

Socially Optimal Energy Usage via Adaptive Pricing

Jiayi Li, Matthew Motoki and Baosen Zhang

Electrical and Computer Engineering, University of Washington
{lji9712, mmotoki, zhangbao}@uw.edu

Abstract—A central challenge in using price signals to coordinate the electricity consumption of a group of users is the operator’s lack of knowledge of the users due to privacy concerns. In this paper, we develop a two-time-scale incentive mechanism that alternately updates between the users and a system operator. As long as the users can optimize their own consumption subject to a given price, the operator does not need to know or attempt to learn any private information of the users for price design. Users adjust their consumption following the price and the system redesigns the price based on the users’ consumption. We show that under mild assumptions, this iterative process converges to the social welfare solution. In particular, the cost of the users need not always be convex and its consumption can be the output of a machine learning-based load control algorithm.

Index Terms—Decentralized Demand Response, Iterative Algorithms, Pricing, Social Welfare Optimization

□

I. INTRODUCTION

We study the coordination of the electricity usage of a group of users by an operator. This question has been studied extensively by the community, for example, in the context of demand response, customer aggregation, and virtual power plants (see [1], [2], [3], [4] and the references within). The common setup is where each user is endowed with a cost (or utility) function, and the operator seeks to minimize a global cost function that is made up of the individual costs of the users and social welfare considerations.

A central challenge in these problems is that the cost functions of the users may not be known to the operator. Furthermore, users themselves may not be able to provide an analytical description. For example, using machine learning to schedule and manage demand has been a very active area of research (see, e.g. [5], [6], [7], [8], [9] and the references within). However, most learning algorithms minimize some cost or discomfort by changing the behavior of the users or their devices and do not provide a cost function in terms of power that can be easily optimized. It’s important to note that we do not mean that a cost function does not exist, rather, it is often not revealed when learning algorithms are applied. In many existing algorithms, however, the operator needs to somehow learn the cost functions of the users [10], [11], [12]. This restricts the users to having simple (typically quadratic) cost functions and is often too restrictive to be implemented in practice.

In this paper, we overcome this challenge by deploying a two-time-scale incentive mechanism that alternatively updates between the operator and the users. More concretely, the actions of the users are characterized by their electricity consumption and the action of the operator is in selecting a price vector. This price is an externality to the users, and a user optimizes its actions by minimizing its own internal cost and the external cost induced by the presence of a price. The operator adaptively updates the price to influence the behavior of the users to achieve a socially beneficial solution. The adaptive updates can be treated as a negotiation process: the customer and the system operator are negotiating on a future price that optimizes the social welfare of the system. Note that as long as users can optimize their actions given a price, it suffices for the operator to just observe the actions of the users, and the operator does not need to know or learn users’ private information.

These types of incentive designs where an operator sets a payment (or sometimes called a tax) have received significant interest. Typically, the goal is to provide incentives to selfish users such that the resulting equilibrium is closer to the socially optimal solution [13], [14], [15]. In the context of electricity markets, a large number of demand response strategies have been proposed, ranging from alert/text-based signals [16], to pricing [17], to direct load control [18]. In this paper, we focus on developing pricing mechanisms. The incentive mechanism we adopt is from [19], which is a type of dynamic incentive that evolves with the actions of the users. The key technical questions in these types of mechanisms are twofold: whether the price update converges, and if so, whether it leads to socially optimal user behaviors.

Existing results in answering these two questions often require strong assumptions that do not readily apply to electricity markets. In [20], the cost function of the users and the system cost are assumed to be linear. The results in [21] generalize the class of cost functions but assume the users’ strategies are convex in the price. This assumption often does not hold even in relatively simple games induced by the price. The work of [22] is similar to ours but makes much stronger assumptions that the system and users’ costs are convex and smooth. Moreover, our algorithm circumvents the need for each user to determine the incremental demand at each step, which is challenging, especially for learning-based methods. The work in [19], which acts as the impetus to this paper, shows that the price converges and leads to socially optimal behavior assuming strict convexity of the cost functions.

In this paper, we relax the assumptions needed in [20], [21], [19], and show that the two-time-scale algorithm leads to price

The authors are partially supported by NSF award CNS-1931718 and the Washington Clean Energy Institute

dynamics that both converge and induce optimal social behavior of the users. For example, we show that in some settings, the only assumption needed is that a user's action (electricity consumption) is decreasing in price. This assumption is intuitive because higher price naturally leads to lower consumption. This result would then apply to users who use learning algorithms to manage their demands. We also generalize the type of social costs that operators could consider. For example, in addition to widely used quadratic functions and time-of-use pricing, our framework provides a way to implement peak pricing through designing the social cost.

In our proof of convergence, we construct a continuous-time dynamical system and its associated Lyapunov functions. Then we show that the price leads to user actions that are optimal for the global optimization problem by analyzing the optimality conditions. The paper is organized as follows. In Section II we describe the setup of our model. In Section III we provide the theoretical results, and Section IV illustrates these results with several simulated case studies. We conclude and give some thoughts in future directions in Section V.

II. PROBLEM FORMULATION AND PRICING UPDATES

In this section, we formally state the problem of obtaining socially optimal energy usage and describe an iterative procedure to compute the price seen by each of the individual users.

