{p})|\le C$ are satisfied for (9)–(10) when $\mathcal{P}$ is compact. Lipschitz continuity follows either from compactness of $\mathcal{P}$, or from boundedness of l^t, $\boldsymbol{\theta}^t$ and m^t, both of which are readily satisfied in practice. Proposition 2 asserts that even with bandit feedback, a sublinear regret can be achieved, although the speed of convergence deteriorates to $O(T^{3/4})$ from $O(\sqrt{T})$ of the full information counterpart.

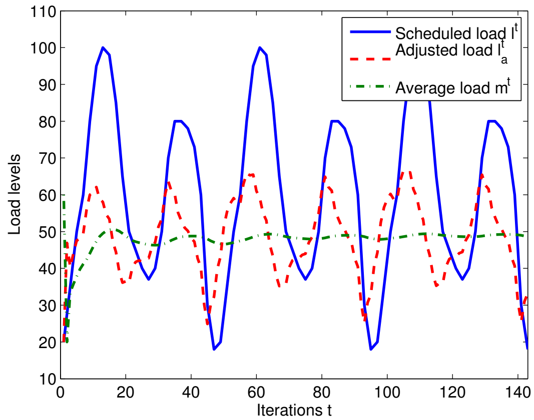

Fig. 2. Load curves before and after real-time DR.

Remark 3. In the case of *oblivious* adversary, where a sequence of functions $\{\phi^t\}$ is fixed *prior* to time $t = 1$, thus without the knowledge of $\{\tilde{\mathbf{p}}^t\}$, it can be shown that a regret bound similar to (26) holds [32]; i.e.,

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(\phi^t(\tilde{\mathbf{p}}^t) + r(\tilde{\mathbf{p}}^t)\right)\right] - \min_{\mathbf{p}\in\mathcal{P}}\sum_{t=1}^{T}\left(\phi^t(\mathbf{p}) + r(\mathbf{p})\right) = O\left(\rho_{\text{out}} CK\left(L + 2C/\rho_{\text{in}}\right)T^{\frac{3}{4}}\right). \tag{27}$$

Remark 4. As the proposed pricing schemes do not explicitly incorporate incentive compatibility as its design goals, it is possible that the consumers may find incentives to act untruthfully [39]. For example, they may pretend that their θ_k^t are small, and later when the demand is high, increase their load to benefit from the low price. This aspect needs further investigation, but a number of comments are in order. First, unlike auction scenarios, where there are negotiation stages through bidding, consumers in our setup must actually suffer from the penalty of consuming more at high prices and less at low prices, in order to forge low elasticity. Moreover, the proposed algorithms can track changes in elasticity, preventing consumers from benefiting from untruthful actions for a long time. Finally, it is worth emphasizing that the OCO framework is inherently robust in the sense that the regret guarantees are unaltered by strategic actions of the consumers.

V. Numerical Tests

To verify the effectiveness of the proposed pricing strategies, numerical tests were performed. A full information case involving $K = 100$ consumers with constant elasticity parameters $\theta_k^t = \theta_k$ is considered first, where θ_k for $k = 1, \ldots, 80$ were picked uniformly at random from $[0, 0.5]$, while $\theta_k = 0$ for $k = 81, \ldots, 100$. A maximum price deviation of $\overline{P}_k = -\underline{P}_k = 5$ was used. The update was performed every half an hour over a three-day horizon. The values of λ and μ were set to 0.1 and 0.5, respectively.

In Fig. 2, the solid curve depicts the base aggregate load l^t, and the dashed one corresponds to the overall adjusted load l_a^t, which reflects the price-induced load adaptation. The dash-dotted curve represents the mean level m^t. It is observed that the proposed pricing scheme significantly reduces load variations.

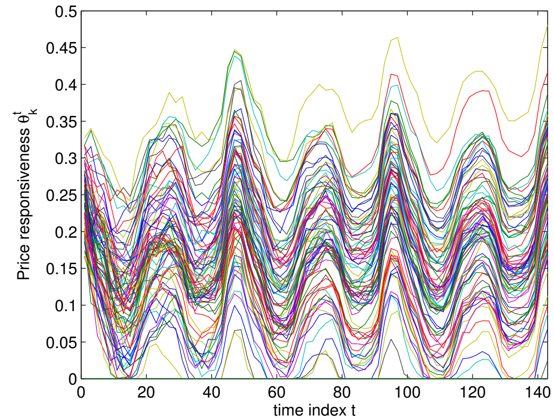

Fig. 3. Time-varying price responsiveness for $K = 100$ customers.

