Optimal Price-Controlled Demand Response with Explicit Modeling of Consumer Preference Dynamics*

Bowen Zhang[1], Michael C. Caramanis[2], and John Baillieul[3]

***Abstract*— This paper describes a new approach to modeling consumers' utility preferences in price-controlled demand systems, such as the demand for mobile service or electric energy. Relaxing the assumption that the consumer group size is infinite and individual utility preferences are static, we explicitly model their dynamics through a non-uniform and time varying probability distribution characterized by a well-defined dynamically changing parameter. This parameter is embedded into a stochastic dynamic programming problem used to solve for the optimal price policy. An analytic characterization of the optimal policy is derived based on the differential cost function which leads to an assisted value iteration approach that reduces computational complexity. Numerical results are provided to verify and elaborate that optimal policies conform to claims established in rigorous analytic investigations.**

I. Introduction

The increasing penetration of renewable energy, such as wind and solar, makes electricity generation less controllable due to the intermittent nature of these resources. One way to stabilize the grid, in addition to acquiring additional reserves from conventional generators, is to allow consumers to participate into the regulation reserve market and provide real time regulation service reserves (RSR). Extensive literature has been published on the promise of demand management including direct load control of thermostatic appliances [1], [2], [3], decentralized scheduling of vehicle-to-grid integration [4], [5], and the optimal coordination of flexible loads in a micro-grid [6].

This paper introduces explicit modeling of dynamic smart building occupant preferences that influence the provision of RSRs. It develops a tractable stochastic dynamic programming (DP) problem for the minimum cost provision of RSRs that includes these dynamic preferences in its state dynamics. Broadly speaking, it also contributes to the real time price demand control literature by pioneering the relaxation of the usual assumption that demand for service is adequately represented in the short term by a static probability distribution of preferences from large group of consumers [7], [8], [9], [10]. For example, the demand for an alternative mobile service provider's bandwidth, or the demand for turning on your cooling appliance, has been assumed routinely as uniformly distributed across some closed set that remains unchanged regardless of whether the mobile service or energy provider has been broadcasting repeatedly a high price in the recent past or whether the room temperature is close to or far from the thresholds that a cooling zone occupant has selected to represent its comfort temperature zone. Undoubtedly, this assumption is restrictive and inaccurate, particularly when the number of users is finite as is the case with duty cycle appliances in a smart building or cell phones trying to connect to a specific base station.

In this work we explicitly model the dynamic nature of specific appliance preferences and their short term evolution in response to past control and state trajectories. In other words, we consider demand preferences to be dynamic and include them in the system state. Focusing for simplicity of exposition on a multiple cooling appliance smart building load, we use a dynamic probability distribution to represent cooling zone occupant preferences to transition their cooling appliance from an idle to an active state. We account for the fact that a sustained broadcast of high prices that has discouraged turning on idle cooling appliances will increase the likelihood that the temperature of a typical cooling zone will be high and raise the occupant's preference to turn on an idle appliance and commence cooling. The opposite is naturally true after a sustained period of broadcasting low prices that encourages idle appliances to turn on. This dynamic probability distribution can be reasonably characterized by a single dynamically changing parameter that we embed in the state space and dynamics of the DP problem. Although modeling dynamic preferences leads to non-concavity of the expected utility, we provide an analytical expression of the locally optimal policy and show that it is globally optimal. This analytical optimal policy is then used to design an assisted value iteration (AVI) algorithm that reduces the solution time of the standard DP. The AVI algorithm also allows a continuous control space avoiding the need to resort to computationally undesirable control space discretization in order to obtain numerical solutions.

The paper proceeds as follows. Sec.II formulates the problem where we develop the state dynamics, period cost, the time varying dynamic preferences, and the relevant Bellman equation. Sec.III compares the consumers' utility realization in the traditional time invariant model and our time varying utility model. It further relates analytically the optimal policy with the differential cost function. Sec.IV proposes an assisted value iteration approach and compares it to the standard approach. Sec.V concludes.

*The authors gratefully acknowledge support of the U.S. National Science Foundation under EFRI Grant 1038230.

[1]B. Zhang is with the Division of Systems Engineering, Boston University, Boston, MA, 02215 USA e-mail: bowenz@bu.edu.

[2]M.C. Caramanis is with the Department of Mechanical Engineering and the Division of Systems Engineering, Boston University, Boston, MA, 02215 USA e-mail: mcaraman@bu.edu.

[3]J. Baillieul is with the Department of Electrical and Computer Engineering, the Department of Mechanical Engineering, and the Division of Systems Engineering, Boston University, Boston, MA, 02215 USA e-mail: johnb@bu.edu.