A. Global Optimization Problem

We consider a system with N users. The action of each user is their energy consumption over T time periods. For example, $T = 24$ if we are considering hourly loads of the users in a day. Let $\mathbf{x}_i \in \mathbb{R}^T$ denote the electricity consumption of user i .¹ We assume $\mathbf{x}_i$ takes value in a compact set $\mathcal{X}_i \subset \mathbb{R}^T$.² The notation $x_{i,t}$ denote the energy consumption of user i at the t 'th period. Each user is interested in optimizing its cost function, $f_i : \mathbb{R}^T \rightarrow \mathbb{R}$. For example, if $f_i(\mathbf{x}_i) = \frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_i)^2$, the user can be interpreted as minimizing its energy usage to some predetermined setpoint $\bar{\mathbf{x}}_i$. Note that in this paper we do not assume any symmetry between the users, nor do we restrict f_i to be in particular parametric forms. Of course, some assumptions on f_i 's are required, and these will be stated closer to the technical results section.

We also place a cost on the system in serving the total energy demanded by the users. We do not distinguish between which user is using the energy, and the system cost is in the form of $g\left(\sum_{i=1}^N \mathbf{x}_i\right)$. Note the sum in g is over the users, and g is a function from $\mathbb{R}^T$ to $\mathbb{R}$. Throughout this paper, we will assume that g is convex or strictly convex. Several examples of g include $g\left(\sum_i \mathbf{x}_i\right) = c\left(\sum_i \mathbf{x}_i\right)$, where $c(\cdot)$ is the cost of procuring energy; $g\left(\sum_i \mathbf{x}_i\right) = \|\sum_i \mathbf{x}_i\|^2$, where the system tries to reduce its total energy consumption; and $g\left(\sum_i \mathbf{x}_i\right) = \max\left(\sum_i x_{i,1}, \dots, \sum_i x_{i,T}\right)$, which represent a cost on the peak energy demand [23].

¹We use bold notation to indicate vectors and matrices.

²The constraints that a user has can be encoded in this set.

The system operator is interested in solving a global social welfare problem:

$$\min_{\mathbf{x}} \sum_i f_i(\mathbf{x}_i) + g\left(\sum_i \mathbf{x}_i\right). \quad (1)$$

The first term represents the sum of the users' costs, while the second represents a social or system cost. This problem has been studied in many settings. When the operator has full information about f_i 's and g , and can set the demand of the users, this is essentially a variant of the economic (load) dispatch problem. This has been extensively studied and readers can refer to [1], [24] and the references within.

In practice, it is often the case that the system operator does not know the exact cost functions of the users. This could stem from privacy concerns, where the user does not communicate their cost to the operator. Another way that incomplete information can arise is that the users themselves do not have a closed-form description of their own costs. Learning algorithms are becoming increasingly popular for demand management, where a user's action is trained without explicitly learning a cost model [25], [8], [7]. In this setting, an iterative algorithm is often used, where local steps and global steps alternate and try to converge to the solution of the social welfare problem in (1).

When f_i 's are convex (or strictly convex), differentiable and frequent communications are possible, primal/dual type of gradient algorithms can be used to solve it. More efficient algorithms exist when f_i 's are assumed to be in particular parametric forms (mostly quadratic [10], [11]). Drawbacks of these approaches appear when the local problem is more complicated. For example, suppose user i uses tabular Q -learning to determine its actions. Then it is difficult to apply existing distributed optimization algorithms.

In this paper, we study a strategy to iteratively solve (1), where we can weaken the existing assumptions on the users. We do this through price updates and more details are given next.

B. User's Optimization Problem

From a user's perspective, it receives a price vector, $\mathbf{p} \in \mathbb{R}^T$ from the system operator. We will define how the operator arrives at this price in the next section, but for now, we treat it as a given vector. User i solves the following optimization problem:

$$\min_{\mathbf{x}_i \in \mathcal{X}_i} f_i(\mathbf{x}_i) + \mathbf{p}^\top \mathbf{x}_i. \quad (2)$$

We denote the solution of this problem as $\mathbf{x}_i^*(\mathbf{p})$. We do not consider the details of how user i solves (2) to obtain $\mathbf{x}_i^*(\mathbf{p})$. If f_i is available, the user may explicitly solve (2). In [25], a user represents a water heater and has the goal of balancing the discomfort of receiving underheated or overheated water with the cost of power. Since the discomfort is easily expressed in terms of the temperature of the water but not in terms of power, a Q -learning algorithm is used to look up $\mathbf{x}_i^*(\mathbf{p})$.

In this paper, we make the following assumption:

Assumption 1. Given a price $\mathbf{p}$, the solution $\mathbf{x}_i^*(\mathbf{p})$ to (2) is unique for all users i .

This says that the behavior of a user is uniquely determined by the price it sees from the system operator. This condition is true under a wide range of conditions, and we note it is much weaker than the ones made in existing literature. For example, it holds if f_i is strictly convex (as assumed in [19], [20], [26], [22]), or if the user used a deterministic algorithm regardless of convexity (not covered by the conditions in [19], [20], [26], [22]). If a probabilistic algorithm is used, the results in the paper hold with respect to the averaged solution, but the analysis becomes much more cumbersome. A more refined understanding of possible stochastic behaviors is an important future direction for us.

It's worth mentioning that this assumption is not strictly necessary for the adaptive algorithm to find socially optimal incentives. However, without this assumption, the updates may oscillate between several different solutions. When the adaptive algorithm does not converge to a unique solution, the notion of convergence becomes more difficult to characterize. In response to the potential oscillations that destabilize the system, the system operator could develop a consistent tie-breaking procedure (e.g. consuming at the earliest hour) so the system would converge to a fixed point. In addition, we could expand the allowed behavior of the agents by considering ideas such as convergence in expectation.

