

Boston University**OpenBU****<http://open.bu.edu>**

BU Open Access Articles

BU Open Access Articles

Optimal provision of distributed reserves under dynamic energy service preferences

This work was made openly accessible by BU Faculty. Please [share](#) how this access benefits you. Your story matters.

Version	
Citation (published version):	B Zhang, MC Caramanis, J Baillieul. "Optimal Provision of Distributed Reserves Under Dynamic Energy Service Preferences."

<https://hdl.handle.net/2144/29264>*Boston University*

Optimal Provision of Regulation Service Reserves Under Dynamic Energy Service Preferences

Bowen Zhang, *Member, IEEE*, Michael C. Caramanis, *Senior Member, IEEE*, and John Baillieul, *Life Fellow, IEEE*

Abstract—We propose and solve a stochastic dynamic programming (DP) problem addressing the optimal provision of bidirectional regulation service reserves (RSR) by pools of duty cycle appliances. Using our previously introduced concept of packetized energy, the approach (i) models the dynamics of the utility $U(T)$ that each consumer associates with a block/packet of energy when the prevailing temperature is T in the appliance specific comfort zone, and (ii) derives a dynamic pricing policy that closes the loop between time varying appliance specific utilities and a regulation signal broadcast by the Independent System Operator (ISO) every 4 seconds. Since T evolves dynamically as a function of past energy block purchases and heat losses, T and $U(T)$ are time varying. Consumer utility levels $U(T)$ are modeled by the probability distribution of utility levels across the ensemble of all idle appliances at time t . This probability distribution provides a statistic of the detailed appliance specific utility information which is sufficient for the determination of an optimal pricing policy. Moreover, it can capture appliance specific diversity within comfort zone limits and temperature-utility function relationships. Dynamic pricing policies investigated in the literature assume that this ensemble probability distribution is time invariant and in fact uniform. A contribution of the work that follows is the replacement of this unrealistic simplifying assumption with the introduction of a new state variable modeling the aforementioned dynamic probability density function. The new state variable evolves as a function of the dynamic pricing policy which aims to track the centrally broadcast regulation signal trading off optimally between tracking error and appliance energy consumption utility. We specifically model active and idle appliances by means of a closed queuing system with price-controlled transitions from the idle to the active queue. Empirical evidence indicates that the dynamically changing frequency distribution of appliance specific utilities is closely approximated by a trapezoid-shape distribution function characterized by a single dynamic parameter which is in fact the aforementioned sufficient statistic. The optimal provision of RSR is addressed within a DP framework where we (i) derive an analytic characterization of the optimal pricing policy and the differential cost function, and (ii) prove optimal policy monotonicity and value function convexity. These properties enable us to implement a smart assisted value iteration (AVI) and an approximate DP (ADP) algorithm that exploit related functional approximations. Numerical results demonstrate (1) the validity of the solution techniques and the computational advantage of the proposed ADP on realistic, large-state-space problems, and (2) that the RSR provision based on the optimal control that respects

dynamic consumer preferences has higher tracking accuracy relative to approaches assuming static uniformly distributed consumer preferences.

Index Terms—Approximate dynamic programming, smart grid, electricity demand response, regulation service reserves, demand response under dynamic preferences.

I. INTRODUCTION

The urgently needed reduction of CO₂ emissions will rely on the adoption of significant renewable electricity generation, whose volatility and intermittency will result in commensurate increase in the provision of secondary or Regulation Service Reserve (RSR) required to ensure power system stability through adequate Frequency Control (FC) and Area Control Error (ACE) management [1], [2], [3], [4]. The provision of these additional RSRs from centralized generation units, until now has been the primary approach, is expensive and poses a major challenge to massive renewable integration. The option of employing much lower cost demand-control-based RSR is rightly a major Federal Energy Regulatory Commission target [5]. Not surprisingly, there is extensive literature on demand management addressing, for example, direct load control (DLC) of thermostatic appliances to provide frequency reserve or load shedding [6], [7], [8], [9], [10], [11], [12], [13], the optimal coordination of distributed resources in a micro-grid flexible loads [14], [4], [15], and real time pricing approaches to reduce peak demand [16], [17], [18]. In contrast to DLC where consumers transfer the control of powering their appliances to a centralized decision maker, dynamic price directed demand response by individual consumers can achieve similar system benefits with lower consumer utility loss. The PNNL Olympic Peninsula project [19] is a successful pilot study where consumer specified price elasticities have been used to control thermostatic appliance by dynamic prices that proved superior to fixed or peak/off-peak prices.

Customer acceptance of price directed demand response, however, requires that customer preferences are modeled accurately and that dynamic prices are adapted to these preferences. In this paper we focus on the deployment of bidirectional regulation service reserves (RSR) by a pool of duty cycle cooling appliances in a smart building or neighborhood micro-grid. Based up-to-date consumer preferences and the current price signal, the microgrid operator will purchase and allocate a block or packet of energy at the beginning of each cycle. Dynamic prices encourage (or discourage) idle appliances to purchase a packet of energy and start a cycle. Two distinct time scales, one hour and 4 seconds, are involved in promising and

B. Zhang is a visiting scholar at Department of Mechanical Engineering, Boston University, Boston, MA, 02215 USA e-mail: bowenz@bu.edu.

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.

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.

Research reported here was supported by NSF grant 1038230.

Manuscript received *****; revised *****.

deploying RSRs. A smart building promises to the Independent System Operator (ISO) a certain quantity of up and down RSR in the hour ahead market and is compensated at the hour ahead markets RSR clearing price. During this hour, however, the smart building must deploy these reserves to track the RSR signal that is rebroadcast by the ISO at 4 second intervals taking values in the interval from -100% to +100%. Optimal deployment of the RSRs at the 4 second time scale is defined as the deployment driven by dynamic prices decided at the same 4 second time scale by the smart building operator, so as to maximize the average utility of cooling appliances over the hour. The smart building operator can not anticipate the ISO RSR signal, but it has access to a statistical model describing its stochastic dynamics during the relevant hour [7], [20].

