Fast Distributed Demand Response with Spatially- and Temporally-Coupled Constraints in Smart Grid

Abstract—As the next generation power grid, smart grid is characterized as an informationized system, and demand response is one of its important features to deal with the ever increasing peak energy usage. However, the supply capacity and required demand make the demand response problem with both spatially- and temporally-coupled constraints which, to the best of our knowledge, has not been thoroughly investigated in a distributed manner. The complexity lies in how to guarantee privacy and convergence of distributed algorithms. Aiming at this challenge, in this paper, we firstly propose a distributed algorithm which is based on dual decomposition and does not require each user to reveal his/her private information. Then, the convergence analysis is conducted to provide guidance on how to choose the proper step size; through which, we notice that the convergence speed of the subgradient projection method is not fast enough, and it is highly dependent on the choice of the step size. Therefore, to increase the convergence rate of the distributed algorithm, we further propose a fast approach based on binary search. Finally, the distributed algorithms are illustrated by numerical simulations, and the extensive comparison results validate the better performance of the fast approach.

Index Terms—Convergence, demand response, distributed algorithm, smart grid, spatially-/temporally-coupled constraint.

I. INTRODUCTION

SMART grid has been widely regarded as the informationization of the traditional power grid, which leverages information and communications technology to fully upgrade the energy generation, transmission (including substation), and distribution [1]–[3]. The novel features in the context of smart grid are many and varied, which include demand response, renewable energy sources, vehicle-to-grid capability, advanced metering infrastructure, microgrid, and so on. Demand response, which is made feasible by the penetration of smart meters and introduction of two-way communications, is a critical component of smart grid [4]–[6]. This capability enables power users to adapt their energy consumptions in response to the fluctuations of electricity prices or economic incentives provided by utilities, further helping reduce the load at peak hours or during system contingencies [7]–[9]. In the long run, the ever-increasing spike energy usage could also be addressed by the demand response approach, instead of purely operating expensive generators or building more power plants [10]–[12].

This paper will focus on the demand response problem in a smart distribution grid, with the goal of maximizing the total welfare of all users and over all time slots. Specifically, we consider a system model with a load-serving entity (LSE) and multiple users (e.g., smart building/community or microgrid, as practical applications), where each user independently makes power consumption schedule. On one hand, since the aggregated energy demand of all users is subject to the supply capacity from the distribution infrastructure limit (e.g., the thermal limit of transformers and feeders), the energy demand of one user is spatially coupled with those of other users to avoid exceeding the supply capacity. On the other hand, since each user requires that the cumulative power consumption exceeds a threshold by a deadline in order to complete the daily task (e.g., an electric vehicle needs to be charged 16 kWh for next-day 40-mile drive [13], or a dishwasher after lunch should finish washing dishes before dinner), the energy demand at one time slot is temporally coupled with those at other time slots to guarantee that the required demand is satisfied. As a result, the problem is demonstrated to have both spatially- and temporally-coupled constraints which, to the best of our knowledge, has not been thoroughly investigated in the context of demand response.

There has been a large amount of literatures on demand response in smart grid. For instance, a distributed and iterative algorithm has been proposed in [14], [15] which is based on utility maximization and can balance between the real-time supply and demand. However, the temporally-coupled constraint has not been taken into account. In [16]–[18], price uncertainty and game interaction have been introduced into energy consumption scheduling, together with the required demand constraint. The authors, however, have not further considered the spatially-coupled constraint incurred by the supply capacity. Residential load control has been investigated in [19] with price prediction to tradeoff between electricity payment and user waiting time minimization. The work has taken both spatially- and temporally-coupled constraints into consideration, yet the problem is only solved in a centralized manner, using the interior-point method [20]. This method may not be applicable for the practical situation, as the central controller needs to know the exact utility function of each user. However, since such information is private and no user wants to reveal, the central controller may not have sufficient information to solve the problem.

In summary, existing works either address the demand response problem without simultaneously considering the supply capacity and required demand constraints, or solve the problem in a centralized manner without preserving users' privacy. To this end, this paper will take both aspects into account and solve the demand response problem in a comprehensive way. Dual decomposition is a standard method used to address a large problem with complex constraints, by decoupling it into simple subproblems that can be solved locally. This method leads to distributed algorithms, without requiring each user to reveal his/her private information. In addition, the dual decomposition approach is usually combined with the subgradient projection method to solve the corresponding dual problem, which has been widely used to maximize network utility in

communication and networking systems [21]–[23]. We base our distributed demand response problem on the above works, although the problem formulation and applications are quite different. Moreover, through convergence analysis, we notice that the convergence rate of the subgradient projection method is not fast enough, and it is highly dependent on the choice of the step size. To overcome the weakness, we further propose a fast approach which searches for the optimal solution more efficiently. The computational time is polynomially bounded, and we differ from the work in [24] by extending to our doubly constrained demand response problem. Specifically, the contributions of this paper are summarized in the following:

1) Problem formulation: We formulate the demand response problem with both spatially- and temporally-coupled constraints;
2) Distributed algorithm design: We solve the problem in a distributed manner, which is based on dual decomposition and can preserve every user's privacy;
3) Convergence improvement: We further utilize the binary search technique to propose a fast approach, which can speed up the convergence of the distributed algorithm.

The remainder of this paper is organized as follows. We describe the system model and formulate the demand response problem with spatially- and temporally-coupled constraints in Section II. In Section III, we propose a distributed algorithm based on dual decomposition to solve the problem. In Section IV, we further propose a fast approach to increase the convergence rate. Simulations are conducted in Section V, and concluding remarks are drawn in Section VI with future work. We refer to Appendix A for detailed load models, and Appendix B for convergence analysis.

II. System Model and Problem Formulation

A. System Model

Consider a smart distribution grid with an LSE and a set $\mathcal{N} \triangleq \{1, \ldots, N\}$ of users. The cycle of a day is comprised of a set $\mathcal{T} \triangleq \{1, \ldots, T\}$ of (discrete) time slots, e.g., $T = 24$. The household appliances in general lie in two categories of "must-run" and "shiftable" respectively. The must-run appliances include refrigerators that always need to be on during the day, and those that need to be used during the specific time interval, e.g., cooking at cook time, or illumination at night. The shiftable (also called controllable/dispatchable/interruptible/deferrable/flexible/elastic) appliances include electric vehicle, washer, dryer, etc. For these appliances, users are only concerned about whether the task can be finished within a time period, e.g., an electric vehicle needs to be charged 16 kWh for next-day 40-mile drive, or a dishwasher after lunch should finish washing dishes before dinner.

Let x_i^t denote the energy demand of user i at time slot t. The energy demand is bounded by

$$\underline{x}_i^t \leq x_i^t \leq \overline{x}_i^t \quad \forall i, t \tag{1}$$

where $\underline{x}_i^t$ and $\overline{x}_i^t$ denote the lower and upper energy consumption bounds respectively of user i at time slot t. In specific, $\underline{x}_i^t$ represents the "baseline" demand, which is dependent on user

and time, but independent of electricity price. The baseline demand comes from must-run appliances. By contrast, $\overline{x}_i^t$ represents the gross energy consumptions of all the appliances in the house on the maximum power level.