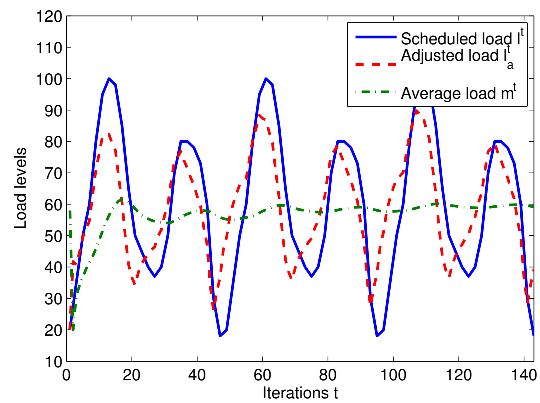

Fig. 4. Load curves for time-varying elasticity.

Another full-information case employing time-varying $\boldsymbol{\theta}^t$ was considered, where the employed values of $\{\theta_k^t\}$ are shown in Fig. 3. This case is challenging as the consumers tend to have low responsiveness during the peak periods. Before the morning commute for example, consumers would not defer running certain electric appliances such as the water heaters or electric stoves. The corresponding load curves in Fig. 4 still ex-

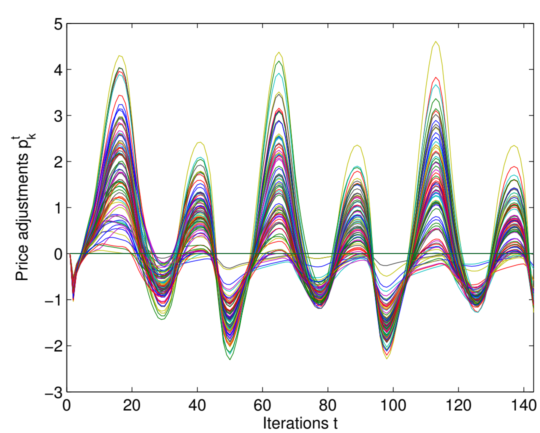

Fig. 5. Price adjustments in the full information case.

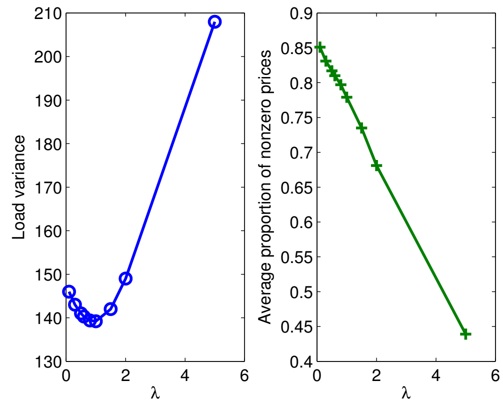

Fig. 6. Effect of λ on load variance and proportion of nonzero price adjustments.

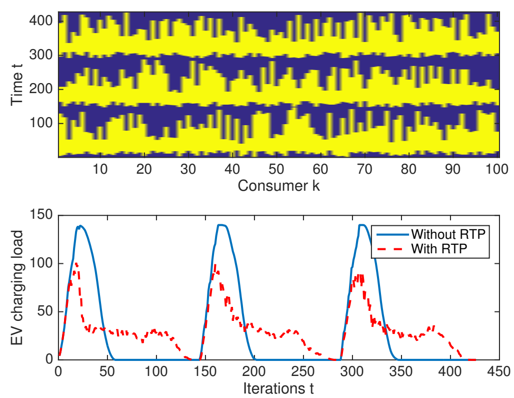

Fig. 7. The case of EV charging customers. (Top) Requested EV charging start and end times. (Bottom) EV charging load with and without the proposed pricing scheme.

