Real-Time Electricity Pricing for Demand Response Using Online Convex Optimization

Seung-Jun Kim and Geogios B. Giannakis
Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455, USA
E-mail: {seungjun,georgios}@umn.edu

***Abstract*—Real-time electricity pricing strategies for demand response in smart grids are proposed. By accounting for individual consumers' responsiveness to prices, adjustments are made so as to induce desirable usage behavior and reduce peaks in load curves. An online convex optimization framework is adopted, which provides performance guarantees with minimal assumptions on the dynamics of load levels and consumer responsiveness. 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 load levels are revealed. Fairness and sparsity constraints are also incorporated. Numerical tests verify the effectiveness of the proposed approach.**

I. Introduction

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 [1]. Such a coordination can be effected either directly by allowing the load serving entities to exert direct control over consumer loads, or indirectly by adjusting electricity prices over time, driving consumers to adapt their loads accordingly. 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. In any case, the idea is to raise the price during the periods of peak usage, and lower it at the valley 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 the 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 [2], [3]. 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.

This work was supported by NSF grant 1202135, and by the Institute of Renewable Energy and the Environment (IREE) at the University of Minnesota, under grant No. RL-0010-13.

Attempts to capture consumer behavioral patterns in the context of smart grids have appeared recently. In [4], individual consumer responsiveness to price was estimated in a linear regression framework, where the shift in the total load was fit to the price changes announced to individual consumers. An algorithm to learn EV charging preferences was developed based on a conditional random field (CRF) model in [5], where spatial dependencies among consumers were also accounted for. The premise 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, the problem of price setting is tackled, where learning consumer price responsiveness is implicit. In particular, our focus is on real-time pricing, where adjustments to the prices announced prior to a scheduling horizon are made to cope with the discrepancy in the actual consumer behaviors and load levels from forecast and planning. Thus, the algorithm must be able to accommodate unmodeled variations in the underlying consumer behavior and load levels. The relevant measurements may be limited due to privacy or technical reasons, 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 under adversarial setups [6], [7]. 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 case, only the total load is revealed corresponding to the price set, levying the additional task of estimating the gradient. Desirable structures in the generated prices, such as fairness and sparsity, are incorporated in the form of appropriate regularizers.

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 time-of-use (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. 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 $\bar{l}^t$ [8]. However, day-ahead scheduling inherently involves many uncertainties that cannot be always predicted with sufficient accuracy.

Once the time horizon of interest emerges, the actual load level l^t is revealed at each time t. In order to reflect real-time supply and demand imbalance, the utilities can make adjustments of the prices through real-time pricing (RTP). Let p_k^t denote the price adjustment for consumer k at time t such that the actual price becomes $\bar{p}_k^t + p_k^t$. Since individual consumers have different constitution of elastic loads and distinct preferences, the load-to-price sensitivities vary. Making a first-order approximation that the load change is linear in the price variation, the increase in consumer k's load demand at time 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 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 the aggregate load $l^t + \sum_{k=1}^K d_k^t$ can be observed at the end of time t. Upon defining $\boldsymbol{\theta}^t := [\theta_1^t, \ldots, \theta_K^t]^\top$ and $\mathbf{p}^t$ likewise, the aggregate load can be equivalently expressed as $l^t - {\boldsymbol{\theta}^t}^\top \mathbf{p}^t$, where $\cdot^\top$ denotes transposition.

An important goal of the utility is to minimize load variations over time by judicious price setting [2], [3]. This can be achieved by minimizing the variance of the aggregate load around prescribed average load levels $\{m^t\}$, given by [9]

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

One issue with adjusting individual customer prices is fairness. It is undesirable to have large adjustments only in few customers' prices. To ensure that the price variation among customers remains small, one can adopt a penalty term discouraging large deviation, such as $\sum_t ||\mathbf{p}^t||_2^2 := \sum_t \sum_{k=1}^K (p_k^t)^2$. 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 [4]. For many consumers, θ_k^t will remain zero, and thus choosing nonzero p_k^t will have no effect in inducing desired load changes. Therefore, it is natural to incorporate this prior knowledge by promoting *sparsity* in $\mathbf{p}^t$ in each time t, which can be achieved by incorporating the penalty term $\sum_t ||\mathbf{p}^t||_1$, where $||\mathbf{p}^t||_1 := \sum_{k=1}^K |p_k^t|$. Upon introducing non-negative weights λ and μ for the sparsity and fairness terms, the overall cost per time is given by

$$c^t(\mathbf{p}^t) := \frac{1}{2}\left(l^t - {\boldsymbol{\theta}^t}^\top \mathbf{p}^t - m^t\right)^2 + \lambda ||\mathbf{p}^t||_1 + \frac{\mu}{2}||\mathbf{p}^t||_2^2. \tag{3}$$