Besides, let e_i denote the "elastic" demand of user i, which is dependent on electricity price. The elastic demand comes from shiftable appliances, which requires that the cumulative power consumption exceeds a threshold by a deadline in order to complete the daily task. If we define the *required demand* from all appliances of user i to be $r_i \triangleq \sum_t \underline{x}_i^t + e_i$. That is, the energy demand at one time slot is temporally coupled with those at other time slots to guarantee that the required demand is satisfied. Thus, we have the *temporally-coupled* constraint, which couples the energy demand over all time slots together:

$$\sum_t x_i^t \geq r_i \quad \forall i \tag{2}$$

Remark 1: Note that the proposed constraints (1) and (2) are in rather general form capable of modeling most, but not all, appliances. A notable exception is to model appliances whose elastic demands need to be satisfied without being interrupted. More detailed load models for different home appliances are elaborated in Appendix A.

Finally, the aggregated electricity consumption of all users is subject to the *supply capacity*, denoted by c^t at time slot t. The supply capacity comes from the distribution infrastructure limit, such as the thermal limit of transformers and feeders. That is, the energy demand of one user is spatially coupled with those of other users to avoid exceeding the supply capacity. Thus, we have the *spatially-coupled* constraint, which couples the energy demand of all users together:

$$\sum_i x_i^t \leq c^t \quad \forall t \tag{3}$$

For ease of presentation, we now define $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ (with entries x_i^t) as the demand matrix. We also define $\underline{\boldsymbol{X}}$ and $\overline{\boldsymbol{X}}$ (with entries $\underline{x}_i^t$ and $\overline{x}_i^t$) as the demand lower and upper bound matrixes respectively. The tth column of $\boldsymbol{X}$, denoted by $\boldsymbol{x}^t \in \mathbb{R}^{N \times 1}$, represents the demand snapshot of all users at time slot t. Similarly, the ith row of $\boldsymbol{X}$, denoted by $\boldsymbol{x}_i \in \mathbb{R}^{1 \times T}$, represents the demand schedule of user i over all time slots. We distinguish between $\boldsymbol{x}^t$ (demand snapshot) and $\boldsymbol{x}_i$ (demand schedule) by their superscript and subscript. In the rest of this work, we also use the following mathematical notations: $(\cdot)^+$ denotes $\max\{\cdot, 0\}$, $\cdot\big|_a^b$ denotes $\min\{\max\{\cdot, a\}, b\}$, $f'(\cdot)$ denotes the first derivative of function $f(\cdot)$, and $f^{-1}(\cdot)$ denotes the inverse of $f(\cdot)$.

B. Problem Formulation

The energy demand of each user varies at different times of the day. Such different behaviors of users are modeled by different choices of utility functions. Formally, with user i at time slot t, we associate an increasing, concave, and differentiable utility function $U_i^t(\cdot)$. The utility derived by energy demand x_i^t is $U_i^t(x_i^t)$, quantifying the user obtained comfort as a function of his/her energy consumption. The quadratic utility function is usually considered, which corresponds to linear decreasing marginal benefit [14], [15].

Let p^t denote the real-time electricity price at time slot t. Thus, the welfare of user i at time slot t is the obtained utility minus the electricity payment: $W_i^t(x_i^t) = U_i^t(x_i^t) - p^t x_i^t$, and the total welfare, of all users and over all time slots, is defined as $W(X) \triangleq \sum_{i,t} W_i^t(x_i^t)$. We define the demand response problem to schedule the energy demand of all users and over all time slots, with the aim of maximizing the total welfare, subject to spatially- and temporally-coupled constraints:

$$\max_X \quad W(X) \tag{4}$$
$$\text{s.t.} \quad (1), (2) \text{ and } (3)$$

Remark 2: The user's welfare function models his behaviors, including his response to various congestion prices. And different users may have different welfare functions. If two users have exactly the same response to a certain given congestion price, they will each curtail or promote the same amount of demand. This makes sense since we assume the continuous energy consumption. Note that the framework proposed in this paper accommodates discrete scenarios, but the solution would become as more complicated, of which the details will be considered as our future work.

III. DUAL DECOMPOSITION APPROACH

Since problem (4) has spatially-coupled constraints (3) and temporally-coupled constraints (2), which couple the energy demand of all users and over all time slots, the problem cannot be directly tackled. However, by means of dual decomposition [20], we can decouple (4) into simple subproblems which can be solved independently. The primal problem and its dual problem are strictly equivalent, due to its strong duality. We derive and solve its dual problem as follows.

A. Dual Problem

Define the (partial) Lagrangian of primal problem (4):

$$\mathcal{L}(X, \lambda, \mu) = \sum_{i,t} \left[W_i^t(x_i^t) - (\lambda^t - \mu_i) x_i^t \right] + \Gamma$$

where we relax the supply capacity constraint (3) by introducing Lagrangian multiplier $\lambda^t \geq 0$ at time slot t, and relax the required demand constraint (2) by introducing Lagrangian multiplier $\mu_i \geq 0$ for user i, while $\lambda \triangleq (\lambda^1, \ldots, \lambda^T), \mu \triangleq (\mu_1, \ldots, \mu_N)$ are the Lagrangian multiplier (dual variable) vectors, $\Gamma \triangleq \sum_t \lambda^t c^t - \sum_i \mu_i r_i$.

The dual function (the objective function of the dual problem) is the supreme of the Lagrangian over the demand matrix X:

$$\mathcal{D}(\lambda, \mu) = \sup_{\underline{X} \leq X \leq \overline{X}} \mathcal{L}(X, \lambda, \mu) = \sum_{i,t} \left[\mathcal{S}_i^t(\lambda^t, \mu_i) \right] + \Gamma$$

where $\mathcal{S}_i^t(\lambda^t, \mu_i)$ is defined as the $(i,t)^{\text{th}}$ subproblem to be solved by user i at time slot t. Note that each subproblem is independent, without spatially- nor temporally-coupled constraints:

$$\mathcal{S}_i^t(\lambda^t, \mu_i) \triangleq \max_{x_i^t} \quad W_i^t(x_i^t) - (\lambda^t - \mu_i) x_i^t \tag{5}$$
$$\text{s.t.} \quad \underline{x}_i^t \leq x_i^t \leq \overline{x}_i^t$$

The subproblem function $\mathcal{S}_i^t(\lambda^t, \mu_i)$ is convex as it is the pointwise supreme of affine functions [20]. Thus the dual function $\mathcal{D}(\lambda, \mu)$ is also convex.

The dual problem is to minimize the dual function over the Lagrangian multiplier vectors λ, μ:

$$\min_{\lambda, \mu \geq 0} \mathcal{D}(\lambda, \mu) \tag{6}$$

The optimal point of primal problem (4) is upper bounded by any feasible value of dual problem (6), i.e., for any $\lambda, \mu \geq 0$ and any feasible X satisfying (1), (2) and (3), we have $\mathcal{D}(\lambda, \mu) \geq W(X)$. From λ^*, μ^*, the optimal solution of dual problem (6), we can get X^*, the optimal solution of primal problem (4), by solving subproblem (5).