C. Price Update

Now we consider how $\mathbf{p}$ should be updated by the system operator. Following the scheme in [19], we define the externality $\mathbf{e}$ as the marginal social cost arising from the term g :

$$\mathbf{e}(\mathbf{x}) = \nabla_{\mathbf{z}} g(\mathbf{z})|_{\mathbf{z}=\sum_i \mathbf{x}_i} \quad (3)$$

where $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^N$. The price is updated as

$$\mathbf{p}_{k+1} = (1 - \beta_k)\mathbf{p}_k + \beta_k \mathbf{e}(\mathbf{x}^*(\mathbf{p}_k)). \quad (4)$$

where $\mathbf{x}^*(\mathbf{p}_k) = \{\mathbf{x}_i^*(\mathbf{p}_k)\}_{i=1}^N$ and the following conditions hold $\sum_{k=1}^{\infty} \beta_k = \infty$ and $\sum_{k=1}^{\infty} \beta_k^2 < \infty$; e.g., $\beta_k = 1/k$. These assumptions about the step size are standard assumptions that allow us to prove the convergence of the price updates.³

The main technical results of the paper are to show that under mild assumptions, the dynamics in (4) converge to a unique $\mathbf{p}^*$ and $\mathbf{x}_i^*(\mathbf{p}^*)$ is the minimizer of (1). This price update scheme is attractive because the operator doesn't need to know or learn about the users.⁴ As long as the sum of decisions $\mathbf{x}_i$'s are observed, a unique $\mathbf{p}^*$ can be found to induce globally optimal behavior.

³We note that step sizes of $1/k$ are required for theoretical reasons; however, we find that experimentally that other step sizes can lead to the same fixed point with fewer iterations.

⁴In fact, the operator does not even need to know how many users there are in the system since it simply broadcasts the price to all users.

Algorithm 1 Adaptive Pricing: An iterative method that solves the global optimization problem (1).

```

1: Initialize  $\mathbf{p} \in \mathbb{R}^T$ 
2: for  $k = 1, 2, \dots$  do
3:   // Solve Local Optimization
4:   for  $i = 1, 2, \dots, N$  do
5:      $\mathbf{x}_i^*(\mathbf{p}) = \arg \min_{\mathbf{x}_i \in \mathcal{X}_i} f_i(\mathbf{x}_i) + \mathbf{p}^\top \mathbf{x}_i$ 
6:   end for
7:   // Price Update
8:   Choose step size:  $\beta = 1/k$ 
9:   Compute externality:  $\mathbf{e} = \nabla g\left(\sum_{i=1}^N \mathbf{x}_i^*(\mathbf{p})\right)$ 
10:  Compute new price:  $\mathbf{p} = (1 - \beta)\mathbf{p} + \beta \mathbf{e}$ 
11: end for

```

III. THEORETICAL RESULTS

The first step to analyzing the system in (4) is to approximate it as a continuous system. A standard result in dynamical systems is that as $k \rightarrow \infty$ and the step-sizes β_k are asymptotically going to zero, (4) behaves as (5) [19], [27]:

$$\dot{\mathbf{p}} = \mathbf{e}(\mathbf{x}^*(\mathbf{p})) - \mathbf{p}. \quad (5)$$

We want (5) to have two properties. The first is that it has a unique equilibrium, and the second is that this equilibrium induces the solutions to the local optimization problem in (2) that's simultaneously solving the global problem in (1).

The proof of these results involves constructing a Lyapunov function based on suitable assumptions on f_i 's and g . We feel it is easier to break the proof into two parts, and first consider the case of $T = 1$ and then $T > 1$. When $T = 1$, the proof is fairly short and the intuition carries over to the $T > 1$ case, where the math becomes more cumbersome.

A. Single time-period case ($T = 1$)

For a single time-dimensional case, $x_i \in \mathbb{R}$, g is a scalar function, and the dynamical system in (5) is also a scalar system. We have the following result

Theorem 1. Suppose Assumption 1 holds and $x_i^*(p)$ is decreasing for each i and g is convex and differentiable. Then the scalar dynamical system (5) is globally asymptotically stable and has a unique fixed point p^* . The solutions $x_i^*(p^*)$ solve the global optimization problem in (1).

The decreasing condition on $x_i^*(p)$ says that if the price increases, then the energy consumption of user i will not increase (users will not consume more if energy becomes more expensive). This condition is satisfied for most types of cost functions. For example, if f_i 's are convex. The decreasing condition is also satisfied for situations where f_i 's are nonconvex [28]. Therefore, our pricing algorithm is applicable to a wider range of problems than existing pricing schemes that require both f_i 's and g to be convex [19].

Proof. We first show that a unique equilibrium of (5) exists. When $T = 1$, (3) simplifies to $e(x) = g'(\sum_i x_i)$. If g is convex,

its derivative is increasing and e is an increasing function. Since $x_i^*(p)$ is decreasing in p , the function of p obtained by summing over i , $\sum_i x_i^*(p)$, is also decreasing in p . Using the fact that for scalar functions, the composition of an increasing function with a decreasing function is decreasing, we have $e(\sum_i x_i^*(p))$ being decreasing in p .

Looking at the right hand side of (5), $e(x^*(p)) - p$ is a strictly decreasing function ranging from ∞ to $-\infty$ as p goes from $-\infty$ to ∞ . Therefore it takes the value 0 at exactly one point, and we denote that point by p^* , and it is an equilibrium point of the dynamical system in (5).

Next, we show that the dynamical system is globally asymptotically stable around p^* . We do this by constructing the following Lyapunov function:

$$V(p) = - \int_{p^*}^p [e(x^*(q)) - q] dq.$$

Using the fact that the integrand is strictly decreasing, is easy to verify that $V(p) > 0$ if $p \neq p^*$ and $V(p^*) = 0$. The time derivative of V is

$$\dot{V}(p) = V'(p)\dot{p} = -[e(x(p)) - p]^2,$$

which is negative except when $p = p^*$. Therefore, p^* is globally asymptotically stable.