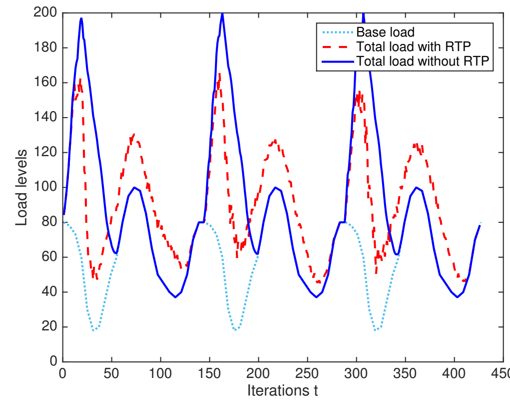

Fig. 8. Overall (base plus EV charging) load levels.

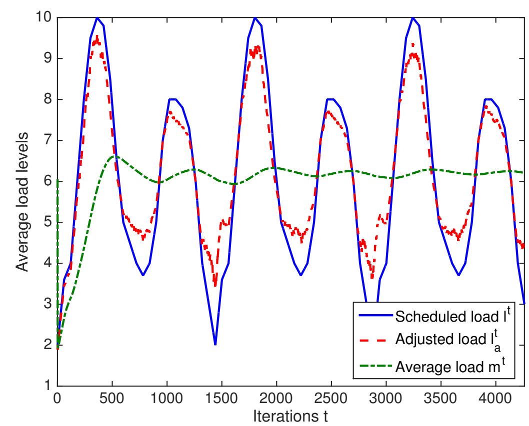

Fig. 9. Load curves in the bandit case averaged over 100 trials.

hibit sizable improvement over the base load, although the DR performance is worse than that of the case with constant $\boldsymbol{\theta}^t$ due to tracking errors. Fig. 5 depicts the prices p_k^t generated from the proposed algorithm. It is seen that consumers with higher elasticity are given larger incentives (price amendments) to adjust their loads. However, the amount of price deviations can be controlled by the "fairness" parameter μ.

To verify the effect of sparsity-promoting regularization, which aims at capturing the inherent sparsity in consumer elasticity, the left panel in Fig. 6 depicts the load variance given by $T^{-1}\sum_{t=1}^{T}\phi^t(\mathbf{p}^t)$ for different values of λ. In the right panel, the proportion of nonzero price adjustments given by $T^{-1}\sum_{t=1}^{T}|\{k|p_k^t \neq 0\}|/K$ versus λ is plotted. It can be seen that $\lambda \approx 1$ achieves the smallest load variance. This value corresponds to making price adjustments for roughly 80% of the consumers in the pool, which matches well with the actual proportion of consumers with nonzero elasticity.

To further verify the efficacy of the proposed algorithm, EV charging is also considered, where the energy demand is typically shifted in time, rather than being compressed. We modeled $K = 100$ customers charging their EVs daily over 3 days, starting approximately at 6-10pm with allowed charging delay uniformly varying over 6 to 21 hours. The allowed EV charging intervals $\mathcal{I}_k$ for consumers $k = 1, 2, \ldots, K$, are depicted in the top panel of Fig. 7 as the yellow region. The energy required for full charge was uniformly distributed over $[6, 10]$ and the charging rate was set to 1.4 per hour. Let E_k^t be the amount of energy still to be charged at time t for consumer k, and D_k^t the time until the deadline. Then, the charging decision $\delta_k^t \in \{1 \text{ (charge)}, 0 \text{ (not charge)}\}$ was modeled as $\delta_k^t = \mathbb{I}\{t \in \mathcal{I}_k \text{ and } \zeta_k E_k^t/(D_k^t + 10^{-3}) + \eta_k^t > p_k^t\}$, where $\mathbb{I}\{\cdot\}$ is the indicator function, yielding 1 if condition $\{\cdot\}$ is satisfied, and 0, otherwise; ζ_k models the priority of charging early for consumer k; and η_k^t captures the randomness in decisions. In the simulation, ζ_k was assumed uniform over $[25, 26]$, and η_k zero-mean Gaussian with variance 0.25. Note the load characteristic is discontinuous, since the demand switches to the charging state abruptly. The values of θ_k^t was estimated by dividing the EV charging load of each consumer by $p_k^t + 0.1$. The iterations were done every 10 minutes.