B. Subgradient Projection

From the above, solving primal problem (4) is equivalent to solving its dual problem (6). For the differentiable dual function $\mathcal{D}(\lambda, \mu)$, the subgradient projection method can be employed to iteratively calculate the optimal solution of dual problem (6). The Lagrangian multipliers λ^t, μ_i are updated in an opposite direction to the subgradient of the dual function:

$$\begin{cases} \lambda^{t,k+1} = \left[\lambda^{t,k} - \gamma \frac{\partial \mathcal{D}(\lambda^k; \mu)}{\partial \lambda^{t,k}} \right]^+ \\ \mu_i^{k+1} = \left[\mu_i^k - \gamma \frac{\partial \mathcal{D}(\mu^k; \lambda)}{\partial \mu_i^k} \right]^+ \end{cases}$$

where $\gamma > 0$ is the step size to adjust the convergence rate, and $k \in \mathbb{N}^+$ denotes the index of iterations. Convergence towards the optimal solution is guaranteed for a sufficiently small step size, such that the dual function satisfies the Lipchitz continuity condition. The convergence analysis in Appendix B provides guidance on how to choose the step size.

From the above, the first term of the dual function consists of $N \times T$ independent subproblems. In other words, by dual decomposition, the global optimization problem (4) has been decoupled into $N \times T$ local optimization subproblems (5) of each user at each time slot. Different users at different time slots interact through the Lagrangian multipliers λ^t, μ_i.

For each local optimization subproblem (5), given λ^t, μ_i, the local optimal energy demand:

$$\tilde{x}_i^t \triangleq x_i^{t*}(\lambda^t, \mu_i) = \left(W_i^{t\prime} \right)^{-1} (\lambda^t - \mu_i) \Big|_{\underline{x}_i^t}^{\overline{x}_i^t} \tag{7}$$

is unique due to the concavity of $U_i^t(\cdot)$. Under arbitrary λ, μ, the locally optimal solution $X^*(\lambda, \mu)$ may not be globally optimal. However, by the duality theory [20], there exists dual optimal λ^*, μ^* such that $X^*(\lambda^*, \mu^*)$ is the globally optimal solution X^*.

From the above, given λ^*, μ^* from the dual problem (6), each user can calculate (7) independently without the need to coordinate with other users nor time slots. In this sense, the Lagrangian multiplier serves as a coordination signal that aligns local optimality (7) with global optimality (4).

With solution (7) to subproblem (5), the dual function is simplified and the subgradient of the dual function is

$$\begin{cases} \frac{\partial \mathcal{D}(\lambda^k; \mu)}{\partial \lambda^{t,k}} = c^t - \sum_i \tilde{x}_i^t \\ \frac{\partial \mathcal{D}(\mu^k; \lambda)}{\partial \mu_i^k} = \sum_t \tilde{x}_i^t - r_i \end{cases} \tag{8}$$

We obtain the following Lagrangian multiplier update rule:

$$\begin{cases} \lambda^{t,k+1} = \left[\lambda^{t,k} - \gamma \left(c^t - \sum_i \tilde{x}_i^t \right) \right]^+ & \text{(9a)} \\[2em] \mu_i^{k+1} = \left[\mu_i^k - \gamma \left(\sum_t \tilde{x}_i^t - r_i \right) \right]^+ & \text{(9b)} \end{cases}$$

C. Distributed Algorithm

The dual decomposition approach to the demand response problem (4) has two folds. On one hand, each subproblem $\mathcal{S}_i^t (\lambda^t, \mu_i)$ is locally solved, returning daily energy demand schedule $\tilde{x}_i$ of user i. These subproblems are defined by (5) and solved by (7). On the other hand, the dual problem $\mathcal{D}(\lambda, \mu)$ is iteratively solved, returning the dual optimal Lagrangian multipliers λ^*, μ^*. The dual problem is defined by (6) and solved by (9a) and (9b).

The Lagrangian multipliers have the following interpretations. Following the classical concept in economics, λ^t is the *congestion price* at time slot t to leverage the balance between supply and demand. For example, if the total demand of all users $\sum_i \tilde{x}_i^t$ exceeds the supply capacity c^t (aggressive case), the congestion price λ^t will rise (see (9a)) to capture the fact that it is expensive to use electricity at that time slot, which will in turn decrease the demand $\tilde{x}^t$ (see (7)). In such a way the supply and demand will reach balance, and vice versa. Similarly, μ_i is the *coordination parameter* of user i to coordinate his demand schedule, such that the required demand is met. For example, if the total demand over all time slots $\sum_t \tilde{x}_i^t$ is less than the required demand r_i (conservative case), the coordination parameter μ_i will rise (see (9b)), which will in turn increase the demand $\tilde{x}_i$ (see (7)). In such a way the required demand will get satisfied, and vice versa.
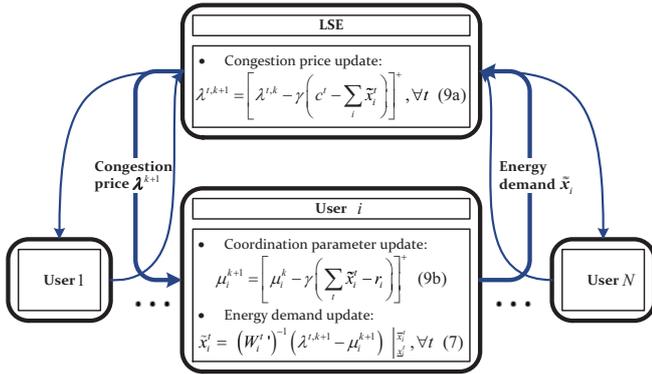

Fig. 1. The illustration of information exchange between LSE and users based on the dual decomposition approach.

Algorithm 1: operated at LSE

---

1 **repeat**
2   receive total demand of all users $\sum_i \tilde{x}_i^t$, $\forall t$;
3   update congestion price λ^{k+1} (9a);
4   send λ^{k+1} to all users;
5 **until** λ converges as $\left| \lambda^{t,k+1} - \lambda^{t,k} \right| < \varepsilon$, $\forall t$;

Algorithm 2: operated at user i

---

1 **repeat**
2   update coordination parameter μ_i^{k+1} (9b);
3   receive congestion price λ^{k+1} from LSE;
4   update energy demand $\tilde{x}_i$ (7);
5   send $\tilde{x}_i$ to LSE;
6 **until** μ_i converges as $\left| \mu_i^{k+1} - \mu_i^k \right| < \varepsilon$;

The interaction between LSE and users based on the dual decomposition approach is illustrated in Fig. 1. The distributed algorithms at LSE and each user are summarized in **Algorithm 1** and **2**, respectively, where ε is the error tolerance (stopping criterion). All the updates (9a), (9b) and (7) are performed based on local information. The congestion price λ is updated at LSE, based on the total demand of all users and the supply capacity. Then the price is announced to each user, aiming to balance between supply and demand. The coordination parameter μ_i is updated at each user, based on the total demand over all time slots and the required demand. Each user also updates his energy demand $\tilde{x}_i$, based on the congestion price and coordination parameter, and feeds it back to LSE. Compared with existing centralized methods, the distributed algorithm does not reveal the exact utility function of each user, and thus preserves their privacy [1].