II. Problem Formulation

We consider an advanced energy management building with N cooling appliances. The smart building operator (SBO) has contracted to regulate in real-time its electricity consumption within an upper and a lower limit $\{\bar{n}-R, \bar{n}+R\}$ agreed upon in the hour-ahead market, where $\bar{n}$ is the constant energy rate that the SBO purchased in the hour ahead market and R is the maximum up or down RSR that the SBO promised to provide. Moreover, the SBO has assumed the responsibility to modulate its energy consumption to track $\bar{n}+y(t)R$ with $y(t) \in [-1,+1]$ specified by the ISO in almost real time as the RSR signal (usually every 2 or 4 seconds). To this end the SBO broadcasts a real-time price signal $\pi(t)$ to all cooling appliances in order to modulate the number of connected appliances and hence the resulting aggregated power consumption. Appliances respond according to their utility function $U(T)$ for cooling service which depends on their current cooling zone temperature $T \in [T_{\min}, T_{\max}]$. We assume $U(T)$ is a monotonic increasing function of T. The variables $T_{\min}$ and $T_{\max}$ specify the threshold temperature value of the comfort zone. Denote $u(t)$ the threshold temperature value obtained by solving $U(u(t)) = \pi(t)$, idle appliances with surrounding temperature $T \geq u(t)$ will consider to connect. Since the mapping between $\pi(t)$ and $u(t)$ is bijective, for the rest of the paper we use $u(t)$ as the control policy. Deficient ISO RSR signal tracking penalties and occupant utility realizations constitute period costs. The objective is to find a state feedback optimal policy that minimizes the associated infinite horizon discounted cost. Individual cooling zone preferences are modelled by a dynamically evolving probability distribution of idle-appliance-zone temperatures $p_t(T)$. We proceed to derive the system dynamics and formulate the period cost of the relevant dynamic programming problem.

A. State Dynamics

At each time t, the state variables contain the number of connected appliances $i(t)$, the ISO RSR signal value $y(t)$, the direction of the RSR signal $D(t)$, and $\hat{T}(t)$ that fully characterizes $p_t(T)$. Queues $i(t)$ and $N-i(t)$ constitute a closed queueing network where the service rate of one queue determines the arrival rate into the other. Queue $N-i(t)$ behaves like an infinite server queue with each server exhibiting a stochastic Markov modulated service rate that depends on the control $u(t)$ and the probability distribution $p_t(T)$. Queue $i(t)$ also behaves as an infinite server queue with each server exhibiting a constant service rate μ. The dynamics of $y(t)$ and the dependent state variable $D(t) = \text{sgn}(y(t) - y(t-\tau_y))$ are characterized by transitions taking place in short but constant time intervals, τ_y[1], resulting in $y(t)$ staying constant, increasing or decreasing by a typical amount of $\Delta y = \tau_y/150sec$ [11]. These transitions are outputs of a proportional integral filter employed in practice by ISOs to convert system frequency deviations and Area Control Error (ACE) to Regulation Reserve Signals broadcasted to RSR providers. Since frequency deviation and ACE are arguably white noise processes resulting from the stochastic imbalance between demand and supply, $y(t)$ is a Markovian random variable. We can therefore approximate $y(t)$ by a continuous time jump Markovian process that allows us to uniformize the DP problem formulation. To uniformize the DP problem we introduce a control update period of $\Delta_t << \tau_y$ which assures that during the period Δ_t, the probability that more than one event can take place is negligible. We further set the time unit so that $\Delta_t = 1$, and scale transition rate parameters accordingly. The following state dynamics follow.

1) Dynamics of $y(t)$: The transition probabilities of the discrete time Markov process $y(t)$ depend on $y(t)$ and $D(t)$. Statistical analysis on historical PJM data on $y(t)$ trajectories indicate a week dependence on $y(t)$ yielding the reasonable approximation

$$\begin{cases} \text{Prob}(y(t+\tau_y) = y(t) + \Delta y | D(t) = 1) = 0.8 \\ \text{Prob}(y(t+\tau_y) = y(t) - \Delta y | D(t) = 1) = 0.2 \\ \text{Prob}(y(t+\tau_y) = y(t) - \Delta y | D(t) = -1) = 0.8 \\ \text{Prob}(y(t+\tau_y) = y(t) + \Delta y | D(t) = -1) = 0.2 \end{cases}.$$

Denoting by γ_1^u (γ_1^d) the rate at which $y(t)$ will jump up by Δy during a control update period when $D(t) = 1$ ($D(t) = -1$), and by γ_2^u (γ_2^d) the corresponding rate that $y(t)$ will jump down when $D(t) = 1$ ($D(t) = -1$), we have the following uniformized dynamics of $y(t)$ for $\Delta_t = 1$ after time rescaling from τ_y to $\Delta_t = 1$

$$\begin{cases} \text{Prob}(y(t+1) = y(t) + \Delta y | D(t) = 1) = \gamma_1^u \\ \text{Prob}(y(t+1) = y(t) - \Delta y | D(t) = 1) = \gamma_2^u \\ \text{Prob}(y(t+1) = y(t) - \Delta y | D(t) = -1) = \gamma_2^d \\ \text{Prob}(y(t+1) = y(t) + \Delta y | D(t) = -1) = \gamma_1^d \end{cases}.$$

Since $\Delta_t << \tau_y$, we will have very small value of the above rates compared with RSR rate change without normalization.