The bottom panel of Fig. 7 shows that the proposed pricing scheme can effectively spread the peaky EV charging load across time. Note that 99.97% of the EV charging demand was still satisfied under the RTP. Fig. 8 shows the overall load, which is the inelastic base load plus the EV charging load. It can be seen that under the proposed pricing, the overall load curve became considerably smoother.

Fig. 9 presents the load curves for the bandit setting averaged over 100 independent trials. A case of $K = 10$ consumers with time-varying elasticity was considered. The values of λ and μ were set to 1 and 0, respectively. The updates were performed every 60 seconds, reflecting the fact that the bandit case takes much longer time to track compared to the full information case. However, the bandit feedback has a privacy advantage and requires markedly reduced communication overhead; cf. Remark 2. Since, in practice, real-time pricing is overlaid on top of day-ahead offline scheduling, severe dynamics used in the present tests may be unusual. Also, incorporation of appropriate dynamic models might mitigate this issue, which will be explored in our future work.

VI. Conclusion

Algorithms for real-time electricity pricing were developed for DR in smart grids. Price responsiveness of individual consumers and loads is taken into account to set optimal prices for inducing desired power consumption behaviors to curb load variations. Structural constraints such as fairness and sparsity of prices have also been incorporated. The proposed strategies provide performance guarantees based on an OCO framework with minimal modeling assumptions on the dynamics of load levels and consumer preferences, capable of coping even with adversarial (strategic) actions of consumers. Pricing strategies for both full and limited information setups have been proposed. Numerical tests demonstrated that the novel approach could reduce load variation significantly through implicitly learning load elasticity in an online fashion.

Appendix

In order to prove Proposition 2, analogous to the proof technique in [37], we need the following lemma.

Lemma 2: *Let $\phi^1, \phi^2, \ldots, \phi^T : \mathcal{P} \to \mathbb{R}$ be a sequence of differentiable convex functions, where ϕ^t can potentially depend on previous iterates $\mathbf{p}^1, \ldots, \mathbf{p}^t$, and $r : \mathcal{P} \to \mathbb{R}$ a convex function with $r(\mathbf{0}) = 0$. Also, let ψ be a ζ-strongly convex function with respect to a norm $\|\cdot\|$. Consider iterates $\{\mathbf{p}^t\}$ defined by $\mathbf{p}^1 = \mathbf{0}$, and*

$$\mathbf{p}^{t+1} = \arg\min_{\mathbf{p}\in\mathcal{P}} \eta\langle \mathbf{g}^t, \mathbf{p}\rangle + D_\psi(\mathbf{p}, \mathbf{p}^t) + \eta r(\mathbf{p}) \tag{28}$$

where $\mathbf{g}^1, \ldots, \mathbf{g}^T$ are random vectors satisfying

(a) $\mathbb{E}[\mathbf{g}^t | \mathbf{p}^1, \phi^1, \mathbf{p}^2, \phi^2, \ldots, \mathbf{p}^t, \phi^t] = \nabla\phi^t(\mathbf{p}^t)$
(b) $\|\mathbf{g}^t\|_* \le G_*$ *for some* $G_* > 0$[3]

Then, for $c^t := \phi^t + r$, $t = 1, 2, \ldots, T$, and any $\mathbf{p}^* \in \mathcal{P}$, it holds that

$$\mathbb{E}\left[\sum_{t=1}^T c^t(\mathbf{p}^t)\right] - \mathbb{E}\left[\sum_{t=1}^T c^t(\mathbf{p}^*)\right] \le \sqrt{2D_\psi(\mathbf{p}^*, \mathbf{0})G_*^2/\zeta}\sqrt{T}. \tag{29}$$

Proof: Define $h^t : \mathcal{P} \to \mathbb{R}$ as

$$h^t(\mathbf{p}) := \phi^t(\mathbf{p}) + \langle \mathbf{p}, \mathbf{g}^t - \nabla\phi^t(\mathbf{p}^t)\rangle. \tag{30}$$

[3]Note that $\|\cdot\|_*$ denotes the dual norm of $\|\cdot\|$. It is defined as $\|\mathbf{v}\|_* := \sup\{\langle \mathbf{u}, \mathbf{v}\rangle | \|\mathbf{u}\| \le 1\}$.