Remark 3: Note that the proposed algorithm is not totally distributed since LSE has to be involved. In this paper, we nevertheless term it as a "distributed" method, following that in some references (e.g., [14], [15]), to differentiate it from existing centralized methods where the central controller (LSE) needs to know the exact utility function of each user. In the proposed algorithm, instead of transmitting private data to LSE, each user operates local computation and makes his own decision, and then only sends the energy demand decision to LSE.

IV. FAST APPROACH

A. Convergence Analysis

Although the subgradient projection method is a standard way to solve a differentiable convex problem, the convergence rate is not fast enough, and it is highly dependent on the choice of the step size. As analyzed in Appendix B, this method converges for a sufficiently small step size $0 < \gamma < 2/K$, where K is the Lipschitz constant. If the step size is chosen too large, the distributed algorithm will diverge; whereas if it is too small, the convergence rate will be slow. Figure 2 illustrates the convergence of the subgradient projection method, started with $\lambda^1, \mu^1 = 0$. The corresponding parameters are set up according to TABLE I and elaborated in Section V. We vary the step size value from 0.01 to 0.04 to evaluate how it will impact on the convergence rate. The figure indicates that as the step size increases, the algorithm converges fast. But when it

---

[1]Since the users' utility functions could infer their energy usage patterns, they are typically regarded as users' private information [25], [26]. The proposed algorithm protects users' utility functions and hence, helps protect privacy of the users.

exceeds a threshold, the output will oscillate around a certain equilibrium. Therefore, to overcome the weakness, we propose a fast approach based on binary search, which does not depend on the step size and searches for the optimal solution more efficiently.
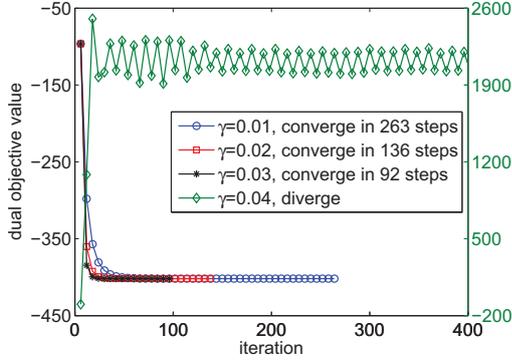

Fig. 2. The impact of the step size on the convergence rate of the subgradient projection method.

Firstly, the Karush-Kuhn-Tucker (KKT) conditions associated with (1)-(4) are [20, Section 5.5.3]

Stationarity:

$$\nabla_{\boldsymbol{X}} \mathcal{L}\left(\boldsymbol{X}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = \boldsymbol{0}$$

i.e.,

$$W_i^{t\prime}\left(x_i^t\right) - \lambda^t + \mu_i = 0 \quad \forall i, t \tag{10}$$

Complementary slackness:

$$\lambda^t\left(c^t - \sum_i x_i^t\right) = 0 \quad \forall t \tag{11}$$

$$\mu_i\left(\sum_t x_i^t - r_i\right) = 0 \quad \forall i \tag{12}$$

Dual feasibility:

$$\lambda^t \geq 0 \quad \forall t$$
$$\mu_i \geq 0 \quad \forall i$$

Primal feasibility:

$$\text{plus (1), (2) and (3)}$$

Consider the solution (7) as a function of λ^t, μ_i. For any selection of λ^t, μ_i, clearly $x_i^{t*}\left(\lambda^t, \mu_i\right)$ satisfies (1) and (10). Hence, to solve original problem (4), we need to find the optimal λ^{t*}, μ_i^* such that $x_i^{t*}\left(\lambda^{t*}, \mu_i^*\right)$ satisfies (2), (3), (11) and (12).

B. Subproblem One

Given $\boldsymbol{\lambda}$, at each user i, define

$$g_i\left(\mu_i; \boldsymbol{\lambda}\right) \triangleq \sum_t x_i^{t*}\left(\mu_i; \lambda^t\right)$$

Then, for (2) and (12), we need to find μ_i^* such that

$$\begin{cases} g_i\left(\mu_i^*; \boldsymbol{\lambda}\right) - r_i \geq 0 \\ \mu_i^*\left[g_i\left(\mu_i^*; \boldsymbol{\lambda}\right) - r_i\right] = 0 \end{cases} \tag{13}$$

Consider two cases as follows:
1) If $g_i\left(0; \boldsymbol{\lambda}\right) > r_i$, then $\mu_i^* = 0$;
2) Else, we need to find $\mu_i^* > 0$ such that $g_i\left(\mu_i^*; \boldsymbol{\lambda}\right) = r_i$.

This is equivalent to the convergence stopping criteria in (9b), i.e., $\mu^{t,k+1}$ converges in the following two cases:
1) If $\sum_t \tilde{x}_i^t = r_i$, then $\mu^{t,k+1} > 0$;
2) If $\sum_t \tilde{x}_i^t > r_i$, then $\mu^{t,k+1} = 0$.

From (7), $x_i^{t*}\left(\mu_i; \lambda^t\right)$ can be expressed as the following piecewise function:

$$x_i^{t*}\left(\mu_i; \lambda^t\right)$$
$$= \begin{cases} \underline{x}_i^t, & \mu_i \leq \underline{\mu}_i^t\left(\lambda^t\right) \triangleq \lambda^t - W_i^{t\prime}\left(\underline{x}_i^t\right) \\ \left(W_i^{t\prime}\right)^{-1}\left(\lambda^t - \mu_i\right), & \text{otherwise} \\ \overline{x}_i^t, & \mu_i \geq \overline{\mu}_i^t\left(\lambda^t\right) \triangleq \lambda^t - W_i^{t\prime}\left(\overline{x}_i^t\right) \end{cases}$$

Clearly, each $x_i^{t*}\left(\mu_i; \lambda^t\right)$ is piecewise and increasing over μ_i, and thus $g_i\left(\mu_i; \boldsymbol{\lambda}\right)$ is the same. A simple example of $g\left(\mu\right)$ is illustrated in Fig. 3 for ease of understanding.
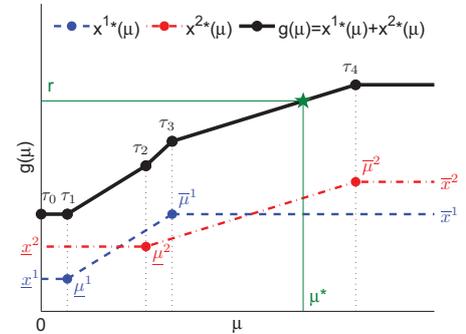

Fig. 3. The illustration of a simple example of $g\left(\mu\right) = x^{1*}\left(\mu\right) + x^{2*}\left(\mu\right)$.