To obtain the optimal dynamic pricing policy, we propose and solve an infinite horizon discounted cost stochastic dynamic programming (DP) problem. The hour long horizon is considered for all intents and purposes to be infinite relative to the 4 second time scale characterizing the regulation signal and price updates. State variables consist of the RSR signal level $y(t)$, the direction $D(t)$ of RSR signal defined as the sign of $y(t) - y(t-1)$, the number of connected appliances $i(t)$, and the sufficient statistic $\hat{T}(t)$ characterizing the probability distribution of utility level. More specifically, we model the dynamics of the utility that each space conditioning appliance enjoys from the consumption of a block/packet of energy purchased when the prevailing temperature is at a specific level $T \in [T_{\min}, T_{\max}]$ within the appliance specific temperature comfort zone. The utility at time t is modeled as an affine function of the temperature T prevailing at time t . The utility of purchasing a packet of energy by a cooling appliance will be low when its temperature is close to T_{\min} , and high when it is close to T_{\max} . To model diversity among appliances, coefficients are appliance specific independent random variables and is normalized by $|T_{\max} - T_{\min}|$. The variable T at a specific appliance evolves dynamically as a function of that appliance's past trajectory of energy block purchases and heat losses, and hence T and $U(T)$ are time varying. The optimal pricing policy, however, does not require individual appliance utility levels at time t . The sufficient information is captured by the probability distribution of utility levels across the ensemble of all idle appliances at time t . This probability distribution is therefore the sufficient statistic that replaces appliance specific utilities in the stochastic DP state vector.

The principal contribution in what follows is a modeling extension in which the system state is a time varying probability density function that we argue provides important information about the dynamics of aggregated effects of individual consumer utility functions. This represents a significant improvement over the more traditional models (e.g. [7], [21], [20]) that assume a virtually infinite pool of homogeneous users and appliances whose preferences are static and independent of the recent history of price control signals. Our approach provides a more appropriate level of granularity that will enhance the value of RSRs that come from finite pools of duty-cycle appliances. The concepts to be described may also have useful applications in related areas such as dynamic pricing policies in the context of Internet service and mobile

telephony bandwidth management [22], [23], [24], [25]. The bottom line message of the paper is that in communication between consumers and service providers, it is important to stop relying on the uniform and time invariant assumption and thereby avoid ratepayer revolt [26].

Focusing on a smart building with a finite number of duty cycle cooling appliances, we use a dynamic probability distribution to represent cooling zone specific occupant preferences to transition their appliance from an idle to an active state. This is achieved by employing a dynamically evolving sufficient statistic that captures the ensemble of idle cooling zone appliance preferences, namely the recent-price-trajectory-dependent probability distribution of a sampled cooling zone occupant's reservation price. Moreover, we improve the tractability of the DP formulation by (i) deriving an analytic characterization of the optimal policy and the differential cost function, and (ii) proving useful monotonicity and convexity properties. These properties motivate the use of appropriate basis functions to construct a parametrized analytic approximation of the value function used to design an approximate DP (ADP) algorithm [27] that estimates optimal value function approximation parameters and near optimal policies. To support the relative advantage of our approach, we compare the real time RSR tracking performance of our dynamic utility based optimal dynamic pricing policy to that of a uniform static utility based dynamic pricing policy. The tracking of our policy will be shown to be clearly superior.

The rest of the paper proceeds as follows. In Sec.II we establish a DP problem by describing the state dynamics, the sufficient statistic of cooling zone preferences, the period costs, and the Bellman Equation. Time is quantized into 4-second intervals, and the electricity used by the appliances is also quantized using our previously announced concept of packetized energy. Sec.III compares the static uniform distribution representation of cooling zone preferences adopted in past work, with the proposed dynamic and non-uniform preference distribution formulation. More specifically, it uses the Bellman equation to prove analytical expressions relating the optimal policy to partial differences in the value function. Sec.IV proves monotonicity properties of the differential cost function and the optimal policy. It also discusses asymptotic behaviour of optimal policy sensitivities as the number of appliances becomes large. Sec.V utilizes the proven properties of the value function and the optimal control policy to propose and implement (i) an assisted value iteration algorithm, and (ii) a parametrized approximation of the value function. It also develops an ADP approach for estimating the value function approximation parameters. Numerical results demonstrate the superior performance of dynamic utility versus static utility modeling pricing policies. Sec.VI concludes the paper and proposes future directions.

II. PROBLEM FORMULATION

We consider an advanced energy management building with N cooling appliances. The smart building operator (SBO) contracts 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 consumption rate that the SBO purchased in the hour ahead market, and R is the maximum up or down RSRs that the SBO agreed to provide in the hour-ahead market. \bar{n} equals the average number of active appliances, which is related to the average number of duty cycles per hour required to maintain the cooling zone temperature within its comfort zone for the prevailing outside temperature and building heat transfer properties. R is chosen as a fraction of N that can be determined based on the appliance response property, such as cooling rate. In real time, the ISO would broadcast the RSR signal $y(t)$ to the SBO, which is the output of a pre-specified proportional integral filter of observed Area Control Error (ACE) and System Frequency excursions. The RSR signal $y(t)$ has an average value of 0 during each hour and its dynamics can be modeled as a Markov process [28]. Given these three values, the SBO assumes 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 (usually every 2 or 4 seconds). To this end the SBO broadcasts a real time price signal $\pi(t)$ to all cooling appliances to modulate the number of connected appliances $i(t)$ and hence control the aggregated consumption. Appliances will detect price at rate λ to resume cooling by comparing its utility $U(T)$ to $\pi(t)$.¹ Appliances utility values depend on their corresponding cooling zone temperature T . In this paper $U(T)$ is a monotonically increasing function of T over $T \in [T_{\min}, T_{\max}]$. Once $U(T) \geq \pi(t)$, the appliance will connect and consume an packet of energy having departure rate μ . ($1/\mu$ equals the average time it takes a cooling appliance to complete the consumption of an energy packet.) See [8], [29] where the associated electric power consumption is defined as a packet of electric energy needed to provide the work required by a single cooling cycle.