The goal becomes to attain small total cost $\sum_{t=1}^T c^t(\mathbf{p}^t)$ by setting $\{\mathbf{p}^t\}$ appropriately.

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.

Thus, the problem of interest is as follows. At each time t, based on the past prices $\mathbf{p}^\tau$, $\tau = 1, 2, \ldots, t-1$, and the corresponding observations, set prices $\{p_k^t\}_{k=1}^K$ 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 work. In the *full* information case, it is assumed that the utilities can acquire the values of l^τ and $\boldsymbol{\theta}^\tau$ (possibly through observing $\{d_k^\tau\}_{k=1}^K$) explicitly at the end of each time slot τ. In the *partial* information or *bandit* case, only the aggregate load $(l^\tau - {\boldsymbol{\theta}^\tau}^\top \mathbf{p}^\tau)$ can be observed at each time τ.

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 [6]. In our setup, the utility plays the role of the learner, while consumers are the adversaries. In round t, the learner plays an action $\mathbf{p}^t \in \mathcal{P}$, and subsequently the adversary chooses a convex function $c^t : \mathcal{P} \to \mathbb{R}$. Then, the learner suffers cost $c^t(\mathbf{p}^t)$ and receives some feedback information (observations) $\mathcal{F}^t$. The goal of the learner is to minimize the so-termed *regret* $R_c(T)$ over T rounds, given by

$$R_c(T) := \sum_{t=1}^{T} c^t(\mathbf{p}^t) - \min_{\mathbf{p}\in\mathcal{P}} \sum_{t=1}^{T} c^t(\mathbf{p}) \tag{4}$$

which represents how well the learner performed compared to a single best strategy 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 $\mathbf{p}^t$, 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 [6]. 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

It is worth paying a special attention to cases where the per-time objective functions c^t are *composite* and consist of two parts: $c^t = \phi^t + r$, 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 $\mathbf{p}^{t+1}$ [10]. 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 [11], [12]. In Sec. IV-B, an 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, the so-termed composite objective mirror descent (COMID) algorithm of [11] will be adapted. To this end, define first

$$\phi^t(\mathbf{p}) := \frac{1}{2}\left(l^t - \boldsymbol{\theta}^{t\top}\mathbf{p} - m^t\right)^2 \quad (5)$$

$$r(\mathbf{p}) := \lambda||\mathbf{p}^t||_1 + \frac{\mu}{2}||\mathbf{p}^t||_2^2\ . \quad (6)$$

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

$$D_\psi(\mathbf{p},\mathbf{p}') := \psi(\mathbf{p}) - \psi(\mathbf{p}') - \langle\nabla\psi(\mathbf{p}'),\mathbf{p}-\mathbf{p}'\rangle \quad (7)$$

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)

$$\mathbf{p}^{t+1} = \arg\min_{\mathbf{p}\in\mathcal{P}}\ \eta\langle\nabla\phi^t(\mathbf{p}^t),\mathbf{p}\rangle + D_\psi(\mathbf{p},\mathbf{p}^t) + \eta r(\mathbf{p}). \quad (8)$$

Choosing $\psi(\mathbf{p}) = \frac{1}{2}||\mathbf{p}||_2^2$ for simplicity, and plugging (5)–(6) into (8) yields[1]

$$\mathbf{p}^{t+1} = \arg\min_{\mathbf{p}\in\mathcal{P}}\Big[-\eta(l^t - \boldsymbol{\theta}^{t\top}\mathbf{p}^t - m^t)\boldsymbol{\theta}^{t\top}\mathbf{p} + \frac{1}{2}||\mathbf{p}-\mathbf{p}^t||_2^2 + \eta\left(\lambda||\mathbf{p}||_1 + \frac{\mu}{2}||\mathbf{p}||_2^2\right)\Big]. \quad (9)$$