Since there are a number T of $x_i^{t*}\left(\mu_i; \lambda^t\right)$, $t = 1, \ldots, T$, thus the breakpoints of the piecewise and increasing function $g_i\left(\mu_i; \boldsymbol{\lambda}\right)$ occur at the $2T$ points $\underline{\mu}_i^t\left(\lambda^t\right)$ and $\overline{\mu}_i^t\left(\lambda^t\right)$. Let $\tau_1, \ldots, \tau_H$ denote all positive breakpoints, where $H \leq 2T$ and $\tau_0 = 0 < \tau_1 \leq \tau_2 \leq \ldots \leq \tau_H$. We present **Algorithm 3** for solving (13), based on binary search for the bracket between two breakpoints. This algorithm is considered more efficient than (9b), since it searches for μ_i^* in one step, and its computational complexity is polynomially bounded by the time horizon T. Taking Fig. 3 ($T = 2$) for illustration, since there are two of $x^{1*}\left(\mu\right)$ and $x^{2*}\left(\mu\right)$, thus the breakpoints of the piecewise and increasing function $g\left(\mu\right) = x^{1*}\left(\mu\right) + x^{2*}\left(\mu\right)$ occur at the four points $\underline{\mu}^1, \overline{\mu}^1$ and $\underline{\mu}^2, \overline{\mu}^2$. We sort them in an increasing order as $\tau_0 = 0 < \tau_1 = \underline{\mu}^1 < \tau_2 = \underline{\mu}^2 < \tau_3 = \overline{\mu}^1 < \tau_4 = \overline{\mu}^2$, and thus the search range is between τ_0 and τ_4. Through binary search, if we find, for instance, that μ^* lies in the bracket between the two breakpoints τ_3 and τ_4, then it can be directly calculated by solving the equation $\overline{x}^1 + \left(W^{2\prime}\right)^{-1}\left(\lambda^2 - \mu\right) = r$ in one step.

C. Distributed Algorithm

Given $\boldsymbol{\mu}$, at each time slot t, define

$$f^t\left(\lambda^t; \boldsymbol{\mu}\right) \triangleq \sum_i x_i^{t*}\left(\lambda^t; \mu_i\right)$$

Algorithm 3: input $\boldsymbol{\lambda}$ and output μ_i^*

1 **if** $g_i(0;\boldsymbol{\lambda}) > r_i$ **then**
2   **return** 0;
3 **else**
4    $left \leftarrow 0, \; right \leftarrow H$;
5   **repeat**
6       $middle \leftarrow [(left + right)/2]$, where $[\cdot]$ denotes the integer part;
7       $M \leftarrow g_i(\tau_{middle};\boldsymbol{\lambda})$;
8   **if** $M = r_i$ **then**
9   **return** τ_{middle};
10   **else if** $M < r_i$ **then**
11          $left \leftarrow middle$;
12   **else**
13          $right \leftarrow middle$;
14   **end if**
15   **until** $right - left = 1$;
16    $\mathcal{T}_1 \leftarrow \left\{ t \big| \overline{\mu}_i^t(\lambda^t) \leq \tau_{left} \right\}$;
17    $\mathcal{T}_2 \leftarrow \left\{ t \big| \underline{\mu}_i^t(\lambda^t) \geq \tau_{right} \right\}$;
18    $\mathcal{T}_3 \leftarrow \mathcal{T} - \mathcal{T}_1 - \mathcal{T}_2$;
19   **return**
$$\underset{\mu_i > 0}{\arg} \left[\sum_{t \in \mathcal{T}_1} \overline{x}_i^t + \sum_{t \in \mathcal{T}_2} \underline{x}_i^t + \sum_{t \in \mathcal{T}_3} (W_i^{t\prime})^{-1}(\lambda^t - \mu_i) = r_i \right];$$
20 **end if**

Then, for (3) and (11), we need to find λ^{t*} such that

$$\begin{cases} c^t - f^t(\lambda^{t*};\boldsymbol{\mu}) \geq 0 \\ \lambda^{t*}[c^t - f^t(\lambda^{t*};\boldsymbol{\mu})] = 0 \end{cases} \quad (14)$$

Consider two cases as follows:

1) If $f^t(0;\boldsymbol{\mu}) < c^t$, then $\lambda^{t*} = 0$;
2) Else, we need to find $\lambda^{t*} > 0$ such that $f^t(\lambda^{t*};\boldsymbol{\mu}) = c^t$.

This is equivalent to the convergence stopping criteria in (9a), i.e, $\lambda^{t,k+1}$ converges in the following two cases:

1) If $\sum_i \tilde{x}_i^t = c^t$, then $\lambda^{t,k+1} > 0$;
2) If $\sum_i \tilde{x}_i^t < c^t$, then $\lambda^{t,k+1} = 0$.

However, here we still let LSE take the subgradient projection method (9a) to update the congestion price $\boldsymbol{\lambda}$, instead of similarly extending **Algorithm 3** to the side of LSE. The reason is that in that case LSE needs to know the utility function of each user to find λ^{t*} satisfying (14), which may not be applicable for the practical situation. Since such information is private and no user wants to reveal, so LSE may not have sufficient information to operate in that way. Therefore, we still apply the conventional subgradient projection method to the side of LSE, which does not require each user to reveal such private information.

The interaction between LSE and users based on the fast approach is illustrated in Fig. 4. The distributed algorithm at LSE and each user are summarized in **Algorithm 1** and **4**, respectively. All the updates (9a), **Algorithm 3** and (7) are performed based on local information. Specifically, the coordination parameter μ_i is updated at each user, based on the congestion price, the utility function and the required demand.
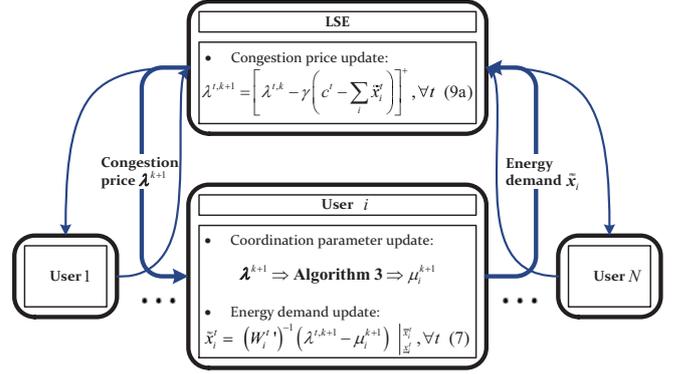

Fig. 4. The illustration of information exchange between LSE and users based on the fast approach.

Algorithm 4: operated at user i

1 **repeat**
2   receive congestion price $\boldsymbol{\lambda}^{k+1}$ from LSE;
3   update coordination parameter μ_i^{k+1} by **Algorithm 3**;
4   update energy demand $\tilde{\boldsymbol{x}}_i$ (7);
5   send $\tilde{\boldsymbol{x}}_i$ to LSE;
6 **until** μ_i converges as $\left| \mu_i^{k+1} - \mu_i^k \right| < \varepsilon$;

V. NUMERICAL RESULTS

We provide numerical examples in this section to illustrate the formulated demand response problem with spatially- and temporally-coupled constraints, and compare the proposed distributed algorithms of dual decomposition and fast approach. We consider a smart distribution grid with $N = 100$ users, and the time horizon is $T = 24$. The must-run appliances (baseline demand) of users are refrigerator-freezer (daily usage: 1.32 kWh), cooking stove (daily usage: 2.01 kWh), television (daily usage: 6 kWh), laptop-desktop computer (daily usage: 13.92 kWh), lighting (daily usage: 1 kWh), and heating (daily usage: 7.1 kWh); while the shiftable appliances (elastic demand) are electric vehicle (daily usage: 16 kWh), dishwasher (daily usage: 1.44 kWh), and clothes washer-dryer (daily usage: 4.44 kWh) [26]. Thus the required demand r_i of user i is chosen randomly from a uniform distribution on $[31.35, 53.23]$. All the following results are obtained by MATLAB R2007b running on a laptop PC with Intel Core i5-3320 CPU @ 2.6 GHz, 4 GB RAM memory, and 32-bit Windows 7 OS.