The RSR signal tracking error, $e(t) = i(t) - \bar{n} - y(t)R$, 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 modeled by a dynamically evolving probability distribution of idle appliance zone temperatures $p_t(T)$. Assuming a known relationship connecting cooling zone preferences to temperatures, we derive the probability distribution of preferences $p_t(T)$ (or derived utility levels) for cooling appliance activation as well as the sufficient statistics $\hat{T}(t)$ that characterizes $p_t(T)$. We next describe system dy-

namics, and formulate the period cost function of the relevant stochastic dynamic problem.

A. State Dynamics

The state variables contain $i(t), y(t), D(t)$ and $\hat{T}(t)$. Queues $i(t)$ and $N - i(t)$ constitute a closed queuing 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 $i(t)$ behaves also as an infinite server queue with each server exhibiting a constant service rate μ . The dynamics of $y(t)$ and the dependent state variable $D(t) = \text{sgn}[y(t) - y(t - 4\text{sec})]$ are characterized by transitions taking place in short but constant time intervals, τ_y^2 , resulting in $y(t)$ staying constant, increasing or decreasing by a typical amount of $\Delta y = \tau_y/300\text{sec}$ [28]. These transitions are outputs of a proportional integral filter operated by the ISO whose inputs are system frequency deviations from 60 hz and Area Control Error (ACE). Since the frequency deviation and ACE signal can be approximated by a white noise process resulting from imbalance between stochastic demands and supply, $y(t)$ is then an unanticipated random variable which is described by memoryless transitions that depends only on the current value. Therefore we can 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 $\Delta_t \ll \tau_y$ 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 and derive the following state dynamics.

1) *Dynamics of $y(t)$* : The transition probabilities of the discrete time Markov process $y(t)$ depend on $y(t)$ as well as $D(t)$ [7], [20], which is $y(t)$'s changing direction, either positive or negative, for every 4 seconds. $D(t) = \text{sgn}[y(t) - y(t - 4\text{sec})]$. Statistical analysis on historical PJM data on $y(t)$ trajectories indicate a weak 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 $D(t) = 1$ ($D(t) = -1$), and selecting these rates to correspond to the exponential rates that are consistent with the geometric probability distribution described above we have:

$$\begin{cases} \gamma_1^u = 0.8\Delta_t/\tau_y, \gamma_2^u = 0.2\Delta_t/\tau_y \\ \gamma_2^d = 0.8\Delta_t/\tau_y, \gamma_1^d = 0.2\Delta_t/\tau_y \end{cases}$$

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

¹In this paper we assume consumers to have the same sensitivity preference across zones, namely parameter b in the utility function is fixed such that the mapping between temperature threshold and the price signal is bijective. However, T_{\min} and T_{\max} are not the same for all consumers. $[T_{\min} - T_{\max}]$ is normalized by the sensitivity parameter b . Alternative implementations in the literature rely on similar approaches, but, in our opinion, less likely to be accepted by consumers. For example, the PNNL Olympic Peninsula project [19], has assumed the presence of aggregators, who, typically provide participants with multiple comfort level choices, and participants choose the comfort level based on their own preferences. Aggregators group participants into multiple subgroups based on their choices. Different groups of participants are characterized by different price elasticities and provide different type of reserves. For example, a participant group with the highest elasticity provides regulation service reserves (RSR) with a response time of 4 seconds. Participant groups with medium price elasticity provide spinning reserves with response times of 15 to 30 minutes. Participant groups with low price elasticity provide capacity reserves that have even slower response times.

Setting the control update period as the operative time unit, i.e., $\Delta_t = 1$, we have the following uniformized dynamics of $y(t)$:

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

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 $T \geq u(t)$. As described in the notation subsection, 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 . Therefore the arrival rate into $i(t)$ is

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

namely $a(t)$ 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 departure rate of a cooling cycle. Modeling the consumption of an energy packet as an exponential random variable with rate μ such that $1/\mu$ equals the average cooling cycle duration, the departure rate is

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

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.

Fig. 1 represents the stochastic dynamics of the number of active and idle appliances as a two queue closed queuing network with queue levels summing to N .

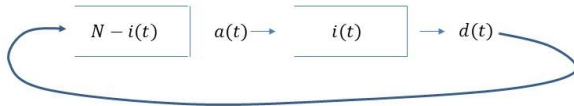


Fig. 1. Dynamics of the closed queuing network with infinite number of servers. The service rate at the active appliance queue is constant, while it is Markov modulated at the idle appliance queue.

3) *Dynamics of $D(t)$* : Recalling that $D(t) = \text{sgn}[y(t) - y(t - 4\text{sec})]$, it is clear that the dynamics of $D(t)$ are fully determined by the dynamics of $y(t)$. We next show that the dynamics of $\hat{T}(t)$ are also determined by the dynamics of $y(t)$.

4) *Dynamics of $\hat{T}(t)$* : Based on related work in HVAC system modelling [9] and [30], 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 simulate two RSR signals from the ISO for the observation of the time varying probability distribution of consumers' temperature. The first is a standard ISO RSR signal trajectory that aspiring RSR market participants must respond to in order to demonstrate that they have the ability to track. This is referred to in the PJM manual as the standard T-50 qualifying test [28]. The second is a real time signal downloaded from PJM [31]. 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. Assuming standard heat transfer relationships, extensive simulation reported below indicates 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. 2. $p_t(T)$ is the probability density function (pdf) of T at time t , $T \in [T_{\min}, T_{\max}]$. T_{\min} and T_{\max} are the threshold values of room temperature that determine the appliance cooling zone occupant's comfort zone. ε^h is of small value and approximated to be zero. Trapezoid has base $(T_{\min}, 0)$ to $(T_{\max}, 0)$ and top side (T_{\min}, h) to (\hat{T}, h) where $h = 2/(T_{\max} + \hat{T} - 2T_{\min})$. Note that \hat{T} and hence h are time varying quantities. We will use $p(T)$ to denote the pdf by dropping the time index in derivation and proof if time is not considered.