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 (9) can be written in closed form per consumer k as

$$p_k^{t+1} = \left[\frac{\eta}{\eta\mu+1}\text{soft\_th}_\lambda\left(\frac{p_k^t}{\eta} + (l^t - \boldsymbol{\theta}^{t\top}\mathbf{p}^t - m^t)\theta_k^t\right)\right]_{\underline{P}_k}^{\overline{P}_k} \quad (10)$$

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

$$\text{soft\_th}_\lambda(x) := \text{sgn}(x)\max\{0,|x|-\lambda\}. \quad (11)$$

To assess performance of the pricing scheme in (10), the following regret bound can be derived straightforwardly based on the results in [11].

[1] 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.

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: $\mathbf{p}^{t+1} = \arg\min_{\mathbf{p}\in(1-\alpha)\mathcal{P}}\Big[\eta\langle\frac{K\epsilon^t}{\delta}\phi^t(\tilde{\mathbf{p}}^t)\mathbf{e}_{k^t},\mathbf{p}\rangle + \frac{1}{2}\lVert\mathbf{p}-\mathbf{p}^t\rVert_2^2 + \eta\left(\lambda\lVert\mathbf{p}\rVert_1 + \frac{\mu}{2}\lVert\mathbf{p}\rVert_2^2\right)\Big]$
7: Next t

Proposition 1: *Suppose $||\boldsymbol{\theta}^t||_2 \le \theta_{\max}$ 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* (9) *yields a regret*

$$R_c(T) = O(\theta_{\max}^2\sqrt{T}||\mathbf{p}^*||_2^2). \quad (12)$$

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 for mirror descent updates 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 [13], 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 [14]. 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 quadratic function $\phi^t(\mathbf{p}) = \mathbf{p}^\top\mathbf{A}^t\mathbf{p} + \mathbf{b}^{t\top}\mathbf{p}$, where $\mathbf{A}^t \succeq \mathbf{0}$. 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

$$\tilde{\mathbf{p}}^t := \mathbf{p}^t + \delta\epsilon^t\mathbf{e}_{k^t}. \quad (13)$$

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

$$\mathbf{g}^t := \frac{K\epsilon^t}{\delta}\phi^t(\tilde{\mathbf{p}}^t)\mathbf{e}_{k^t}. \quad (14)$$

It is easy to verify that $\mathbb{E}[\mathbf{g}^t|\mathbf{p}^t] = \nabla\phi^t(\mathbf{p}^t) = \mathbf{A}^t\mathbf{p}^t + \mathbf{b}^t$.

Now, define

$$(1-\alpha)\mathcal{P} := \{(1-\alpha)\mathbf{p} : \mathbf{p}\in\mathcal{P}\},\quad 0 \le \alpha < 1 \quad (15)$$

$$\rho\mathbb{B} := \{\mathbf{p}\in\mathbb{R}^K : ||\mathbf{p}||_2 \le \rho\}. \quad (16)$$

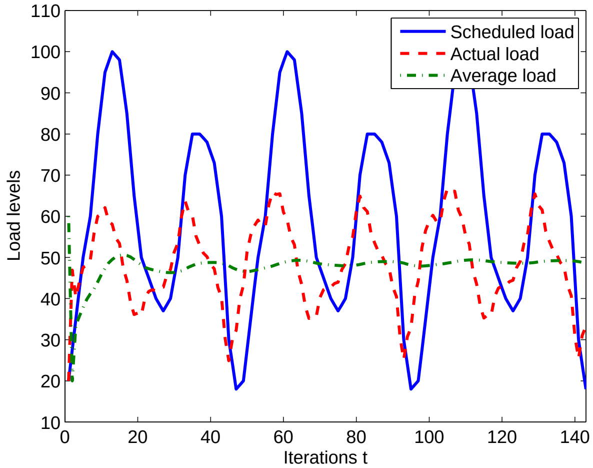

Fig. 1. Load curves before and after real-time DR.

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 (9). 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 $\mathbf{p}$). The regret bound for this algorithm is assessed in the following proposition, whose proof is omitted due to space limitation.