TABLE I
PARAMETER SETUPS

Parameter	Value	Parameter	Value	Parameter	Value
N	100	r_i	$[31.35, 53.23]$	$\boldsymbol{\lambda}^1, \boldsymbol{\mu}^1$	**0**
T	24	c^t	$[130.6, 221.8]$	γ	0.03
$\underline{x}_i^t$	1	y	$[1.3, 2.2]$	$\overline{x}_i^t$	2.5

The utility function associated with user i at time slot t is quadratic: $U_i^t(x_i^t) \triangleq \begin{cases} -(x_i^t - y)^2 & 0 \leq x_i^t < y \\ 0 & x_i^t \geq y \end{cases}$ [16]–
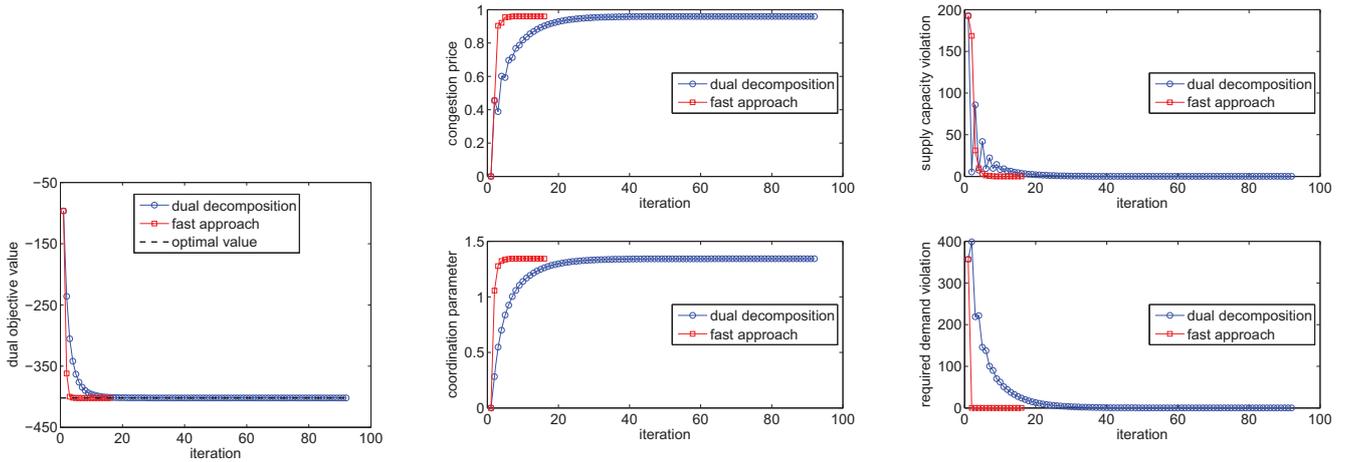
(a) The dual objective value versus iteration. The dashed line shows the optimal value.

(b) The congestion price at a certain time slot (top) and the coordination parameter of a certain user (bottom), versus iteration.

(c) The supply capacity violation (top) and the required demand violation (bottom), versus iteration.

Fig. 5. The comparison between dual decomposition and fast approach to the demand response problem with spatially- and temporally-coupled constraints.

[18], where the target demand y is chosen randomly from a uniform distribution on $[1.3, 2.2]$. We set the lower and upper bounds on demand as $\underline{x}_i^t = 1$ and $\overline{x}_i^t = 2.5$ respectively for all i and t. The hourly-based real-time electricity prices are taken from the real-time pricing program in Illinois, USA on August 4, 2013 [18, Fig. 5(a)]. The supply capacity c^t at time slot t is chosen randomly from a uniform distribution on $[130.6, 221.8]$. For this example, we calculate a valid Lipschitiz constant $K = 149$ for the dual function $\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ using (18) in Appendix B, which implies that the proposed dual decomposition approach will converge as long as the step size is sufficiently small such as $0 < \gamma < 0.0134$. Numerical trials suggest that the algorithm converges for $\gamma \leq 0.037$, and diverges for $\gamma \geq 0.038$. From Fig. 2, we choose $\gamma = 0.03$ to ensure fast convergence rate. The above simulation parameter setups are summarized in TABLE I.

The comparison between dual decomposition and fast approach to the demand response problem with spatially- and temporally-coupled constraints has been detailedly illustrated in Fig. 5. Specifically, Figure 5(a) shows the dual objective value versus iteration, and the optimal value. Figure 5(b) shows the Lagrangian multipliers (the congestion price λ^t and the coordination parameter μ_i), versus iteration. Figure 5(c) shows the supply capacity violation $\sum_t (\mathbf{1} \cdot \boldsymbol{x}^t - c^t)^+$ and the required demand violation $\sum_i (r_i - \boldsymbol{x}_i \cdot \mathbf{1})^+$, versus iteration. It is observed that the fast approach converges much faster than the dual decomposition method. This is because, the subgradient projection method needs to iteratively calculate the optimal $\boldsymbol{\mu}$ to the dual problem (see (9b)); whereas the fast approach searches for the optimal coordination parameter more efficiently in one step (see **Algorithm 3**). Besides, at each iteration, we have a dual feasible point $(\boldsymbol{\lambda}, \boldsymbol{\mu})$; but the corresponding primal point $\boldsymbol{X}$ is generally not feasible. Thus the optimal value of the primal problem is upper bounded by the dual objective value.

Figure 6 shows the total demand of all users $\sum_i x_i^t$, the
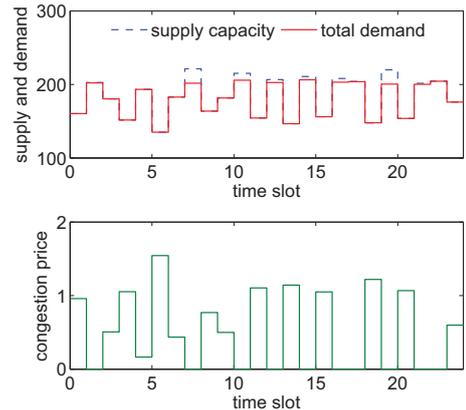

Fig. 6. The supply capacity (top, dashed), the total demand of all users (top, solid), and the congestion price (bottom).
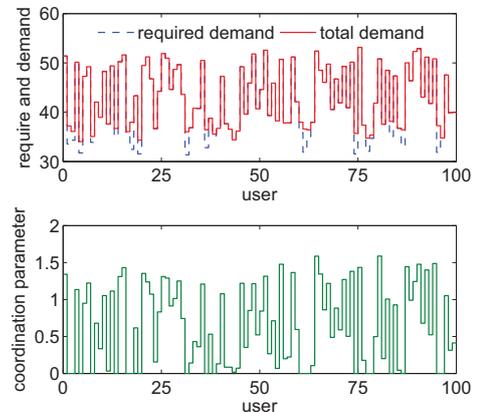

Fig. 7. The required demand (top, dashed), the total demand over all time slots (top, solid), and the coordination parameter (bottom).

supply capacity c^t, and the congestion price λ^t. It is verified that the spatially-coupled constraints are satisfied. The congestion prices are zero whenever the system operates under full capacity. Similarly, Figure 7 shows the total demand over all time slots $\sum_t x_i^t$, the required demand r_i, and the coordination parameter μ_i. It can be seen that the temporally-coupled constraints are satisfied. The coordination parameters are zero whenever the required demand is overqualified. This is due to the complementary slackness conditions (11) and (12), which state that for all inactive constraints the corresponding Lagrangian multipliers should be zero.