Then, it can be seen that $\nabla h^t(\mathbf{p}^t) = \mathbf{g}^t$. Consider applying (12) for $\{h^t\}$ instead of $\{\phi^t\}$. Then, from [29, Theorem 2], one obtains

$$\begin{aligned}\sum_{t=1}^T &\left(h^t(\mathbf{p}^t) + r(\mathbf{p}^t)\right) - \sum_{t=1}^T \left(h^t(\mathbf{p}^*) + r(\mathbf{p}^*)\right)\\ &\le \frac{1}{\eta} D_\psi(\mathbf{p}^*, \mathbf{p}^1) + r(\mathbf{p}^1) + \frac{\eta G_*^2 T}{2\zeta}\\ &\le \sqrt{2D_\psi(\mathbf{p}^*, \mathbf{p}^1)G_*^2 T/\zeta}\end{aligned} \tag{31}$$

where the last inequality follows by noting that $r(\mathbf{p}^1) = 0$, and choosing

$$\eta = \sqrt{\frac{2\zeta D_\psi(\mathbf{p}^*, \mathbf{p}^1)}{G_*^2 T}}. \tag{32}$$

Now, note that $\mathbb{E}[h^t(\mathbf{p}^t)] = \mathbb{E}[\phi^t(\mathbf{p}^t)]$ and $\mathbb{E}[h^t(\mathbf{p}^*)] = \mathbb{E}[\phi^t(\mathbf{p}^*)]$. Thus, taking expectations on (31) yields the desired result. ■

Proof of Proposition 2: First, it is noted that $\tilde{\mathbf{p}}^t \in \mathcal{P}$, provided that

$$\frac{\delta}{\rho_{\text{in}}} \le \alpha < 1. \tag{33}$$

This is because a ball of radius $\alpha\rho_{\text{in}}$ centered at any point in $(1-\alpha)\mathcal{P}$ is contained in $\mathcal{P}$ (cf. [37, Observation 3.2]). Also, for $\mathbf{g}^t$ defined in (18), assumption (a) in Lemma 2 holds as shown in Lemma 1. Moreover,

$$\begin{aligned}\|\mathbf{g}^t\|_* &= K\delta^{-1}|\phi^t(\tilde{\mathbf{p}}^t)|\|\mathbf{e}_{k^t}\|_*\\ &\le K\delta^{-1}C \max_{1\le k\le K} \|\mathbf{e}_k\|_* := G_*\end{aligned} \tag{34}$$

which verifies that assumption (b) is satisfied as well.

Consider now the iterates $\{\mathbf{p}^t\}$ generated by line 6 of Table I, which is equivalent to (28) with $\mathcal{P}$ replaced by $(1-\alpha)\mathcal{P}$ and $\psi(\mathbf{p}) = \frac{1}{2}\|\mathbf{p}\|_2^2$. Note that $\zeta = 1$ for this choice of ψ. Since $D_\psi(\mathbf{p}^*, \mathbf{0}) \le \frac{1}{2}{\rho_{\text{out}}}^2$, the following holds from Lemma 2 for any $\mathbf{p}^* \in \mathcal{P}$

$$\mathbb{E}\left[\sum_{t=1}^T c^t(\mathbf{p}^t)\right] - \mathbb{E}\left[\sum_{t=1}^T c^t((1-\alpha)\mathbf{p}^*)\right] \le \rho_{\text{out}} G_* \sqrt{T}. \tag{35}$$

To relate (35), which is in terms of $\mathbf{p}^t$ and $c^t((1-\alpha)\mathbf{p}^*)$, to the desired bound (26) in $\tilde{\mathbf{p}}^t$ and $c^t(\mathbf{p}^*)$, note first that $c^t(\mathbf{p}^t)$ and $c^t(\tilde{\mathbf{p}}^t)$ are close in the sense that

$$|c^t(\mathbf{p}^t) - c^t(\tilde{\mathbf{p}}^t)| \le \delta L \tag{36}$$

since $\|\mathbf{p}^t - \tilde{\mathbf{p}}^t\| \le \delta$ and c^t is L-Lipschitz. Summing this up over t, one obtains

$$\sum_{t=1}^T c^t(\tilde{\mathbf{p}}^t) - \sum_{t=1}^T c^t(\mathbf{p}^t) \le \delta L T. \tag{37}$$