2) Dynamics of $i(t)$: The dynamics of $i(t)$ is governed by the following arrival and the departure rates.

The arrival rate $a(t)$ depends on the policy $u(t)$. Denote by $p_{u(t)}$ the proportion of idle appliances with cooling zone temperature $T \geq u(t)$. Since idle appliances observe the price broadcast by the SBO at a rate λ and decide to connect and resume cooling when the price is smaller than their utility for cooling at time t, the arrival rate into $i(t)$ is

$$a(t) = [N - i(t)]\lambda p_{u(t)} = (N - i(t))\lambda \int_{u(t)}^{T_{\max}} p(T)dt. \quad (1)$$

This says that $a(t)$ equals the product of the number of idle appliances that observe the broadcast price times the probability that $T \geq u(t)$.

The departure rate $d(t)$ is independent of $u(t)$, It equals the product of active appliances times the inverse of the average duration of a cooling cycle. Modelling the cooling cycle duration as an exponential random variable with rate μ such that $1/\mu$ equals the average energy consumption per energy packet transaction [3], we have

$$d(t) = i(t)\mu. \quad (2)$$

[1] this varies across ISOs. In PJM it is either 2 or 4 seconds depending on the type of regulation service offered

The stochastic dynamics of $i(t)$ in the homogenized model is thus given by $i(t+1) = i(t) + \tilde{i}$ where the random variable $\tilde{i}$ satisfies the following probability relations

$$\begin{cases} p(\tilde{i}=1) = a(t) \\ p(\tilde{i}=-1) = d(t) \\ p(\tilde{i}=0) = 1-a(t)-d(t)-\gamma \end{cases},$$

where $\gamma = \gamma_1^u + \gamma_2^u = \gamma_1^d + \gamma_2^d$ is the total probability that the ISO RSR signal will change in one interval.

3) Dynamics of $D(t)$: Recalling that $D(t) = \text{sgn}(y(t) - y(t-1))$, it is clear that the dynamics of $D(t)$ are fully determined by the dynamics of $y(t)$, namely $D(t)$ equals to the directional changes of $y(t)$. We next argue that the dynamics of $\hat{T}(t)$ are also determined by $y(t)$.

4) Dynamics of $\hat{T}(t)$: By means of extensive simulation reported below, the feature parameter characterizing the consumers' dynamic preferences, $\hat{T}(t)$, is shown to be a linear function of $y(t)$ over a reasonable range of inputs.

Based on related literature [12] and [13], we use standard energy transfer relations to simulate the dynamics of the frequency histogram of idle appliance cooling zone temperatures, which appear to conform to a three parameter functional representation $p_t(T) = f(\hat{T}(t), T_{\min}, T_{\max})$. We use two RSR signal trajectories and associate them with a reasonable RSR tracking policy to simulate the corresponding dynamics of idle appliance temperature trajectories. The first is a standard ISO RSR signal trajectory that aspiring RSR market participants must demonstrate that they have the ability to track. This is referred to in the PJM manual as the standard T-50 qualifying test [11]. The second is an actual historical RSR signal downloaded from the PJM web site [14]. We record the temperature levels prevailing across the N cooling zones when a control trajectory is applied that results in near-perfect tracking of the ISO RSR requests implied by the aforementioned two signals. Simulation results indicate that the time evolution of the probability distribution of cooling zone temperatures conforms to a dynamically changing trapezoid characterized fully by $\hat{T}(t)$; see Fig. 1.

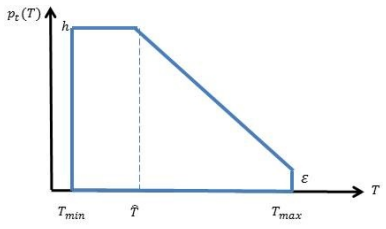

Fig. 1. Trapezoid probability distribution function $p(T)$ with $T \in [T_{\min}, T_{\max}]$ parametrized by a single parameter $\hat{T}$. Height of the trapezoid is $h = 2/(T_{\max} + \hat{T} - 2T_{\min})$.