VI. CONCLUSION

In this paper, the demand response problem with spatially- and temporally-coupled constraints has been investigated. We first proposed a distributed algorithm adopting dual decomposition to solve the problem, which preserves every user's privacy. Since the convergence rate of the subgradient projection method is not fast enough, and it is highly dependent on the choice of the step size, we further introduced a fast approach which can search for the optimal solution more efficiently. Numerical results were conducted, which verify the theoretical analysis and also demonstrate the outperformance of the fast approach.

As aforementioned in Section II, in this paper we have assumed the user's energy consumption as continuous variables. To be more realistic, we will extend the former setting to the scenario of appliances with discrete power levels. This, however, will lead the problem to be formulated into a mixed integer optimization problem. Getting its solution would become much more complicated. Better formulation of the problem and more efficient algorithm design specifically for discrete scenarios therefore will be of our future research interest.

APPENDIX A
DETAILED LOAD MODELS

The specific load models of commonly-used electric appliances in households are introduced in the followings.

1) Type 1: Let $\mathcal{A}_1$ denote must-run appliances like lighting or cooking that must be on for a certain period of time. For an appliance $a \in \mathcal{A}_1$, let $\mathcal{T}_a \triangleq [\alpha_a, \beta_a]$ denote the time slots that the appliance must run. Such appliances have their strictly defined constraints:

$$\begin{cases} x_a^t \equiv b_a^t & \forall t \in \mathcal{T}_a \\ x_a^t = 0 & \text{otherwise} \end{cases}$$

where b_a^t denotes the baseline demand that the appliance a must consume at time slot t.

2) Type 2: Let $\mathcal{A}_2$ denote shiftable appliances, e.g., electric vehicles, for which the users only concern whether the task can be finished within a certain time period. For an appliance $a \in \mathcal{A}_2$, let $\mathcal{T}_a \triangleq [\alpha_a, \beta_a]$ denote its working time slot, where β_a is the deadline for the appliance a to be finished. Such appliances are subject to the following constraints:

$$\begin{cases} 0 \le x_a^t \le \overline{x}_a^t & \forall t \in \mathcal{T}_a \\ x_a^t = 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \sum_{t=\alpha_a}^{\beta_a} x_a^t \ge e_a$$

where $\overline{x}_a^t$ denotes the maximum energy level that the appliance a can consume in time slot t, and e_a denotes the elastic demand that the appliance a requires to finish a given task. Setting $\overline{x}_a^t = 0$ for $\forall t \in \mathcal{T} \backslash \mathcal{T}_a$, these constraints can be simplified as

$$\begin{cases} 0 \le x_a^t \le \overline{x}_a^t & \forall t \in \mathcal{T} \\ \sum_{t \in \mathcal{T}} x_a^t \ge e_a \end{cases}$$

The proposed constraints (1) and (2) are in general form for types 1 and 2. Since $x_i^t \triangleq \sum_{a \in \mathcal{A}} x_a^t$, we have that for $\underline{x}_i^t \triangleq \sum_{a \in \mathcal{A}_1} x_a^t$, $\overline{x}_i^t \triangleq \underline{x}_i^t + \sum_{a \in \mathcal{A}_2} \overline{x}_a^t$; since $e_i \triangleq \sum_{a \in \mathcal{A}_2} e_a$, there exists the relationship that $\sum_t \overline{x}_i^t \ge \sum_t \underline{x}_i^t + e_i$.

3) Type 3: Let $\mathcal{A}_3$ denote the special subset of shiftable appliances, like washer and drier, whose elastic demands need to be satisfied without being interrupted. Similar to that for type 2 appliances, for an appliance $a \in \mathcal{A}_3$, setting $\underline{x}_a^t = \overline{x}_a^t = 0$ for $\forall t \in \mathcal{T} \backslash \mathcal{T}_a$, we have

$$\begin{cases} \underline{x}_a^t \le x_a^t \le \overline{x}_a^t & \forall t \in \mathcal{T} \\ \sum_{t \in \mathcal{T}} x_a^t \ge e_a \end{cases}$$

Since the appliance is non-interruptible, the task should be finished within a consecutive time period. Let r_a^t denote the remaining demand that the appliance a still requires at the beginning of time slot t. We have,

$$r_a^t = \begin{cases} e_a & t = 1 \\ e_a - \sum_{\tau=1}^{t-1} x_a^\tau & t = 2, \ldots, T \end{cases}$$

At the beginning of time slot t, if the task has not started yet, i.e., $r_a^t = e_a$, then the appliance can choose to wait or start. Conversely, if the task has started, i.e., $r_a^t < e_a$, then the appliance must continue working until completing the task, i.e., $r_a^t = 0$. That is, we need to set additional constraints [27], [28]:

$$\begin{cases} \underline{x}_a^t \ge 0 & r_a^t = e_a \\ \underline{x}_a^t > 0 & 0 < r_a^t < e_a \\ \underline{x}_a^t = 0 & r_a^t = 0 \end{cases}$$

However, these constraints cannot be trivially generalized to the proposed constraints (1) and (2), since the condition r_a^t is dependent on the variable x_a^t. Handling the non-interruptible constraints therefore essentially resorts to a dynamic programming problem. We do not include such a special case in this paper, but will further consider it in our future work.

B CONVERGENCE ANALYSIS

Here, the convergence of the distributed algorithm is analyzed, which will provide guidance on how to choose the step size. As we know, a standard result is that the subgradient projection method converges for a sufficiently small step size $0 < \gamma < 2/K$, where K is the Lipschitz constant [29]. Thus

in the following we derive a valid Lipschitz constant for the dual function $\mathcal{D}(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

We first define a single Lagrangian multiplier $\boldsymbol{\nu} \triangleq (\boldsymbol{\lambda}, \boldsymbol{\mu})$. From (8), thus,

$$\begin{cases} \nabla_{\lambda^t} \mathcal{D}(\boldsymbol{\nu}) = c^t - \mathbf{1}^{1 \times N} \cdot \boldsymbol{x}^{t*}(\boldsymbol{\nu}) \\ \nabla_{\mu_i} \mathcal{D}(\boldsymbol{\nu}) = \boldsymbol{x}_i^*(\boldsymbol{\nu}) \cdot \mathbf{1}^{T \times 1} - r_i \end{cases}$$

By construction of $\nabla \mathcal{D}(\boldsymbol{\nu})$, we have

$$\begin{aligned} &\|\nabla \mathcal{D}(\boldsymbol{\nu}_1) - \nabla \mathcal{D}(\boldsymbol{\nu}_2)\|_2 \\ &\leq \left(\sqrt{N} + \sqrt{T}\right) \|\boldsymbol{X}^*(\boldsymbol{\nu}_1) - \boldsymbol{X}^*(\boldsymbol{\nu}_2)\|_F \end{aligned} \tag{15}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm.