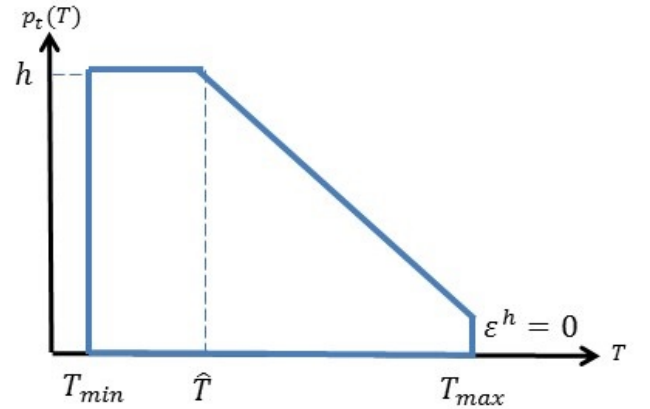


Fig. 2. Trapezoid probability distribution function $p(T)$ with $T \in [T_{\min}, T_{\max}]$ parametrized by a single parameter \hat{T} . Simulation shows that ε^h is of small value and assumed to be zero throughout paper. The height of the trapezoid is $h = 2/(T_{\max} + \hat{T} - 2T_{\min})$.

Fig. 3 shows the accuracy of using a trapezoid probability distribution to represent idle appliance cooling zone temperatures obtained from a Monte Carlo Simulation of the price controlled system. We discretize the temperature into 20 states and simulate a total number of 16000 appliances to observe a smooth probability distribution. The duty cycle on and off time are both 10 minutes. The price detection rate from idle appliances is 1 minute. PJM's RSR signal is broadcast every 4 seconds. In Fig. 3, the numerous black curves are the actual

probability distribution recorded at different time stamps for trapezoids characterized with $\hat{T}(t) = 5, 6, 7, 8$. The red curve is the mean value of the set of black curves taken at each temperature state. The red curve is then approximated by the trapezoid green curve proposed in Fig. 2. $\hat{T}(t)$ and the trapezoid approximation are then the mean statistics of the actual frequency distribution of the temperature distribution. Note that the trapezoids are completely specified by two static quantities, T_{\min} and T_{\max} , forming the base of the trapezoid, and the time varying quantity $\hat{T}(t)$ that determines the height of the top horizontal side. It should be noted that the trapezoid approximation is based on the assumption of (1) small value of $\varepsilon^h = 0$ and (2) equal probability $p_i(T)$ for temperature $T \in [T_{\min}, \hat{T}]$. In this paper, we model the trapezoid distribution based on the single sufficient statistics \hat{T} , future work will consider modeling the preference distribution with additional statistics to fully capture the distribution shape.

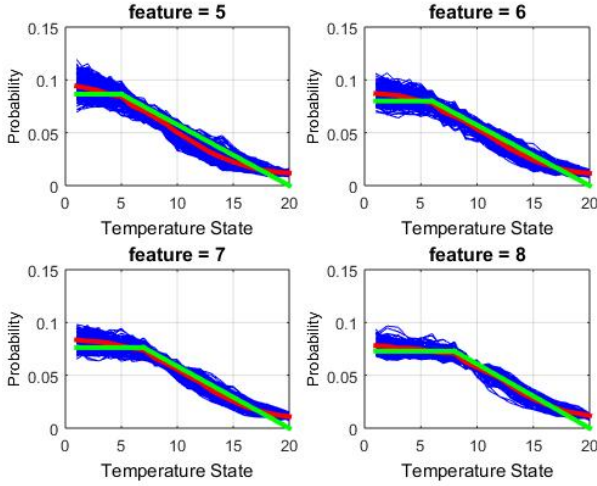


Fig. 3. Trapezoid probability distribution function $p(T)$ with $T \in [T_{\min}, T_{\max}]$ parametrized by a single parameter \hat{T} . Height of the trapezoid is $h = 2/(T_{\max} + \hat{T} - 2T_{\min})$.

The upper left plot in Fig. 4 shows the time evolution of the trapezoids representing the idle appliance cooling zone temperature histograms throughout time when we are simulating the RSR tracking for the real signal. The upper right plot is the contour plot of the number of appliances where we can see clearly a time varying trapezoid distribution shaping the preferences. We filter the contour plot to get \hat{T} in lower right figure. Based on the time series recorded for $y(t)$ (lower left figure) and $\hat{T}(t)$, we find a strong anti-correlation between the two vectors in both simulation for the T-50 standard signal and the real RSR signal that are given by -0.9833 and -0.8106, respectively. We propose the regression of $\hat{T}(t)$ on $y(t)$ with the following linear function

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

where α_0 corresponds to the value of $\hat{T}(t)$ when the building's energy consumption level is \bar{n} , $\alpha_1 < 0$, and ω is a zero mean symmetrically distributed error. These results not only explain most of the variability but are also sensible and conform with our expectations. Indeed, large values of $y(t)$ indicate

a history of repeated broadcasts of low prices to achieve high consumption levels requested by the ISO. Small values of $\hat{T}(t)$ approaching T_{\min} are observed for high $y(t)$ levels, while for low levels of $y(t)$, $\hat{T}(t)$ is large. 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. The verification of expectations by actual observations justifies the adoption of a tractable dynamic utility model conforming to the dynamics $\hat{T}(t) = \alpha_0 + \alpha_1 y(t) + \omega$ where ω is a zero mean symmetric random variable. Since the SBO is able to observe the actual cooling zone temperatures through its access to Building Automation Control (BAC), the dynamics above are adequate for optimal control estimation. We finally note that the mapping of temperature to consumption utility allows the dynamic and past-control-dependent distribution of cooling area zone temperatures to provide a dynamic and past-control-dependent distribution of cooling area consumption utility levels. In the end, Fig. 5 shows the actual \hat{T} (black curve), the predicted \hat{T} (red curve) based on linear regression, and the error between the two values. It can be seen that the error is zeros mean and symmetrically distributed that is consistent with our assumption in proposing (3).