Fig. 2 shows the accuracy of using a trapezoid function to model the dynamic frequency distribution of idle appliance cooling zone temperatures. We discretize temperature into 20 intervals and perform a Monte Carlo simulation involving 16000 appliances to generate a relatively smooth frequency distribution. The on and off duty sub-cycles are both assumed to be 10 minutes long. The idle appliance price-detection-rate is assumed to be 1 detection per minute. PJM's RSR signal is broadcast every 4 seconds. In Fig. 2, blue curves represent the simulated probability distributions as observed at different times. Trapezoids with $\hat{T}(t) = 5$ and $\hat{T}(t) = 6$ are shown in the upper and the lower figure, respectively. The red curve represents the average of the blue curves at the same temperature. The red curve is then approximated by an exact trapezoid shown by the green curve in Fig. 1. Note that $\hat{T}(t)$ and the trapezoid approximation are the mean statistics of the observed frequency distributions. Note that for a fixed ε, the trapezoids are completely specified by two static quantities, $T_{\min}$ and $T_{\max}$, and the time varying parameter $\hat{T}(t)$.

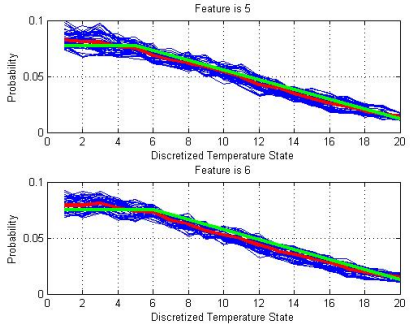

Fig. 2. Monte Carlo simulation is used to determine the probability distribution of idle appliances' temperature. We observe that mean value of these curves conform to a time varying trapezoid shape featured by $\hat{T}(t)$.

Based on the observed time series of $y(t)$ and $\hat{T}(t)$, we find a significant negative correlation in both simulations with values -0.9833 and -0.8106, respectively. The simulation results support the reasonableness of using the following relationship between $\hat{T}(t)$ and $y(t)$:

$$\hat{T}(t) = \alpha_0 + \alpha_1 y(t) + \omega, \tag{3}$$

where α_0 corresponds to the value of $\hat{T}(t)$ when the building's energy consumption level is $\bar{n}$, $\alpha_1 < 0$ is the sensitivity of $\hat{T}$ to the change in the normalized RSR signal $y(t)$, and ω is a zero mean symmetrically distributed error. These findings support our a priori expectation that $y(t)$ is a reasonable sufficient statistic of past state and control trajectories in the information vector available at time t. This a priori expectation is based on the fact that $y(t)$ levels are in fact integrators of recent price control trajectories.

Remark 1 The intuition behind the trapezoid distribution is as follows: The magnitude of the elbow point (temperature) $\hat{T}$ is negatively correlated with the RSR signal $y(t)$. When a consumers temperature is below $\hat{T}$, they have no incentive to have their cooling appliance leave the idle state. This results in a flat distribution $p(T)$ for $T < \hat{T}$. For $T > \hat{T}$, on the other hand, consumers are increasingly likely to have their cooling appliances leave the idle state. It is interesting to note that our simulation data shows $p(T)$ to be a decreasing linear function for $T > \hat{T}$. It remains to be seen whether the idealization of the trapezoid characterization of consumer

utility remains valid as parameters such as cooling appliance duty cycle are varied.

B. Period Cost

The period cost rate consists of two parts: a penalty for inaccurate ISO RSR signal tracking and the utility realized by connected appliances. The tracking penalty is defined as

$$g(i(t),y(t)) = K[\frac{i(t)-\bar{n}-y(t)R}{R}]^2, \tag{4}$$

where K is the penalty per unit of inaccurate tracking. Defining $\kappa = K/R^2$ we can write the tracking penalty rate as

$$g(i(t),y(t)) = \kappa[i(t)-\bar{n}-y(t)R]^2, \tag{5}$$

The expected utility rate realized by an idle cooling appliance zone occupant who decides to resume cooling by paying $\pi(t)$ corresponding to threshold temperature $u(t)$ is

$$U_u = \frac{\int_{u(t)}^{T_{\max}} U(T)p(T)dT}{\int_{u(t)}^{T_{\max}} p(T)dT}. \tag{6}$$

Noting that the probability that an idle appliance will decide to resume cooling is $a(t)$, the expected realized utility is

$$\begin{aligned} a(t)U_u &= (N-i(t))\lambda\, p_u U_u, \\ &= (N-i(t))\lambda \int_{u(t)}^{T_{\max}} P(T)dt \frac{\int_{u(t)}^{T_{\max}} U(T)p(T)dT}{\int_{u(t)}^{T_{\max}} p(T)dT}, \\ &= (N-i(t))\lambda \int_{u(t)}^{T_{\max}} U(T)p(T)dT. \end{aligned} \tag{7}$$