Proposition 2: *Let $\mathcal{P} \subset \mathbb{R}^K$ be a compact 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 $|\phi^t(\mathbf{p})+r(\mathbf{p})| \leq C$ for all t and $\mathbf{p} \in \mathcal{P}$. Since $\mathcal{P}$ is compact, c^t is Lipschitz continuous with Lipschitz constant L. Then, for sufficiently large T, the algorithm in Table I with*

$$\delta = \sqrt{\frac{\rho_{\text{out}} CK}{L + 2C/\rho_{\text{in}}}} T^{-\frac{1}{4}} \tag{17}$$

$$\alpha = \delta/\rho_{\text{in}} \tag{18}$$

$$\eta = \sqrt{\frac{{\rho_{\text{out}}}^3}{CK(L+2C/\rho_{\text{in}})}} T^{-\frac{3}{4}} \tag{19}$$

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

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(\phi^t(\tilde{\mathbf{p}}^t) + r(\tilde{\mathbf{p}}^t)\right)\right] - \min_{\mathbf{p}\in\mathcal{P}} \sum_{t=1}^{T}\left(\phi^t(\mathbf{p}) + r(\mathbf{p})\right) = O\left(\rho_{\text{out}} CK\left(L + 2C/\rho_{\text{in}}\right) T^{\frac{3}{4}}\right). \tag{20}$$

Note that condition $|\phi^t(\mathbf{p})+r(\mathbf{p})| \leq C$ is satisfied for (5)–(6) 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.

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 $k =$

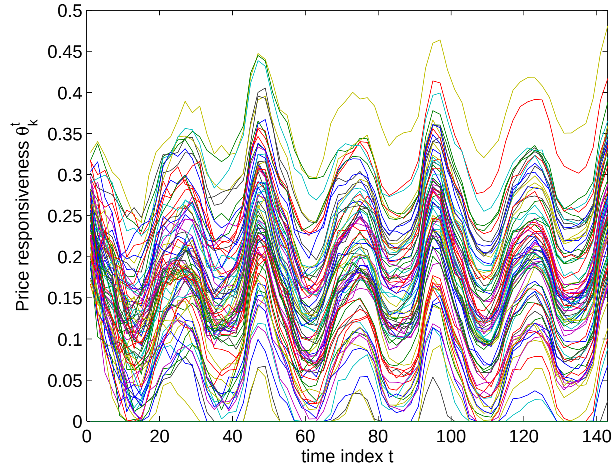

Fig. 2. Time-varying price responsiveness.

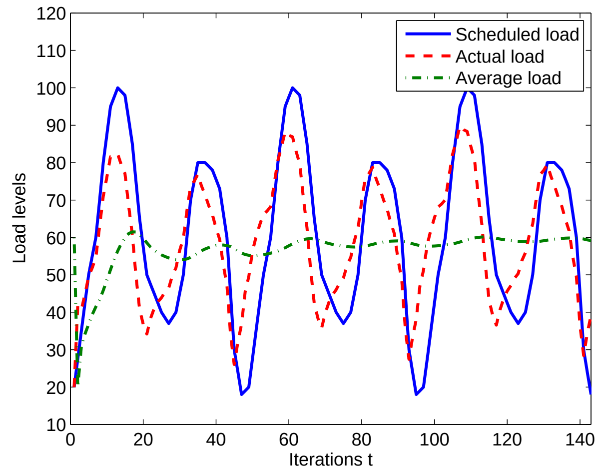

Fig. 3. Load curves for time-varying elasticity.

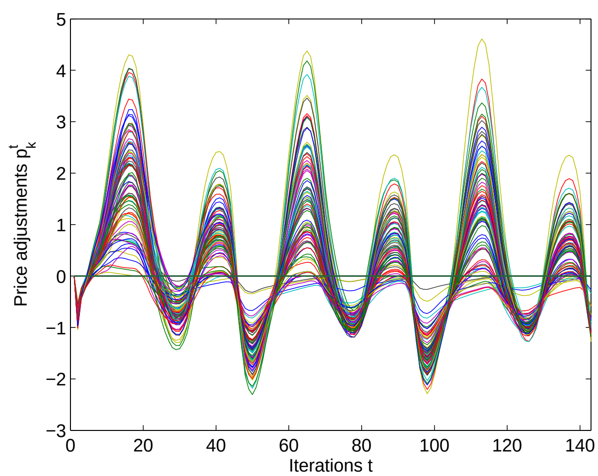

Fig. 4. Price adjustments in the full information case.