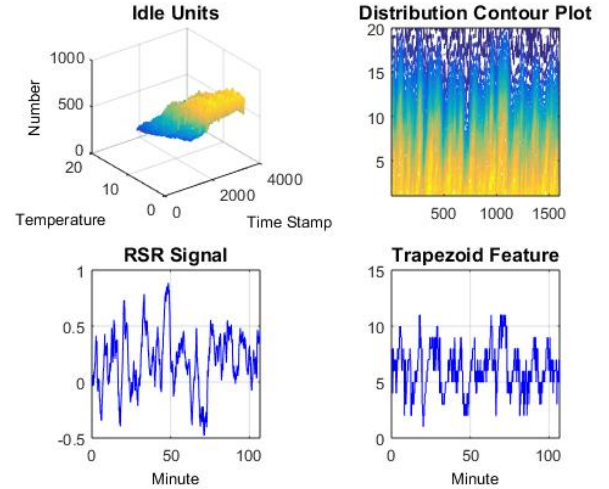


Fig. 4. Monte Carlo simulation is used to determine the time evolution of temperature histograms in idle appliance cooling areas. We observe that these histograms conform to a time varying trapezoid shape with $\hat{T}(t)$ a linear function of the ISO RSR signal $y(t)$.

B. Period Cost

The period cost consists of two parts: a penalty for deficient ISO RSR signal tracking and the utility realized by appliance users. The penalty for deficient tracking rate at time t is defined as:

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

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

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

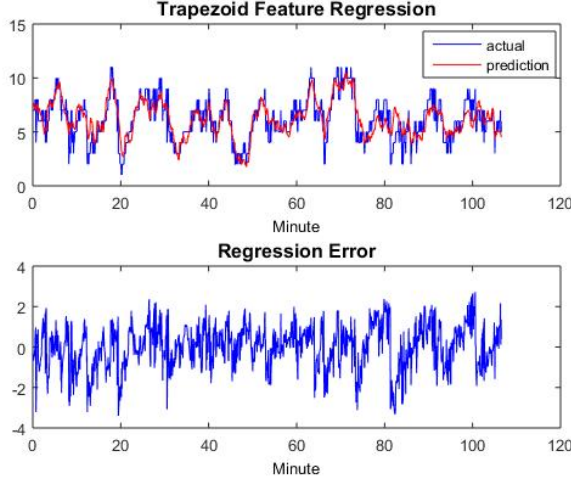


Fig. 5. Original \hat{T} and the regression based prediction of \hat{T} are shown in the upper figure. The linear regression explains 81.06% of the uncertainty of \hat{T} . Errors are observed to be Gaussian noise satisfying the assumption of regression.

The expected utility realized from a group of idle appliances can be calculated by three components: (1) the number of idle appliances, $N - i(t)$, that can potentially connect, (2) the connection probability of each appliance, $p_{u(t)}$, that depends on the dynamic frequency distribution $p_t(T)$, and (3) the utility rate per connection, which is the expected monetary value per connection calculated by consumers' temperature based utility function $U(T)$ and the probability distribution of room temperature T . It should be noted that while the utility for a given consumer is determined by his room temperature T , for a group of consumers the aggregated utility is based on the frequency distribution of room temperatures of the group, and therefore depends on $p_t(T)$ instead of T .

For a given price signal $\pi(t)$, there is a threshold temperature value $u(t)$ obtained by solving $\pi(t) = U(u(t))$. Idle appliances having temperature at least $u(t)$ will connect. The consequent expected utility value realized by each connecting cooling appliance is

$$U_{u(t)} = \frac{\int_{u(t)}^{T_{\max}} U(T) p_t(T) dT}{\int_{u(t)}^{T_{\max}} p_t(T) dT}. \quad (6)$$

Denoting $a(t)$ as the expected number of connections from the idle appliance group, the period expected utility value is

$$\begin{aligned} a(t)U_{u(t)} &= [N - i(t)]\lambda p_{u(t)}U_{u(t)}, \\ &= [N - i(t)]\lambda \int_{u(t)}^{T_{\max}} p_t(T) dt \frac{\int_{u(t)}^{T_{\max}} U(T) p_t(T) dT}{\int_{u(t)}^{T_{\max}} p_t(T) dT}, \quad (7) \\ &= [N - i(t)]\lambda \int_{u(t)}^{T_{\max}} U(T) p_t(T) dT. \end{aligned}$$

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

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

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 depend on $u(t)$. For notation simplicity we let $\tilde{u}(t) = \{y(t), D(t) = 1, \hat{T}(t)\}$ ($\tilde{d}(t) = \{y(t), D(t) = -1, \hat{T}(t)\}$) to be the state variables that make up the complement of $i(t)$ when the RSR signal is going up (down), so that $\{i(t), \tilde{u}(t)\}$ is the representation of the full state vector. 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 $D(t) = -1$.

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

$J(i, \tilde{d})$ is the value function satisfying the Bellman equation, with α denoting the discount factor. For notational simplicity we denote by $\tilde{u} + \Delta y$ 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 $\tilde{d} - \Delta y$ 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) standing for upwards (downwards) RSR signals. $J(i, \tilde{u})$ can be written similarly with minor notational changes.

III. UTILITY REALIZATION AND OPTIMAL POLICY

A. Uniform Utility Probability Distribution Model

For illustration purposes, we start with a simple utility function that represents a linear relationship between cooling zone temperature and utility enjoyed by activating an idle appliance and allowing it to embark on a cooling cycle

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

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 can obtain that the expected period utility rate

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

is a concave function of the policy u .