Equations (5) and (7) imply that the total period cost rate is

$$\begin{aligned} c(i(t),y(t),u(t)) &= \kappa[i(t)-\bar{n}-y(t)R]^2 - \\ &\quad (N-i(t))\lambda \int_{u(t)}^{T_{\max}} U(T)p(T)dT. \end{aligned} \tag{8}$$

C. Bellman Equation

The state variables can be grouped according to their dependence on $u(t)$: $i(t)$ depends explicitly on $u(t)$. $\hat{T}(t)$ is also dependent on the past trajectory of controls, but, to the extent that this trajectory is consistent with a reasonable tracking the ISO RSR signal, it can be considered as a function of $y(t)$, which, as discussed earlier, is the sufficient statistic of this trajectory. We can thus consider all state variables, other than $i(t)$, to have dynamics that do not depends on $u(t)$. For notation simplicity we let $\bar{i}^u(t) = \{y(t), D(t)=1, \hat{T}(t)\}$ ($\bar{i}^d(t) = \{y(t), D(t)=-1, \hat{T}(t)\}$) to be the state variables that make up the complement of $i(t)$ when the RSR signal is going up (down), so that $\{i(t), \bar{i}^{u(d)}(t)\}$ ($\{i(t), \bar{i}^d(t)\}$) is the representation of the full state vector when the RSR signal goes up (down). Given the cost function and dynamics described above, we can formulate an infinite horizon discounted cost problem with the following Bellman equation for states including $D(t)=-1$.

$$\begin{aligned} J(i,\bar{i}^d) &= \min_{u\in[T_{\min},T_{\max}]} \{g(i,\bar{i}^d) - a(t)U_u \\ &\quad +\alpha[a(t)J(i+1,\bar{i}^d) + d(t)J(i-1,\bar{i}^d) \\ &\quad +\gamma_1^d J(i,\bar{i}^u+\Delta y) + \gamma_2^d J(i,\bar{i}^d-\Delta y) \\ &\quad +(1-a(t)-d(t)-\gamma_1^d-\gamma_2^d)J(i,\bar{i}^d)]\}. \end{aligned} \tag{9}$$

$J(i,\bar{i}^d)$ is the value function satisfying the Bellman equation, with α denoting the discount factor. For notational simplicity we denote by $\bar{i}^u + \Delta y$ the new state realized when the regulation signal increases from $y(t)$ to $y(t+1) = y(t)+\Delta y$ rendering $D(t+1)=1$, while the rest of the state variables remain unchanged. Similarly we denote by $\bar{i}^d - \Delta y$ the new state when the regulation signal decreases from $y(t)$ to $y(t+1) = y(t)-\Delta y$ rendering $D(t+1)=-1$, while the rest state variables remain unchanged. The superscripts u (d) stand for upwards (downwards) RSR signals. $J(i,\bar{i}^u)$ can be written similarly with minor notational changes.

III. Consumer Utility Realization and the Optimal Policy

A. Generalized Utility Probability Distribution Model

Without loss of generality, we select the following utility function which represents a linear relationship between cooling zone temperature rise and utility enjoyed by activating an idle appliance and allowing it to start a cooling cycle

$$U(T) = b(T - T_{\min}). \tag{10}$$

The utility increases proportionately to the cooling zone temperature T. If $p(T)$ were selected to be a static and uniform probability distribution, as is the case with work published so far, the expected period utility rate would be a conveniently concave function of u. Indeed, using (7) we would obtain

$$\begin{aligned} a(t)U_u &= (N-i(t))\lambda \int_{u}^{T_{\max}} U(T)p(T)dT, \\ &= (N-i(t))\lambda \int_{u}^{T_{\max}} b(T-T_{\min})\frac{1}{T_{\max}-T_{\min}}dT, \\ &= (N-i(t))\lambda \frac{b(T_{\max}-u)(T_{\max}+u-2T_{\min})}{2(T_{\max}-T_{\min})}. \end{aligned} \tag{11}$$

This concavity property, no longer holds true under the more realistic modeling of $p(T)$ by a dynamic trapezoid characterized additionally by the time varying quantity $\hat{T}$. Indeed, the trapezoidal representation implies,

$$p(T) = \begin{cases} \frac{2}{T_{\max}+\hat{T}-2T_{\min}}, & T \le \hat{T}, \\ \frac{2(T-T_{\max})}{(\hat{T}-T_{\max})(T_{\max}+\hat{T}-2T_{\min})}, & T \ge \hat{T}. \end{cases}$$

which yields the following expected period utility rate

$$\begin{aligned} &a(t)U_u \\ &= \begin{cases} [N-i(t)]\lambda \frac{2b(C_1-\frac{1}{2}u^2+T_{\min}u)}{T_{\max}+\hat{T}-2T_{\min}}, & u \le \hat{T}, \\ [N-i(t)]\lambda \frac{2b(C_2-\frac{1}{3}u^3+\frac{T_{\min}+T_{\max}}{2}u^2-T_{\min}T_{\max}u)}{(\hat{T}-T_{\max})(T_{\max}+\hat{T}-2T_{\min})}, & u \ge \hat{T}. \end{cases} \end{aligned} \tag{12}$$

with some constants C_1 and C_2.