Finally, we show that the p^* induces globally optimal behavior. That is, the solutions of the local problems $x_i^*(p^*)$ solves the global optimization problem in (1). To show this, let $\hat{x}$ denote an optimal solution to (1). For notational simplicity, define $\hat{x}_s = \sum_i \hat{x}_i$ and $x_s^* = \sum_i x_i^*(p^*)$. Suppose $\hat{x} \neq x^*(p^*)$, and $\hat{x}$ achieves a lower social welfare cost. Then we have

$$\sum_i f_i(\hat{x}_i) + g(\hat{x}_s) < \sum_i f_i(x_i^*(p^*)) + g(x_s^*). \quad (6)$$

But $x_i^*(p^*)$ are the minimizers of the local optimization problem in (2), we have

$$f_i(x_i^*(p^*)) + g'(x_s^*)x_i^*(p^*) < f_i(\hat{x}_i) + g'(x_s^*)\hat{x}_i, \quad \forall i, \quad (7)$$

where we use the fact that $p^* = g'(x_s^*)$ and the inequality is strict because of the uniqueness assumption in Assumption 1. Adding (6) and (7) and simplifying the expressions, we get

$$g(\hat{x}_s) < g(x_s^*) + g'(x_s^*)(\hat{x}_s - x_s^*).$$

This is the first-order condition of a strictly *concave* function, and it contradicts the fact that g is *convex*. Hence the assumption that there exists a more optimal solution to (1) than x^* is not possible. $\square$

B. Multi-Time Period Case ($T > 1$)

The analysis in the single time period case relies on the monotonicity of several functions. To extend that to the multi-time period case, we extend the notion of monotonicity to vector-valued functions:

Definition 1. Consider a vector-valued function $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$. We say that h is increasing if $[h(\mathbf{x}) - h(\mathbf{x}')]^T (\mathbf{x} - \mathbf{x}') \geq 0$. It is strictly increasing if the inequality is strict when $\mathbf{x} \neq \mathbf{x}'$.

A function is (strictly) decreasing if its negation is (strictly) increasing.

We adopt Definition 1 since it is a natural generalization of monotonicity for a scalar function in the following sense: the gradient of a differentiable convex function is increasing.⁵

In the $T = 1$ setting, the proof for Theorem 1 has three parts. First, we showed that a unique equilibrium price exists, then we show that the dynamical system in (5) is asymptotically stable around this price, and finally that this price induces socially optimal behavior. In the $T > 1$ setting, the first and last parts generalize directly from the $T = 1$ setting with minimal changes. However, asymptotic stability is considerably more difficult since the composition of two monotone functions is no longer monotone.⁶ Therefore, we first state a lemma about the uniqueness and social optimality of the equilibrium price. Then we provide two conditions where the iterative algorithm converges asymptotically to this equilibrium.

Lemma 2. Consider the dynamical system in (5). If $\mathbf{x}_i^*(\mathbf{p})$ is a decreasing function of $\mathbf{p}$, then there exists a unique equilibrium $\mathbf{p}^*$. Furthermore, under Assumption 1, $\mathbf{x}^*(\mathbf{p}^*)$ is the optimal solution to social welfare problem in (1).

Proof. The proof of these two results is analogous to the $T = 1$ case. By sweeping $\mathbf{p}$ from ∞ to $-\infty$ and using the fact that $\mathbf{e}(\mathbf{x}^*(\mathbf{p})) - \mathbf{p}$ is strictly decreasing, we can conclude that there exists a unique $\mathbf{p}^*$ such that $\mathbf{e}(\mathbf{x}^*(\mathbf{p}^*)) - \mathbf{p}^* = 0$. For the second part of the lemma, it suffices to change a derivative to a gradient in (7) and the product between scalars to a dot product between vectors. All other steps remain the same. $\square$

Next, we state two conditions where the dynamical system in (5) converges:

Theorem 3. Suppose the local solutions $\mathbf{x}_i^*(\mathbf{p})$ are decreasing in $\mathbf{p}$. The dynamical system in (5) is asymptotically stable if one of the following conditions are true

- 1) $g(\mathbf{z}) = \frac{1}{2}\mathbf{z}^T \mathbf{B} \mathbf{z}$ for some positive definite $\mathbf{B}$.
- 2) Each f_i is twice differentiable and strictly convex and g is strictly convex and differentiable.

Before presenting the proof, we comment on why these settings are potentially interesting and practical. A quadratic cost is the most common form used in the literature [29] since it includes standard pricing schemes such as flat rates and time-of-use (ToU) rates. The assumption on each of the users is minimal since we just require that an increase in energy price leads to a decrease in energy consumption.

The second setting generalizes the price function by assuming a more strict form of the local costs f_i 's. The reason this result is useful is that it includes peak pricing, which is being adopted by a large number of users [30], [31], [32], [33]. More specifically, peak pricing has the form of

⁵We note there exists other definition of monotonicity for vector-valued functions. For example, sometimes a function is said to be monotone if it is monotone in every coordinate. This definition, however, is too restrictive for our purposes.

⁶This is not even true for linear functions in the vector-valued case.

$g(\mathbf{z}) = c \cdot \max(z_1, \dots, z_T)$, where c is some constant. This function is convex, but not strictly convex nor differentiable. But it can be approximated by the so-called LogSumExp function, defined as

$$\text{LSE}(z_1, \dots, z_T) = \frac{1}{\alpha} \log(\exp(\alpha z_1) + \dots + \exp(\alpha z_T)). \quad (8)$$

The LSE function is strictly convex if at least one of its arguments is positive and approximates the max function arbitrarily as α grows [34].