In general, this concavity property of the expected period utility rate holds for broader class of utility functions $U(T, b, \varepsilon)$ – where T is the temperature, b is the utility sensitivity, and ε is a consumer specific value to characterize individual utility preference. We require two properties hold for the utility function: (i) we need ε to be a random parameter independent of T , and (ii) $E_\varepsilon[U(T, b, \varepsilon)]$ is a monotonically increasing function of T . We prove this property by the following reasoning. Taking expectation on U_u in (6), we have

$$U_u = E_\varepsilon \left[\frac{\int_{u(t)}^{T_{\max}} U(T, b, \varepsilon) p(T) dT}{\int_{u(t)}^{T_{\max}} p(T) dT} \right] = \frac{\int_{u(t)}^{T_{\max}} E_\varepsilon[U(T, b, \varepsilon)] p(T) dT}{\int_{u(t)}^{T_{\max}} p(T) dT}. \quad (12)$$

Taking the derivative of the expected utility rate $a(t)U_u$ with respect to u

$$\begin{aligned} \frac{d}{du} a(t)U_u &= \frac{d}{du} (N-i(t))\lambda \int_{u(t)}^{T_{\max}} U(T)p(T)dT, \\ &= -\frac{(N-i(t))\lambda}{T_{\max}-T_{\min}} E_\varepsilon[U(u, b, \varepsilon)]. \end{aligned} \quad (13)$$

Since $E_\varepsilon[U(u, p, \varepsilon)]$ increases with u , we have $\frac{d}{du} E_\varepsilon[U(u, p, \varepsilon)] > 0$, which leads to

$$\frac{d^2}{du^2} a(t)U_u = -\frac{(N-i(t))\lambda}{T_{\max}-T_{\min}} \frac{d}{du} E_\varepsilon[U(u, p, \varepsilon)] < 0 \quad (14)$$

Hence the expected period utility rate is a concave function of the policy u .

B. Generalized Utility Probability Distribution Model

The concavity property no longer holds true under the realistic modelling of $p(T)$ by a dynamic trapezoid characterized additionally by the time varying quantity \hat{T} . Indeed, the realistic representation implies the following consumers' preferences distribution

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

For example, if we use a linear utility function as in (10), it yields the following expected period utility rate

$$\begin{aligned} a(t)U_u &= \begin{cases} [N-i(t)]\lambda \frac{2b(C_1 - \frac{1}{2}u^2 + T_{\min}u)}{T_{\max} + \hat{T} - 2T_{\min}}, & u \leq \hat{T}, \\ [N-i(t)]\lambda \frac{2b(C_2 - \frac{1}{2}u^3 + \frac{T_{\min} + T_{\max}}{2}u^2 - T_{\min}T_{\max}u)}{(\hat{T} - T_{\max})(T_{\max} + \hat{T} - 2T_{\min})}, & u \geq \hat{T}. \end{cases} \end{aligned} \quad (15)$$

where C_1 and C_2 are some constants. The introduction of a dynamic $\hat{T}(t)$ dependent $p(T)$ removes the concavity of the expected utility rate since the second derivative of the expected utility is

$$\frac{d^2}{du^2} a(t)U_u \propto T_{\min} + T_{\max} - 2u. \quad (16)$$

And therefore the expected period utility rate 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 can be easily obtained since we can set derivative to zero to get a local maximum which is also global. In addition, we proceed to show that a unique optimal policy exists as well under dynamic $p(T)$ circumstances. We do this by showing first that a local maximum exists, and then prove that only one local maximum exists, and hence it is the global maximum as well.

C. Optimal Price Policy

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

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

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

$$\begin{aligned} u(i, \tilde{i}^d) &= \arg \min_u g(i, \tilde{i}^d) - \lambda(N-i)p_u U_u + \\ &\quad \alpha \{ i\mu J(i-1, \tilde{i}^d) + \lambda(N-i)p_u J(i+1, \tilde{i}^d) + \\ &\quad \gamma_1^d J(i, \tilde{i}^u + \Delta y) + \gamma_2^d J(i, \tilde{i}^d - \Delta y) \\ &\quad + [1 - (i\mu + \lambda(N-i)p_u + \gamma_1^d + \gamma_2^d)] J(i, \tilde{i}^d) \}, \\ &= \arg \max_u p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d), \end{aligned} \quad (17)$$

where the second equation is obtained by neglecting terms that are independent of u . Letting

$$f(u, \Delta(i+1, \tilde{i}^d)) = p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d), \quad (18)$$

we can write that the optimal policy must satisfy

$$\max_{u \in [T_{\min}, T_{\max}]} f(u, \Delta(i+1, \tilde{i}^d)). \quad (19)$$

Proposition 1 In the conventional assumption where consumers' utility preference is statically uniformly distributed ($\hat{T} = T_{\max}$), $f(u, \Delta(i+1, \tilde{i}^d))$ is a concave function of the policy u in the allowable control set. The optimal policy is obtained either at the boundary of the allowable control set or at u^* satisfying $\frac{d}{du} f(u, \Delta(i+1, \tilde{i}^d))|_{u=u^*} = 0$.

Proposition 1 is straightforward because the first term in $f(u, \Delta(i, \tilde{i}^d))$ 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 trapezoid, $f(u, \Delta(i+1, \tilde{i}^d))$ no longer possesses the concavity property which under Proposition 1 guaranteed that a local maximum is the global maximum. We therefore proceed to prove existence and uniqueness as follows.

Proposition 2 For trapezoid consumers' preference distribution $p(T)$ with $\hat{T} < T_{\max}$ and non-homogeneous preferences function $U(T, b, \varepsilon)$ having monotonically increasing expected

value $E_\varepsilon[U(T, b, \varepsilon)]$, the optimal policy that solves (19) is described by the following relations

$$u(i, \tilde{i}^d) = \begin{cases} T_{\max}, & \text{if } \alpha\Delta(i+1, \tilde{i}^d) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)] \\ T_{\min}, & \text{if } \alpha\Delta(i+1, \tilde{i}^d) \leq 0 \\ I^{-1}(\alpha\Delta(i+1, \tilde{i}^d)), & \text{otherwise} \end{cases} \quad (20)$$

where $I^{-1}(\alpha\Delta(i+1, \tilde{i}^d))$ is the inverse function of the expected utility function satisfying $E_\varepsilon[U(u, b, \varepsilon)] = \alpha\Delta(i+1, \tilde{i}^d)$.