Let $\rho_i^t \triangleq \lambda^t - \mu_i$ denote the combined Lagrangian multiplier for user i at time slot t, and $\boldsymbol{\rho} \in \mathbb{R}^{N \times T}$ (with entries ρ_i^t) is the combined Lagrangian multiplier matrix. We have $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2$ corresponding to $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2$. From (7), thus,

$$\boldsymbol{X}^*(\boldsymbol{\nu}) = (\boldsymbol{W}')^{-1}(\boldsymbol{\rho}) \Big|_{\underline{\boldsymbol{X}}}^{\overline{\boldsymbol{X}}}$$

If we define $\underline{\rho}_i^t \triangleq W_i^{t\prime}(\overline{x}_i^t)$, $\overline{\rho}_i^t \triangleq W_i^{t\prime}(\underline{x}_i^t)$, and $V_i^t(\cdot) \triangleq (W_i^{t\prime})^{-1}(\cdot)$, thus,

$$\begin{aligned} &\|\boldsymbol{X}^*(\boldsymbol{\nu}_1) - \boldsymbol{X}^*(\boldsymbol{\nu}_2)\|_F \\ &\leq \max_{i,t} \left\{ \left| V_i^{t\prime}\left(\underline{\rho}_i^t\right) \right|, \left| V_i^{t\prime}\left(\overline{\rho}_i^t\right) \right| \right\} \|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_F \end{aligned} \tag{16}$$

Finally, due to $\boldsymbol{\rho} = \mathbf{1}^{N \times 1} \cdot \boldsymbol{\nu}[1:T] - \boldsymbol{\nu}[T+1:T+N] \cdot \mathbf{1}^{1 \times T}$, we have

$$\|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\|_F \leq 2 \max\left\{\sqrt{N}, \sqrt{T}\right\} \|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\|_2 \tag{17}$$

Combining inequalities (15), (16) and (17), there exists

$$\|\nabla \mathcal{D}(\boldsymbol{\nu}_1) - \nabla \mathcal{D}(\boldsymbol{\nu}_2)\|_2 \leq K \|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\|_2$$

where the Lipschitz constant K for the dual function is

$$\begin{aligned} K = &2\left(\sqrt{N} + \sqrt{T}\right) \times \max\left\{\sqrt{N}, \sqrt{T}\right\} \\ &\times \max_{i,t}\left\{ \left| V_i^{t\prime}\left(\underline{\rho}_i^t\right) \right|, \left| V_i^{t\prime}\left(\overline{\rho}_i^t\right) \right| \right\} \end{aligned} \tag{18}$$

REFERENCES

[1] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.

[2] ——, "A survey on smart grid potential applications and communication requirements," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28–42, 2013.

[3] A. I. Sabbah, A. El-Mougy, and M. Ibnkahla, "A survey of networking challenges and routing protocols in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 210–221, 2014.

[4] R. Deng, S. Maharjan, X. Cao, J. Chen, Y. Zhang, and S. Gjessing, "Sensing-delay tradeoff for communication in cognitive radio enabled smart grid," in *Proc. IEEE SmartGridComm*, 2011, pp. 155–160.

[5] R. Deng, J. Chen, X. Cao, Y. Zhang, S. Maharjan, and S. Gjessing, "Sensing-performance tradeoff in cognitive radio enabled smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 302–310, 2013.

[6] R. Deng and Z. Yang, "Cooperative transmission game for smart grid communication," in *Proc. IEEE ICCA*, 2014, pp. 314–319.

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

[8] F. De Angelis, M. Boaro, D. Fuselli, S. Squartini, F. Piazza, and Q. Wei, "Optimal home energy management under dynamic electrical and thermal constraints," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1518–1527, 2013.

[9] F. Kennel, D. Gorges, and S. Liu, "Energy management for smart grids with electric vehicles based on hierarchical MPC," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1528–1537, 2013.

[10] T. Facchinetti and M. L. Della Vedova, "Real-time modeling for direct load control in cyber-physical power systems," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 689–698, 2011.

[11] H. Kumar Nunna and S. Doolla, "Responsive end-user based demand side management in multi-microgrid environment," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1262–1272, 2014.

[12] N. Rahbari-Asr and M.-Y. Chow, "Cooperative distributed demand management for community charging of PHEV/PEVs based on KKT conditions and consensus networks," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 3, pp. 1907–1916, 2014.

[13] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power and Energy Magazine*, vol. 7, no. 2, pp. 52–62, 2009.

[14] P. Samadi, A. Mohsenian-Rad, R. Schober, V. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *Proc. IEEE SmartGridComm*, 2010, pp. 415–420.

[15] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *Proc. IEEE PES General Meeting*, 2011, pp. 1–8.

[16] R. Deng, Z. Yang, and J. Chen, "Load scheduling with price uncertainty and coupling constraints," in *Proc. IEEE PES General Meeting*, 2013, pp. 1–5.

[17] R. Deng, Z. Yang, J. Chen, N. R. Asr, and M.-Y. Chow, "Residential energy consumption scheduling: A coupled-constraint game approach," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1340–1350, 2014.

[18] R. Deng, Z. Yang, J. Chen, and M.-Y. Chow, "Load scheduling with price uncertainty and temporally-coupled constraints in smart grids," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2823–2834, 2014.

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

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[21] E. Wei, A. Ozdaglar, A. Eryilmaz, and A. Jadbabaie, "A distributed newton method for dynamic network utility maximization with delivery contracts," in *Proc. IEEE CISS*, 2012, pp. 1–6.

[22] S. He, J. Chen, D. K. Yau, and Y. Sun, "Cross-layer optimization of correlated data gathering in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 11, pp. 1678–1691, 2012.

[23] R. Deng, Y. Zhang, S. He, J. Chen, and X. S. Shen, "Globally optimizing network utility with spatiotemporally-coupled constraint in rechargeable sensor networks," in *Proc. IEEE Globecom*, 2013, pp. 4810–4815.

[24] Y.-S. Fu and Y.-H. Dai, "Improved projected gradient algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds," *Asia-Pacific Journal of Operational Research*, vol. 27, no. 1, pp. 71–84, 2010.

[25] S. Caron and G. Kesidis, "Incentive-based energy consumption scheduling algorithms for the smart grid," in *Proc. IEEE SmartGridComm*, 2010, pp. 391–396.

[26] A. Mohsenian-Rad, V. W. 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 Transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.

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

[28] P. Samadi, H. Mohsenian-Rad, V. W. Wong, and R. Schober, "Tackling the load uncertainty challenges for energy consumption scheduling in smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1007–1016, 2013.

[29] D. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.