Proof. We prove Theorem 3 by constructing appropriate Lyapunov functions. In the first case, we use a quadratic Lyapunov function

$$V(\mathbf{p}) = \frac{1}{2}(\mathbf{p} - \mathbf{p}^*)^\top \mathbf{B}^{-1}(\mathbf{p} - \mathbf{p}^*). \quad (9)$$

It is clear that $V(\mathbf{p}) = 0$ if $\mathbf{p} = \mathbf{p}^*$ and $V(\mathbf{p}) > 0$ otherwise.

In the second setting, we use a different Lyapunov function, defined as

$$\begin{aligned} V(p) = & g\left(\sum_i \mathbf{x}_i^*(\mathbf{p})\right) - g\left(\sum_i \mathbf{x}_i^*(\mathbf{p}^*)\right) \\ & - \nabla g\left(\sum_i \mathbf{x}_i^*(\mathbf{p}^*)\right)^\top \left(\sum_i (\mathbf{x}_i^*(\mathbf{p}) - \mathbf{x}_i^*(\mathbf{p}^*))\right) \\ & - \sum_i (\mathbf{x}_i^*(\mathbf{p}) - \mathbf{x}_i^*(\mathbf{p}^*))^\top (\mathbf{p} - \mathbf{p}^*). \end{aligned} \quad (10)$$

This function is based on the Bregman's divergence of g , and using the fact that g is strictly convex, $V(\mathbf{p}) = 0$ if $\mathbf{p} = \mathbf{p}^*$ and $V(\mathbf{p}) > 0$ otherwise.

In Appendix A, we show that the time derivative of both Lyapunov functions are negative when $\mathbf{p} \neq \mathbf{p}^*$. $\square$

The proof above illustrates the difficulty of showing a general result when $T > 1$ since we need two different Lyapunov functions. However, we believe this is a consequence of our proof technique, and we conjecture that the twice differentiable and strictly convex condition in Theorem 3 can be relaxed to be just $\mathbf{x}_i^*(\mathbf{p})$ is strictly decreasing for all i .

IV. SIMULATION RESULTS

In this section, we conduct numerical studies that illustrate the theoretical results in this paper, in particular the convergence properties of the update in (4). Our code for the numerical simulations can be found at <https://github.com/socially-optimal-energy>.

A. Single Time-Period

The first is a simple single time-period ($T = 1$) example that clearly shows the convergence to the global solution (see Fig. 1). Suppose there are 5 users ($N = 5$) that have actions $x_i \in \mathbb{R}$ which can automatically adapt to changes in price [7]. The user cost functions are $f_i(x_i) = \frac{1}{2}(x_i - \bar{x}_i)^2$ and the global cost is $g(\sum_i x_i) = (\sum_i x_i)^2$. Both actions and the price converge

⁷In this example, we assume that the users and utility are in a negotiating a future price (e.g., a day-ahead price) and the algorithm converges by the time the price needs to be set and the load is realized.

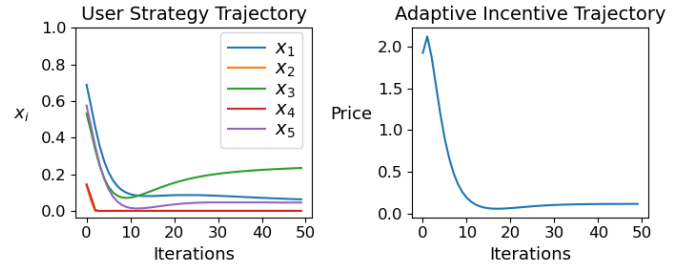


Fig. 1. Convergence of user actions and price incentive for a system with 5 users and a single time-period. Both the actions and the price converge quickly.

quickly, and it is easy to check that they converge to the optimal values (which can be computed in this simple setup).

B. Peak Pricing of Multiple Time-Periods

Here, we consider multiple periods, with $T = 24$. Then user i 's action is $\mathbf{x}_i \in \mathbb{R}^{24}$. For simplicity, let $f_i(\mathbf{x}_i) = \frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_i)^2$ for some given $\bar{\mathbf{x}}_i$'s. Suppose the system adopts peak pricing, approximated as $g(\sum_i \mathbf{x}_i) = \lambda \text{LSE}(\sum_i \mathbf{x}_i)$ for some λ , where the LSE function was defined in (8). Following (3), the externality of each user is calculated as:

$$\mathbf{e}(\mathbf{x}) = \nabla_{\mathbf{z}} g(\mathbf{z})|_{\mathbf{z}=\sum_i \mathbf{x}_i} = \frac{1}{\sum_{t=1}^T \exp(\alpha \mathbf{z}_t)} \exp(\alpha \mathbf{z}).$$

For 10 users, the convergence of price and total demand is shown in Fig. 2, where both converge quickly to their final values. The comparison of initial and final price and load profiles are shown in Fig. 3. The initial prices (chosen uniformly at random) induce a total (summed over the users at each time period) load profile that is uneven. After convergence, load profiles become much more even, which is what we expect to see when the peak system load is minimized. Note the price that achieves this still shows variations across time periods.

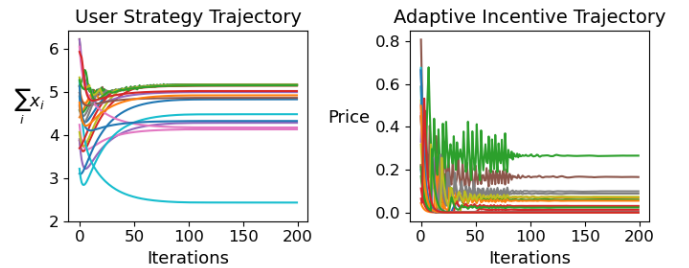


Fig. 2. The top figure shows the convergence in the sum of the users' actions at each time period. The bottom figure shows the convergence of price.