Proof. Let $f(u, \Delta(i+1, \tilde{i}^d)) = p_u(U_u - \alpha\Delta(i+1, \tilde{i}^d))$, for the three conditions in (20) we claim the following:

1) When $\alpha\Delta(i+1, \tilde{i}^d) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)]$, namely $\alpha\Delta(i+1, \tilde{i}^d)$ is no less than the maximum possible utility per connection, we always have $U_u - \alpha\Delta(i+1, \tilde{i}^d) \leq 0$. Since $U_u - \alpha\Delta(i+1, \tilde{i}^d)$ is a monotonically increasing function and $p_u \geq 0$ is a monotonically decreasing function of u , $f(u, \Delta(i+1, \tilde{i}^d))$ reaches its maximum value at $u = T_{\max}$. On the other hand, if $u(i, \tilde{i}^d) = T_{\max}$ which is the optimal policy, we must have $\alpha\Delta(i+1, \tilde{i}^d) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)]$. To see this necessity, assume that $\alpha\Delta(i+1, \tilde{i}^d) < E_\varepsilon[U(T_{\max}, b, \varepsilon)]$, then there exists a policy $u \neq T_{\max}$ such that $E_\varepsilon[U(T_{\max}, b, \varepsilon)] > U_u > \alpha\Delta(i+1, \tilde{i}^d)$ and $p_u > 0$. Hence $f(u, \Delta(i+1, \tilde{i}^d)) > 0 = f(T_{\max}, \Delta(i+1, \tilde{i}^d))$, which is a contradiction to the assumption that $u(i, \tilde{i}^d) = T_{\max}$ is optimal.

2) When $\alpha\Delta(i+1, \tilde{i}^d) \leq 0$, both $-p_u\alpha\Delta(i+1, \tilde{i}^d)$ and $p_u U_u$ are monotonically decreasing function of u . Therefore $f(u, \Delta(i+1, \tilde{i}^d))$ reaches its maximum at $u = T_{\min}$.

3) When $\alpha\Delta(i+1, \tilde{i}^d) \in (0, E_\varepsilon[U(T_{\max}, b, \varepsilon)])$, we can take the derivative of $f(u, \Delta(i+1, \tilde{i}^d))$ as $f(u, \Delta(i+1, \tilde{i}^d))$ is continuously differentiable on (T_{\min}, T_{\max}) .

$$\begin{aligned} & \frac{d}{du} f(u, \Delta(i+1, \tilde{i}^d)), \\ &= \frac{d}{du} [p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d)], \\ &= \frac{d}{du} \left[\int_{T_{\min}}^{T_{\max}} E_\varepsilon[U(T, b, \varepsilon)] p(T) dT - \alpha\Delta(i+1, \tilde{i}^d) \int_{T_{\min}}^{T_{\max}} p(T) dT \right], \\ &= -p(u) [E_\varepsilon[U(u, b, \varepsilon)] - \alpha\Delta(i+1, \tilde{i}^d)]. \end{aligned} \quad (21)$$

Denoting $u(i, \tilde{i}^d)$ as the optimal control that minimizes $f(u, \Delta(i+1, \tilde{i}^d))$, a necessary condition is to have $u(i, \tilde{i}^d)$ be a local maximum of $f(u, \Delta(i+1, \tilde{i}^d))$. Therefore it satisfies the first order condition

$$p(u) [E_\varepsilon[U(u, T, \varepsilon)] - \alpha\Delta(i+1, \tilde{i}^d)]|_{u=u(i, \tilde{i}^d)} = 0. \quad (22)$$

According to the proof in 1), $p(u) = 0$ ($u(i, \tilde{i}^d) = T_{\max}$) if and only if $\alpha\Delta(i+1, \tilde{i}^d) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)]$. Therefore in this case $p(u) \neq 0$. The only solution to satisfy (22) is

$$E_\varepsilon[U(u(i, \tilde{i}^d), b, \varepsilon)] - \alpha\Delta(i+1, \tilde{i}^d) = 0. \quad (23)$$

Based on the definition of the inverse function, we have

$$u(i, \tilde{i}^d) = I^{-1}(\alpha\Delta(i+1, \tilde{i}^d)). \quad (24)$$

For the second order condition, it can be verified that $\frac{d^2}{du^2} f(u, \Delta(i+1, \tilde{i}^d))|_{u=u(i, \tilde{i}^d)} < 0$. Therefore $u(i, \tilde{i}^d)$ is a local maximum. Moreover, given $f(u, \Delta(i+1, \tilde{i}^d))$ is first order differentiable, $\frac{d}{du} f(u, \Delta(i+1, \tilde{i}^d))$ is continuous and has only one critical point inside the allowable control set, then the local maximum is the global maximum for $u \in [T_{\min}, T_{\max}]$, namely $u(i, \tilde{i}^d) = I^{-1}(\alpha\Delta(i+1, \tilde{i}^d))$. \square

Remark 1 The optimal policy characterization between $u(i, \tilde{i}^d)$ and $\Delta(i+1, \tilde{i}^d)$ does not rely on $p(T)$, namely it holds for broader possible realizations of consumers' real time preferences. This is because (23) has only one solution which is the local and global optimal bearing the same argument in the proof. In addition, it holds for broader class of utility function (linear, quadratic, etc) and non-homogeneous utility incorporating consumer specific preference ε .

Remark 2 The optimal policy is determined by balancing (1) the utility rewards from connected consumers, and (2) the differential of the 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 $\Delta(i+1, \tilde{i}^d)$. In the extreme case when $b \rightarrow \infty$, we have $u = T_{\min}$ namely the lowest price is broadcast to guarantee the largest possible utility reward; when $b \rightarrow 0$, the optimal controller is bang-bang depending on the sign of $\Delta(i+1, \tilde{i}^d)$ indicating that consumers become extremely elastic.

Remark 3 The three cases in Proposition 2 correspond to different geometry of $f(u, \Delta(i+1, \tilde{i}^d))$; see Fig. 6. With different choice of utility function and $\Delta(i+1, \tilde{i}^d)$, $f(u, \Delta(i+1, \tilde{i}^d))$ can be a monotonically increasing function of u that leads to the optimal control $u(i, \tilde{i}^d) = T_{\max}$, or it can be a monotonically decreasing function to render $u(i, \tilde{i}^d) = T_{\min}$, or can be a non-concave and non-monotonic function whose local maximum is the global maximum on (T_{\min}, T_{\max}) .