The introduction of a dynamic $\hat{T}(t)$ dependent $p(T)$ removes the concavity of the expected utility rate as the second derivative of the expected utility is

$$\frac{d}{du^2}a(t)U_u \propto T_{\min}+T_{\max}-2u, \tag{13}$$

and therefore the expected utility function is concave for $u \in [T_{\min}, \max(\hat{T}, \frac{T_{\min}+T_{\max}}{2})]$, and convex for $u \in [\max(\hat{T}, \frac{T_{\min}+T_{\max}}{2}), T_{\max}]$.

Under the static uniform probability distribution $p(T)$, the optimal policy is easily shown to exist. However, we proceed to show that a unique optimal policy exists as well in the case of dynamic $p(T)$. We do this by showing first that a local minimum exists, and then prove that only one local minimum exists, and hence it is the global minimum as well.

B. Optimal Price Policy

We define the difference of the value function $J(i,\bar{i}^d)$ w.r.t. the active appliance state variable i as

$$\Delta(i+1,\bar{i}^d) = J(i+1,\bar{i}^d) - J(i,\bar{i}^d).$$

Using the Bellman equation, we express the optimal policy $u(i,\bar{i}^d)$ in terms of $\Delta(i+1,\bar{i}^d)$

$$\begin{aligned} u(i,\bar{i}^d) &= \arg\min_{u\in[T_{\min},T_{\max}]} g(i,\bar{i}^d) - \lambda(N-i)p_u U_u + \\ &\quad \alpha\big\{ i\mu J(i-1,\bar{i}^d) + \lambda(N-i)p_u J(i+1,\bar{i}^d) + \\ &\quad \gamma_1^d J(i,\bar{i}^u+\Delta y) + \gamma_2^d J(i,\bar{i}^d-\Delta y) \\ &\quad +[1-(i\mu+\lambda(N-i)p_u+\gamma_1^d+\gamma_2^d)]J(i,\bar{i}^d)\big\}, \\ &= \arg\max_{u\in[T_{\min},T_{\max}]} p_u U_u - \alpha p_u \Delta(i+1,\bar{i}^d), \end{aligned} \tag{14}$$

where the second equation is obtained by neglecting terms that are independent of u. Letting the remaining terms in (14) be

$$f(u,\Delta(i+1,\bar{i}^d)) = p_u U_u - \alpha p_u \Delta(i+1,\bar{i}^d), \tag{15}$$

we can write that the optimal policy must satisfy

$$\max_{u\in[T_{\min},T_{\max}]} f(u,\Delta(i+1,\bar{i}^d)). \tag{16}$$

Based on the equation derived in (11), we present the following proposition when assuming a uniform static preferences among large group of consumers.

Proposition 1 If the probability distribution of the utility is uniform with $\hat{T} = T_{\max}$, then $f(u,\Delta(i+1,\bar{i}^d))$ is a concave function of u for $u \in [T_{\min}, T_{\max}]$. A local maximum for $f(u,\Delta(i+1,\bar{i}^d))$ is a global maximum that yields an optimal policy.

Proposition 1 is straightforward because the first term in $f(u,\Delta(i,\bar{i}^d))$ is quadratic and the second term is a linear function of u for $\hat{T} = T_{\max}$. When $\hat{T} < T_{\max}$ with $p(T)$ no longer uniform but trapezoidal, $f(u,\Delta(i+1,\bar{i}^d))$ stops possessing the concavity property which under Proposition 1 guarantees that a local maximum is the global maximum. We therefore proceed to prove that a local maximum to solve (16) is also global for $u \in [T_{\min}, T_{\max}]$ in the following proposition.

Proposition 2 For trapezoid $p(T)$ with $\hat{T} < T_{\max}$, the optimal policy that solves (16) is given by

$$u(i,\bar{i}^d) = \begin{cases} T_{\max}, & \text{if} \quad \alpha\Delta(i+1,\bar{i}^d) \geq U(T_{\max}) \\ T_{\min}, & \text{if} \quad \alpha\Delta(i+1,\bar{i}^d) \leq 0 \\ T_{\min} + \frac{\alpha\Delta(i+1,\bar{i}^d)}{b}, & \text{otherwise} \end{cases}. \tag{17}$$

Proof. We sketch the proof of Proposition 2. We refer to [15] for the complete proof.