C. Water Heater Load Optimization with Q-Learning

Now we consider a case where users' actions are determined by a learning algorithm and show that the f_i 's need not be differentiable or convex. We consider a group of water heaters where $\mathbf{x}_i \in \mathbb{R}^{96}$ represents its daily load profile measured every

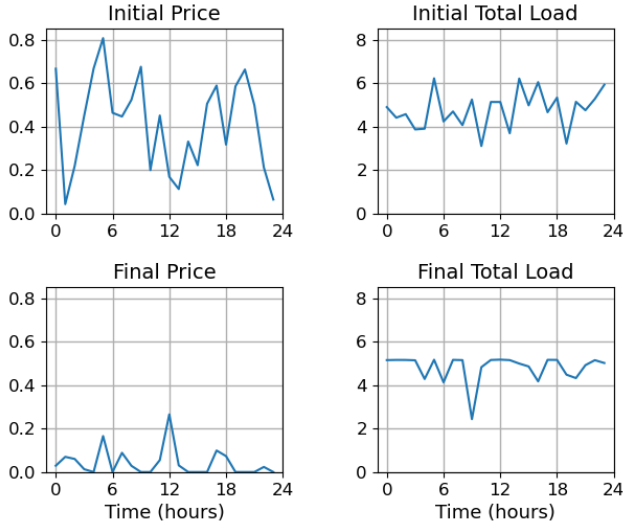


Fig. 3. The initial and converged price and demand profiles demonstrate that this adaptive pricing framework is effective in reducing system peak.

15-minutes, f_i represents user discomfort, and g is the LSE function. In this example, users can flexibly shift their energy consumption earlier to avoid high prices and store the energy (hot water) until it is needed. We model and optimize each user following the learning algorithm in [25] and numerically check that $\mathbf{x}_i^*$ is decreasing in $\mathbf{p}$.

For simplicity, we assume that each user's demand for hot water is a binary vector drawn from a sequence of Bernoulli random variables. We assume the demand and all system parameters are known and so given a price $\mathbf{p}$, we can solve for $\mathbf{x}^*(\mathbf{p})$ exactly using dynamic programming. Fig. 4 shows the initial and final price and the corresponding distributions of total energy consumption of all the users. Initially, the system cost is 12.781 and the user cost is 0.630. At convergence, the system cost decreases to 10.397 and the user cost to 0.627. Overall, the adaptive pricing procedure reduces the social welfare cost by 2.387 (17.8%).

V. CONCLUSION

In this paper, we consider the problem of using price signals to coordinate the electricity consumption of a group of users. We develop a two-time-scale incentive mechanism that alternately updates between the users and a system operator. We assume that users can optimize their actions given a price, and no private information is exchanged between the users and the system operator. In turn, the system operator does not need to learn the cost of the users. We show that the iterative algorithm converges to the socially optimal solution.

The algorithms proposed in this paper can be implemented by any smart device that can receive and respond to information from a system operator. For example, a home with a Nest thermostat could easily negotiate with a system to reach the price equilibrium without any manual intervention of a user. A key assumption in this paper is that each of the users are price takers, in the sense that they lack any market power

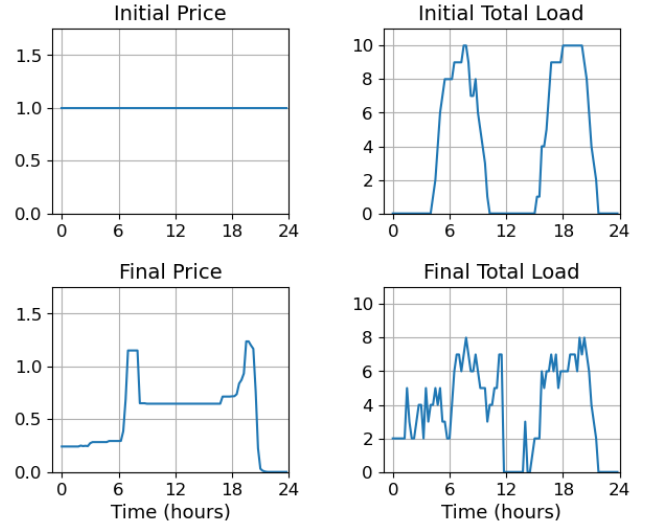


Fig. 4. Simulation results for water heater optimization. There are 10 users minimizing discomfort using a Q-learning algorithm [25], and the system operator tries to minimize the peak load. The initial price is chosen to be flat, which leads to a profile with high peaks. After several iterations, the price becomes uneven, and the final load has much lower peaks.

to explicitly manipulate the prices. Thus an important future direction for us is to consider the price-anticipatory behavior of users.