Furthermore, from the boundedness of $|c^t(\mathbf{p})|$, it can be verified that the minimum of c^t in $(1-\alpha)\mathcal{P}$ is close to the minimum in $\mathcal{P}$. Specifically, it holds that (cf. [37, Observation 3.1])

$$\sum_{t=1}^{T} c^t((1-\alpha)\mathbf{p}^*) \leq \sum_{t=1}^{T} c^t(\mathbf{p}^*) + 2\alpha CT \tag{38}$$

for any $\mathbf{p}^* \in \mathcal{P}$. Combining (35), (37) and (38), and taking the minimum over $\mathbf{p}^* \in \mathcal{P}$, yields

$$\mathbb{E}\left[\sum_{t=1}^{T} c^t(\tilde{\mathbf{p}}^t)\right] - \min_{\mathbf{p}^* \in \mathcal{P}} \mathbb{E}\left[\sum_{t=1}^{T} c^t(\mathbf{p}^*)\right] \leq \rho_{\text{out}} K\delta^{-1} C\sqrt{T} + \delta LT + 2\alpha LT. \tag{39}$$

Finally, choosing δ, α and η as in (23)–(25) leads to the desired bound. ■

References

[1] S. M. Amin and B. F. Wollenberg, "Toward a smart grid," *IEEE Power & Energy Mag.*, vol. 3, no. 5, pp. 34–41, Sep./Oct. 2005.

[2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid – the new and improved power grid: A survey," *IEEE Comm. Survey & Tutorials*, vol. 14, no. 4, pp. 944–980, Fourth Quarter 2012.

[3] G. B. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Sig. Proc. Mag.*, vol. 30, no. 5, pp. 107–128, Sep. 2013.

[4] K. Hamilton and N. Gulhar, "Taking demand response to the next level," *IEEE Power & Energy Mag.*, vol. 8, no. 3, pp. 61–65, May/Jun. 2010.

[5] D. Wu, D. C. Aliprantis, and L. Ying, "Load scheduling and dispatch for aggregators of plug-in electric vehicles," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 368–376, Mar. 2012.

[6] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 120–133, Sep. 2010.

[7] T. T. Kim and H. V. Poor, "Scheduling power consumption with price uncertainty," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 519–527, Sep. 2011.

[8] D. Ban, G. Michailidis, and M. Devetsikiotis, "Demand response control for PHEV charging stations by dynamic price adjustments," in *Proc. of the IEEE PES ISGT Conf.*, Washington, D.C., Jun. 2012.

[9] L. Chen, N. Li, L. Jiang, and S. H. Low, "Optimal demand response: Problem formulation and deterministic case," in *Control and Optimization Methods for Electric Smart Grids*, A. Chakrabortty and M. D. Ilić, Eds. New York, NY: Springer, 2012, pp. 63–85.

[10] L. Gan, U. Topcu, and S. Low, "Optimal decentralized protocol for electric vehicle charging," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 940–951, May 2013.

[11] L. P. Qian, Y. J. Zhang, J. Huang, and Y. Wu, "Demand response management via real-time electricity price control in smart grids," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1268–1280, Jul. 2013.

[12] Q. Li, T. Cui, R. Negi, F. Franchetti, and M. D. Illić, "On-line decentralized charging of plug-in electric vehicles in power systems," Nov. 2011. [Online]. Available: http://arxiv.org/abs/arXiv:1106.5063

[13] E. Sortomme, M. M. Hindi, S. D. J. MacPherson, and S. S. Venkata, "Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 198–205, Mar. 2011.

[14] W. Saad, Z. Han, H. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, smart grid communications," *IEEE Sig. Proc. Mag.*, vol. 29, no. 5, pp. 86–105, Sep. 2012.

[15] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: Pricing methods and optimization algorithms," *IEEE Comm. Survey & Tutorials*, vol. 17, no. 1, pp. 152–178, First Quarter 2015.

[16] A.-H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[17] P. Yang, G. Tang, and A. Nehorai, "A game-theoretic approach for optimal time-of-use electricity pricing," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 884–892, May 2013.