For the case $\alpha\Delta(i+1,\bar{i}^d) \geq U(T_{\max})$ or $\alpha\Delta(i+1,\bar{i}^d) \leq 0$, the discounted differential value function is greater than the maximum possible utility realization $U(T_{\max})$ or less than the minimum possible realization 0. We can show that $f(u,\Delta(i+1,\bar{i}^d))$ is a monotonically increasing function in the former case, and is a monotonically decreasing function is the latter case. This would result in choosing the optimal policy to be either $T_{\max}$ or $T_{\min}$.

When $\alpha\Delta(i+1,\bar{i}^d) \in (0, U(T_{\max}))$, we show that the policy given in (17) is a critical point by solving the first order condition, and further a local maximum by the second order condition. Moreover, we show that $f(u,\Delta(i+1,\bar{i}^d))$ is continuous differentiable and has only one critical point inside the allowable control set, then the local maximum is the global maximum. □

Remark 2 The optimal policy is determined by balancing (1) the utility rewards from connected consumers and (2) the differential optimal cost viewed as an estimate of the value function difference across two adjacent states. Consumers utility sensitivity b plays the following role: When b increases, then the optimal policy will decrease for the same value of $\Delta(i+1,\bar{i}^d)$. In the extreme case when $b \to \infty$, we have $u = T_{\min}$ namely the lowest price is broadcast to guarantee the largest possible utility reward; when $b \to 0$, the optimal controller is bang-bang depending on the sign of $\Delta(i+1,\bar{i}^d)$ indicating that consumers are extremely elastic.

Remark 3 The three cases in Proposition 2 correspond to different geometry of $f(u,\Delta(i+1,\bar{i}^d))$; see Fig. 3. With different combinations of b and $\Delta(i+1,\bar{i}^d)$, $f(u,\Delta(i+1,\bar{i}^d))$ can be a monotonically increasing function of u that leads to the optimal control $u(i,\bar{i}^d) = T_{\max}$, or it can be a monotonically decreasing function to render $u(i,\bar{i}^d) = T_{\min}$, or can be a non-concave and non-monotonic function whose local maximum is the global maximum on $(T_{\min}, T_{\max})$.

IV. Numerical Solution Algorithms

We implement two numerical DP solution algorithms, the first for benchmarking and comparison purposes using the conventional value iteration (CVI) approach [16], and the second by leveraging the optimal policy structure proven in Proposition 2 of Sec. III which we call *assisted value iteration* (AVI) algorithm. The AVI algorithm replaces the computationally inefficient discretization of the allowable policy space and exhaustive search over it at each iteration. We instead use the policy in (17) at each iteration because it is optimal for a given value function at the current iteration resembling policy iteration algorithms. Numerical results from the CVI and AVI algorithms are shown in Fig. 4. We

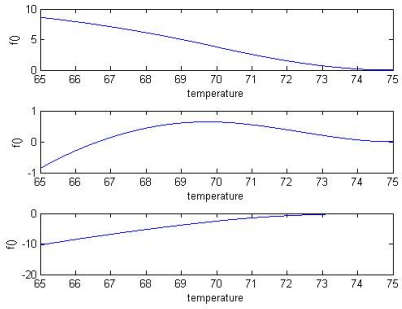

Fig. 3. Geometric features of $f(u,\Delta(i,\bar{i}^d))$. (1) $f(u,\Delta(i,\bar{i}^d))$ is monotonically decreasing when $\alpha\Delta(i+1,\bar{i}^d) \leq 0$. The optimal policy is $u = T_{\min}$. (2) $f(u,\Delta(i,\bar{i}^d))$ has unique global maximum when $\alpha\Delta(i+1,\bar{i}^d) \in (0, b(T_{\max} - T_{\min}))$. The function may not be concave for $\hat{T} \neq T_{\max}$. (3) $f(u,\Delta(i,\bar{i}^d))$ is monotonically increasing when $\alpha\Delta(i+1,\bar{i}^d) \geq b(T_{\max} - T_{\min})$. In this case the optimal policy is $u = T_{\max}$.

find that the CVI algorithm yields policies selected from the discretized allowable policy set and the AVI algorithm provides a smooth and continuous policy. In addition, we observe two monotonicity properties: (*i*) for a given RSR signal $y(t)$, the optimal policy monotonically increases as the aggregated consumption increases, and (*ii*) for a given aggregated consumption, the optimal policy monotonically increases as $y(t)$ decreases. This policy structure is consistent with the smart building operator's objective in reducing the deviation between $y(t)$ and $i(t)$.