APPENDIX A PROOF OF THEOREM 3

Consider the first condition in Theorem 3. Again we let $\mathbf{x}_s$ denote the sum $\sum_i \mathbf{x}_i$. The time derivative of the Lyapunov function in (9) is

$$\begin{aligned}
 \dot{V}(\mathbf{p}) &= \nabla V(\mathbf{p})^\top \dot{\mathbf{p}} \\
 &= (\mathbf{p} - \mathbf{p}^*)^\top \mathbf{B}^{-1}(\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{p}) \\
 &\stackrel{(a)}{=} (\mathbf{p} - \mathbf{p}^*)^\top \mathbf{B}^{-1}(\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{p} - \mathbf{e}(\mathbf{x}_s^*(\mathbf{p}^*)) + \mathbf{p}^*) \\
 &\stackrel{(b)}{=} (\mathbf{p} - \mathbf{p}^*)^\top \mathbf{B}^{-1}(\mathbf{B}(\mathbf{x}_s^*(\mathbf{p}) - \mathbf{x}_s^*(\mathbf{p}^*)) - \mathbf{p} + \mathbf{p}^*) \\
 &= (\mathbf{p} - \mathbf{p}^*)^\top (\mathbf{x}_s^*(\mathbf{p}) - \mathbf{x}_s^*(\mathbf{p}^*)) \\
 &\quad - (\mathbf{p} - \mathbf{p}^*)^\top \mathbf{B}^{-1}(\mathbf{p} - \mathbf{p}^*) \\
 &\stackrel{(c)}{\leq} 0,
 \end{aligned}$$

where (a) follows from the fact that $\mathbf{p}^*$ is the equilibrium; (b) follows from the definition $\mathbf{e} = \nabla g$; (c) follows from the assumption that each $\mathbf{x}_i^*(\mathbf{p}^*)$ is decreasing (and hence their sum is) and $\mathbf{B}$ is positive definite. Furthermore, $V(\mathbf{p}) = 0$ only if $\mathbf{p} = \mathbf{p}^*$.

Next, we consider the second condition in Theorem 3. The

time derivative of the Lyapunov function in (10) is

$$\begin{aligned}\dot{V}(\mathbf{p}) &= \nabla V(\mathbf{p})^T \dot{\mathbf{p}} \\ &= \nabla V(\mathbf{p})^T (\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{e}(\mathbf{x}_s^*(\mathbf{p}^*)) - \mathbf{p} + \mathbf{p}^*) \\ &= (\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{e}(\mathbf{x}_s^*(\mathbf{p}^*)) - \mathbf{p} + \mathbf{p}^*)^T [\nabla_{\mathbf{p}} \mathbf{x}_s^*(\mathbf{p})] \\ &\quad (\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{e}(\mathbf{x}_s^*(\mathbf{p}^*)) - \mathbf{p} + \mathbf{p}^*) \quad (11) \\ &\quad - (\mathbf{x}_s^*(\mathbf{p}) - \mathbf{x}_s^*(\mathbf{p}^*))^T (\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{e}(\mathbf{x}_s^*(\mathbf{p}^*))) \\ &\quad + (\mathbf{x}_s^*(\mathbf{p}) - \mathbf{x}_s^*(\mathbf{p}^*))^T (\mathbf{p} - \mathbf{p}^*).\end{aligned}$$

It's easy to see that $\dot{V}(\mathbf{p}) = 0$ when $\mathbf{p} = \mathbf{p}^*$. We first look at the last two terms. Since $\mathbf{e} = \nabla g$ and g is strictly convex, $\mathbf{e}$ is strictly increasing, and $(\mathbf{x}_s^*(\mathbf{p}) - \mathbf{x}_s^*(\mathbf{p}^*))^T (\mathbf{e}(\mathbf{x}_s^*(\mathbf{p})) - \mathbf{e}(\mathbf{x}_s^*(\mathbf{p}^*))) > 0$ if $\mathbf{p} \neq \mathbf{p}^*$. By assumption, $\mathbf{x}_s^*(\mathbf{p})$ is decreasing and $(\mathbf{x}_s^*(\mathbf{p}) - \mathbf{x}_s^*(\mathbf{p}^*))^T (\mathbf{p} - \mathbf{p}^*) \leq 0$.

Next, consider the term in (11). The object $\nabla_{\mathbf{p}} \mathbf{x}_s^*(\mathbf{p})$ is the gradient of the vector $\mathbf{x}_s^*(\mathbf{p})$, hence it is a $T \times T$ matrix and it suffices to show $\nabla_{\mathbf{p}} \mathbf{x}_s^*(\mathbf{p})$ is negative definite. Using the fact that summations of negative definite matrices are negative definite, it is enough to show that $\nabla_{\mathbf{p}} \mathbf{x}_i^*(\mathbf{p})$ is negative definite for all i . Recall $\mathbf{x}_i^*(\mathbf{p})$ is the optimal solution to the local optimization problem, and it satisfies the first-order condition

$$\nabla_{\mathbf{x}_i} f_i(\mathbf{x}_i) + \mathbf{p} = 0.$$

Taking the derivative of $\mathbf{p}$ on both sides, applying the chain rule, and rearranging leads to:

$$\mathbf{I} = -[\nabla_{\mathbf{p}} \mathbf{x}_i^*(\mathbf{p})] [\mathbf{H}_{\mathbf{x}_i}(\mathbf{x}_i^*(\mathbf{p}))]$$

where $[\mathbf{H}_{\mathbf{x}_i}(\mathbf{x}_i^*(\mathbf{p}))]$ is the Hessian of f_i evaluated at $\mathbf{x}_i^*(\mathbf{p})$. Since f_i are assumed to be strictly convex and twice differentiable, $[\mathbf{H}_{\mathbf{x}_i}(\mathbf{x}_i^*(\mathbf{p}))]$ is positive definite. Rearranging the above equation gives $\nabla_{\mathbf{p}} \mathbf{x}_i^*(\mathbf{p}) = -[\mathbf{H}_{\mathbf{x}_i}(\mathbf{x}_i^*(\mathbf{p}))]^{-1}$. Therefore $\nabla_{\mathbf{p}} \mathbf{x}_i^*(\mathbf{p})$ is negative definite and $\dot{V}(\mathbf{p}) < 0$ when $\mathbf{p} \neq \mathbf{p}^*$.