[18] W. Tushar, W. Saad, H. V. Poor, and D. Smith, "Economics of electric vehicle charging: A game theoretic approach," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1767–1779, Dec. 2012.

[19] Y. Wang, W. Saad, Z. Han, H. V. Poor, and T. Başar, "A game-theoretic approach to energy trading in the smart grid," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1439–1450, May 2014.

[20] V. Gomez, M. Chertkov, S. Backhaus, and H. Kappen, "Learning price-elasticity of smart consumers in power distribution systems," in *Proc. of IEEE SmartGridComm*, Tainan City, Taiwan, Nov. 2012.

[21] O. Corradi, H. Oschsenfeld, H. Madsen, and P. Pinson, "Controlling electricity consumption by forecasting its response to varying prices," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 421–429, Feb. 2013.

[22] M. Alizadeh, Y. Xiao, A. Scaglione, and M. van der Schaar, "Incentive design for direct load control programs," in *Proc. of the 51st Ann. Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Oct. 2013.

[23] P. T. Baboli, M. Eghbal, M. P. Moghaddam, and H. Aalami, "Customer behavior based demand response model," in *Proc. IEEE PES General Meeting*, San Diego, CA, Jul. 2012, pp. 1–7.

[24] N. Y. Soltani, S.-J. Kim, and G. B. Giannakis, "Online learning of load elasticity for electric vehicle charging," in *Proc. of the 5th CAMSAP Conf.*, St. Martin Island, French Caribbean, Dec. 2013.

[25] W. Wang, Y. Xu, and M. Khanna, "A survey on the communication architectures in smart grid," *Comput. Net.*, vol. 55, no. 15, pp. 3604–3629, Oct. 2011.

[26] OpenHAN Task Force, *UCAIug Home Area Network System Requirements Specification*, Ver. 2.0, UCA Int'l. Users Group, Aug. 2010.

[27] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. & Trends in Mach. Learn.*, vol. 4, pp. 107–194, Mar. 2012.

[28] B. Narayanaswamy, T. S. Jayram, and V. N. Yoong, "Hedging strategies for renewable resource integration and uncertainty management in the smart grid," in *Proc. IEEE PES ISGT Europe*, Berlin, Germany, 2012.

[29] J. C. Duchi, S. Shalev-Shwartz, Y. Singer, and A. Tewari, "Composite objective mirror descent," in *Proc. of the 23rd Intl. Conf. on Learning Theory*, Haifa, Israel, Jun. 2010.

[30] L. Xiao, "Dual averaging methods for regularized stochastic learning and online optimization," *J. Mach. Learning Res.*, vol. 11, pp. 2543–2596, Oct. 2010.

[31] H. B. McMahan, "A unified view of regularized dual averaging and mirror descent with implicit updates," Sep. 2011. [Online]. Available: http://arxiv.org/abs/arXiv:1009.3240v2

[32] S.-J. Kim and G. B. Giannakis, "Real-time electricity pricing for demand response using online convex optimization," in *Proc. of the IEEE PES ISGT Conf.*, Washington, D.C., Feb. 2014.

[33] ——, "Scalable and robust demand response with mixed-integer constraints," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2089–2099, Dec. 2013.

[34] Ameren Illinois, "Real Time Prices," https://www2.ameren.com/RetailEnergy/RealTimePrices.aspx.

[35] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[36] N. Srebro, K. Sridharan, and A. Tewari, "On the universality of online mirror descent," in *Advances in Neural Information Processing Systems*, Granada, Spain, Dec. 2011.

[37] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online convex optimization in the bandit setting: gradient descent without a gradient," in *Proc. of the 16th Annual ACM-SIAM Symp. on Discrete Algorithms*, Vancouver, British Columbia, Jan. 2005, pp. 385–394.

[38] J. Abernethy, E. Hazan, and A. Rakhlin, "Competing in the dark: An efficient algorithm for bandit linear optimization," in *Proc. of the 21st Annual Conf. on Learning Theory*, Helsinki, Finland, Jul. 2008.

[39] N. Nisan, *Algorithmic Game Theory*. New York: Cambridge University Press, 2007.