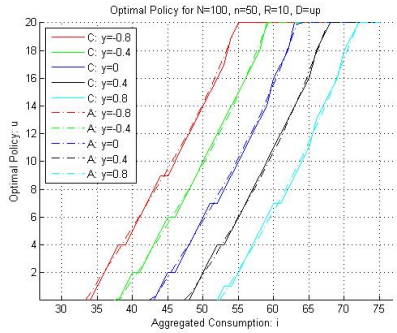

Fig. 4. For fixed value of $y = -0.8, \ldots, 0.8$, the monotonic optimal value of u are displayed as functions of i as indicated by Theorem 1. C (A) stands for CVI (AVI) algorithm. In CVI, the policy space is discretized into 11 possible value from 0 to 10. In AVI, the policy space is continuous.

We compare the computational performance of the CVI and AVI for different state space size problems in Table I. The ADP algorithm assisted by Proposition 2 directly gets the optimal policy at each iteration compared with the optimal policy search in CVI. It is not surprising that computational time is reduced by over 60%.

TABLE I

COMPARISON OF THE COMPUTATIONAL PERFORMANCES

Problem Size ($\|N\| * \|y\| * \|D\|$)	100*20*2	500*40*2	2000*40*2
CVI Computation Time (sec)	168.7	1586.2	6659.7
AVI Computation Time (sec)	14.9	418.2	1319.1

V. CONCLUSION

This paper relaxes the assumption that the utility of market participants is a static (uniform) distribution that is independent of control history. We show that a dynamically changing trapezoid pdf captures the dynamics of market participant preferences in the cooling appliance duty cycle paradigm considered here, proceed to model dynamic preferences, and succeed to overcome the complexity that it introduces. We derive an analytic expressions characterizing the optimal policy which is used to design and implement efficient and scalable numerical solution algorithms. Future work will investigate the structure of the optimal policy and the value function in order to design low cost algorithms.

REFERENCES

[1] Wei Zhang, K. Kalsi, J. Fuller, M. Elizondo, and D. Chassin. Aggregate model for heterogeneous thermostatically controlled loads with demand response. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8, July 2012.

[2] Stephan Koch, Johanna L. Mathieu, and Duncan S. Callaway. Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services. In *the Proceedings of the 17th Power Systems Computation Conference*, 2011.

[3] Bowen Zhang and John Baillieul. A packetized direct load control mechanism for demand side management. In *the 51st IEEE Conference on Decision and Control*, pages 3658–3665, 2012.

[4] Zhongjing Ma, Duncan Callaway, and Ian Hiskens. Decentralized charging control for large populations of plug-in electric vehicles. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 206–212. IEEE, 2010.

[5] Lingwen Gan, Ufuk Topcu, and Steven Low. Optimal decentralized protocol for electric vehicle charging. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 5798–5804. IEEE, 2011.

[6] DJ Hammerstrom, R Ambrosio, J Brous, TA Carlon, DP Chassin, JG DeSteese, RT Guttromson, GR Horst, OM Järvegren, R Kajfasz, et al. Pacific northwest gridwise testbed demonstration projects. *Part I. Olympic Peninsula Project*, 2007.

[7] Michael C. Caramanis, Ioannis Ch. Paschalidis, Christos G. Cassandras, Enes Bilgin, and Elli Ntakou. Provision of regulation service reserves by flexible distributed loads. In *the 51th IEEE Conference on Decision and Control*, pages 3694–3700, 2012.

[8] Ioannis Ch. Paschalidis, Binbin Li, and Michael C. Caramanis. A market-based mechanism for providing demand-side regulation service reserves. In *the 50th IEEE Conference on Decision and Control and European Control Conference*, pages 21–26, December 2011.

[9] Frank P Kelly, Aman K Maulloo, and David KH Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, 49(3):237–252, 1998.

[10] Ioannis Ch. Paschalidis and John N. Tsitsiklis. Congestion-dependent pricing of network services. *IEEE/ACM Trans. Networking*, 8(2):171–184, April 2000.

[11] PJM Balancing Operations: http://www.pjm.com/~/media/documents.

[12] David. P. Chassin and J. C. Fuller. On the equilibrium dynamics of demand response in thermostatic loads. In *the 44th Hawaii International Conference on System Sciences*, 2011.

[13] Johanna L. Mathieu and Duncan S. Callaway. State estimation and control of heterogeneous thermostatically controlled loads for load following. In *the 45th Hawaii International Conference on System Sciences*, pages 2002 – 2011, 2012.

[14] PJM real signal: http://www.pjm.com/markets-and-operations/ancillary-services/mkt-based-regulation/fast-response-regulation-signal.aspx.

[15] Bowen Zhang, Michael Caramanis, and John Baillieul. Control of smart building dynamic energy service preferences for efficient regulation service. http://arxiv.org/abs/1403.4828, March 2014.

[16] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2007.