REFERENCES

- [1] P. Pinson, H. Madsen *et al.*, "Benefits and challenges of electrical demand response: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 686–699, 2014.
- [2] C. Zhang, Y. Xu, Z. Y. Dong, and K. P. Wong, "Robust coordination of distributed generation and price-based demand response in microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4236–4247, 2017.
- [3] S. H. Tindemans, V. Trovato, and G. Strbac, "Decentralized control of the thermostatic loads for flexible demand response," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1685–1700, 2015.
- [4] M. R. Sarker, Y. Dvorkin, and M. A. Ortega-Vazquez, "Optimal participation of an electric vehicle aggregator in day-ahead energy and reserve markets," *IEEE transactions on power systems*, vol. 31, no. 5, pp. 3506–3515, 2015.
- [5] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *2010 First IEEE international conference on smart grid communications*. IEEE, 2010, pp. 409–414.
- [6] H. Li, Z. Wan, and H. He, "Constrained ev charging scheduling based on safe deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427–2439, 2019.
- [7] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied energy*, vol. 235, pp. 1072–1089, 2019.
- [8] B. Wang, Y. Li, W. Ming, and S. Wang, "Deep reinforcement learning method for demand response management of interruptible load," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3146–3155, 2020.
- [9] Y. Shi and B. Zhang, "Multi-agent reinforcement learning in cournot games," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 3561–3566.
- [10] P. Li, H. Wang, and B. Zhang, "A distributed online pricing strategy for demand response programs," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 350–360, 2019.
- [11] K. Khezeli and E. Bitar, "Risk-sensitive learning and pricing for demand response," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6000–6007, 2017.
- [12] S. Zheng, Y. Sun, B. Li, B. Qi, K. Shi, Y. Li, and X. Tu, "Incentive-based integrated demand response for multiple energy carriers considering behavioral coupling effect of consumers," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3231–3245, 2020.
- [13] T. Başar, "Affine incentive schemes for stochastic systems with dynamic information," *SIAM Journal on Control and Optimization*, vol. 22, no. 2, pp. 199–210, 1984.
- [14] Y.-C. Ho, P. B. Luh, and G. J. Olsder, "A control-theoretic view on incentives," *Automatica*, vol. 18, no. 2, pp. 167–179, 1982.
- [15] D. Paccagnan, R. Chandan, B. L. Ferguson, and J. R. Marden, "Incentivizing efficient use of shared infrastructure: Optimal tolls in congestion games," *arXiv preprint arXiv:1911.09806*, 2019.
- [16] M. Peplinski and K. T. Sanders, "Residential electricity demand on caiso flex alert days: a case study of voluntary emergency demand response programs," *Environmental Research: Energy*, vol. 1, no. 1, 2023.
- [17] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: Pricing methods and optimization algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152–178, 2014.
- [18] C. Chen, J. Wang, and S. Kishore, "A distributed direct load control approach for large-scale residential demand response," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2219–2228, 2014.
- [19] C. Maheshwari, K. Kulkarni, M. Wu, and S. S. Sastry, "Inducing social optimality in games via adaptive incentive design," in *Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 2864–2869.
- [20] L. J. Ratliff and T. Fiez, "Adaptive incentive design," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3871–3878, 2020.
- [21] B. Liu, J. Li, Z. Yang, H.-T. Wai, M. Hong, Y. M. Nie, and Z. Wang, "Inducing equilibria via incentives: Simultaneous design-and-play finds global optima," *arXiv:2110.01212*, 2021.
- [22] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *2011 IEEE power and energy society general meeting*. IEEE, 2011, pp. 1–8.
- [23] X. Yan, D. Wright, S. Kumar, G. Lee, and Y. Ozturk, "Enabling consumer behavior modification through real time energy pricing," in *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 2015, pp. 311–316.
- [24] W. Huang, N. Zhang, C. Kang, M. Li, and M. Huo, "From demand response to integrated demand response: Review and prospect of research and application," *Protection and Control of Modern Power Systems*, vol. 4, pp. 1–13, 2019.
- [25] M. Motoki, M. Umeda, M. Fripp, and A. Kuh, "Approximate dynamic programming for control of a residential water heater," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–8.
- [26] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing data center demand response," *ACM SIGMETRICS Performance Evaluation Review*, vol. 42, no. 1, pp. 111–123, 2014.
- [27] V. S. Borkar, "Stochastic approximation with two time scales," *Systems & Control Letters*, vol. 29, no. 5, pp. 291–294, 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167691197900153>
- [28] Y. Wan, T. Kober, and M. Densing, "Nonlinear inverse demand curves in electricity market modeling," *Energy Economics*, vol. 107, 2022.
- [29] D. S. Kirschen and G. Strbac, *Fundamentals of power system economics*. John Wiley & Sons, 2018.
- [30] Federal Energy Management Program, "Demand response and time-variable pricing programs: Western states," U.S. Department of Energy, Tech. Rep., 2024.
- [31] SmartGrid.gov, "Recovery act: Time based rate programs," U.S. Department of Energy, Tech. Rep., 2024.
- [32] G. R. Newsham and B. G. Bowker, "The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review," *Energy policy*, vol. 38, no. 7, pp. 3289–3296, 2010.
- [33] Y. Wang and L. Li, "Time-of-use electricity pricing for industrial customers: A survey of us utilities," *Applied Energy*, vol. 149, pp. 89–103, 2015.
- [34] F. Nielsen and G. Hadjeres, "Monte carlo information-geometric structures," *Geometric Structures of Information*, pp. 69–103, 2019.