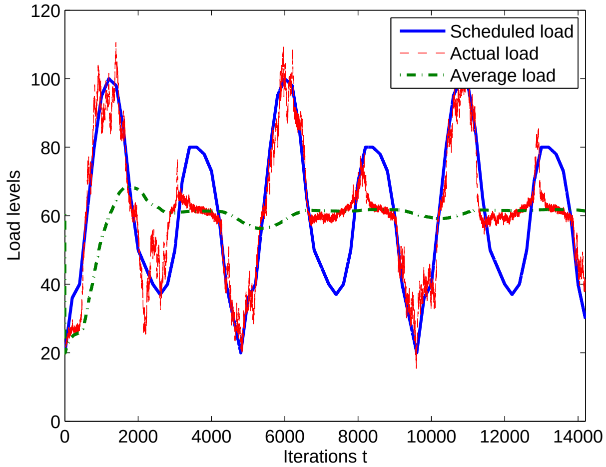

Fig. 5. Load curves in the bandit case.

$1, \ldots, 80$ were picked uniformly at random from $[0, 0.5]$, while $\theta_k = 0$ for $k = 81, \ldots, 100$. A maximum price deviation of $\overline{P}_k = -\overline{P}_k = 5$ 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. The values of m^t were set as $m^{t+1} = (t-1)t^{-1}m^t + t^{-1}(l^t - \boldsymbol{\theta}^{t^\top}\mathbf{p}^t)$. In Fig. 1, the solid curve depicts the base aggregate load l^t, and the dashed one corresponds to the overall load $(l^t - \boldsymbol{\theta}^{t^\top}\mathbf{p}^T)$, 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.

Another full-information case employing time-varying $\boldsymbol{\theta}^t$ was considered, where the employed values of θ^t are shown in Fig. 2. It is noted that the case is challenging as the consumers tend to have low responsiveness during the peak periods. The corresponding load curves in Fig. 3 still exhibit sizable improvement over the base load. Fig. 4 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. It is also apparent that the algorithm correctly identified the subset of non-elastic consumers, and set their price adjustments to zero, thanks to the sparsity-promoting regularizer. Incorporating sparsity may improve tracking abilities, in addition to possibly lowering the signaling overhead in our context.

Fig. 5 presents the load curves for the bandit setting. Again, $K = 100$ consumers with constant random $\theta_k^t \in [0, 0.5]$ for 80 of them were considered. The values of $\lambda = 0.5$ and $\mu = 0$ were used. The updates were performed every 18 seconds. It is seen from the figure that it takes much longer time to track compared to the full information case. In fact, for the larger magnitude peaks, the dynamics of l^t seems to be too fast for the algorithm to track, while for smaller peaks, the algorithm could significantly reduce the peaks. 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 showed that the novel approach could reduce load variation significantly through implicitly learning load elasticity in an online fashion.

References

[1] 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.

[2] 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://http://arxiv.org/abs/arXiv:1106.5063

[3] 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.

[4] 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.

[5] N. Y. Soltani, S.-J. Kim, and G. B. Giannakis, "Online learning of load elasticity for electric vehicle charging," in *Proc. of the 5th Intl. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, St. Martin Island, French Caribbean, Dec. 2013.

[6] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, Mar. 2012.

[7] B. Narayanaswamy, T. S. Jayram, and V. N. Yoong, "Hedging strategies for renewable resource integration and uncertainty management in the smart grid," in *Proc. of the 3rd IEEE Intl. Conf. and Exhibition on Innovative Smart Grid Technologies–Europe*, Berlin, Germany, Oct. 2012.

[8] S.-J. Kim and G. B. Giannakis, "Scalable and robust demand response with mixed-integer constraints," *IEEE Trans. Smart Grid*, 2013, to appear. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6510544

[9] L. Gan, U. Topcu, and S. Low, "Optimal decentralized protocol for electric vehicle charging," in *Proc. of the 50th IEEE Conf. on Decision and Control*, Orlando, FL, Dec. 2011, pp. 5798–5804.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.