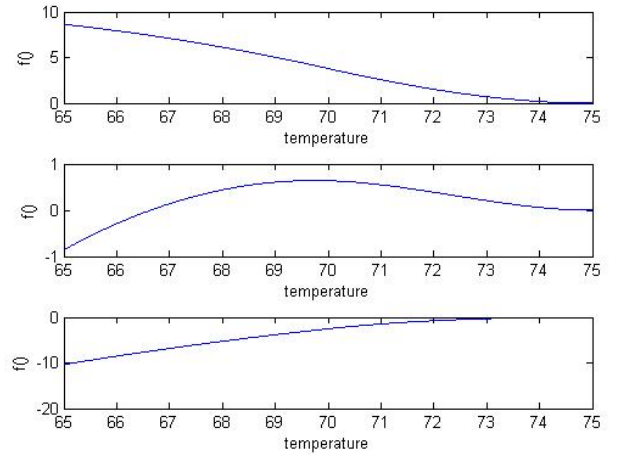


Fig. 6. Geometric features of $f(u, \Delta(i, \tilde{i}^d))$. (1) $f(u, \Delta(i, \tilde{i}^d))$ is monotonically decreasing when $\alpha\Delta(i+1, \tilde{i}^d) \leq 0$. Optimal policy is $u = T_{\min}$. (2) $f(u, \Delta(i, \tilde{i}^d))$ has unique global maximum when $\alpha\Delta(i+1, \tilde{i}^d) \in (0, E_\varepsilon[U(T_{\max}, b, \varepsilon)])$. The function may not be concave for $\hat{T} \neq T_{\max}$. (3) $f(u, \Delta(i, \tilde{i}^d))$ is monotonically increasing when $\alpha\Delta(i+1, \tilde{i}^d) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)]$. Optimal policy is $u = T_{\max}$.

IV. PROPERTIES OF THE OPTIMAL POLICY

Proposition 2 expresses the optimal policy $u(i, \tilde{i}^d)$ as a function of $\Delta(i, \tilde{i}^d)$. To study the properties of $u(i, \tilde{i}^d)$, we focus on the structure of $\Delta(i, \tilde{i}^d)$. In this section we derive key properties of $\Delta(i, \tilde{i}^d)$ in terms of the changes in state space variables that lead to desirable structures for $u(i, \tilde{i}^d)$. There are three state variables that affect $u(i, \tilde{i}^d)$: the aggregate

consumption over all active appliances $i(t)$, the ISO RSR signal $y(t)$, and the tracking error $e(t) = i(t) - \bar{n} - y(t)R$. When two of the three variables are given, the third variable can be expressed accordingly by $i(t) - y(t)R - e(t) = \bar{n}$ since \bar{n} is fixed. To study the structure of $\Delta(i, \tilde{i}^d)$ as a function of $i(t)$, $y(t)$, and $e(t)$, we fix one variable each time and allow the other two to vary. Before proceeding, we prove the following:

Lemma 1 Denote

$$\phi(\Delta(i+1, \tilde{i}^d)) = \max_{u \in [T_{\min}, T_{\max}]} p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d), \quad (25)$$

then $\phi(\Delta(i+1, \tilde{i}^d))$ is a monotonically non-increasing function.

Proof. For saturated optimal control $u(i, \tilde{i}^d) = T_{\min}$ or $u(i, \tilde{i}^d) = T_{\max}$, $p_{u(i, \tilde{i}^d)}$ and $U_{u(i, \tilde{i}^d)}$ are constant and the statements stand. When the optimal control is not saturated, namely $u(i, \tilde{i}^d) = I^{-1}(\alpha \Delta(i+1, \tilde{i}^d))$ as in the last scenario in Proposition 2, we have

$$\begin{aligned} & \frac{d\phi(\Delta(i+1, \tilde{i}^d))}{d\Delta(i+1, \tilde{i}^d)} \\ &= \left[\frac{dp_u U_u}{du} \Big|_{u=u(i, \tilde{i}^d)} - \alpha \Delta(i+1, \tilde{i}^d) \frac{dp_u}{du} \Big|_{u=u(i, \tilde{i}^d)} \right] \frac{du}{d\Delta(i+1, \tilde{i}^d)} \\ &= -\alpha p_{u(i, \tilde{i}^d)} \\ &= -[E[U(u(i, \tilde{i}^d), b, \epsilon)] - \alpha \Delta(i+1, \tilde{i}^d)] p(u(i, \tilde{i}^d)) \frac{du}{d\Delta(i+1, \tilde{i}^d)} \\ &= -\alpha p_u \\ &= -\alpha p_u \leq 0. \end{aligned}$$

Therefore $\phi(\Delta(i+1, \tilde{i}^d))$ is a monotonically non-increasing function of $\Delta(i+1, \tilde{i}^d)$. \square

In addition to the monotonicity properties of $\phi(\Delta(i+1, \tilde{i}^d))$, we derive upper and lower bounds on the change in $\phi(\Delta(i+1, \tilde{i}^d))$ with respect to a change in $\Delta(i+1, \tilde{i}^d)$ shown in the following Lemma:

Lemma 2 $\phi(\Delta(i+1, \tilde{i}^d)) - \phi(\Delta(i, \tilde{i}^d))$ has the following upper and lower bound:

- (1) $\phi(\Delta(i+1, \tilde{i}^d)) - \phi(\Delta(i, \tilde{i}^d)) \leq -\alpha p_{u(i, \tilde{i}^d)} (\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)).$
- (2) $\phi(\Delta(i+1, \tilde{i}^d)) - \phi(\Delta(i, \tilde{i}^d)) \geq -\alpha p_{u(i-1, \tilde{i}^d)} (\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)).$

Proof. (1). The proof is straightforward

$$\begin{aligned} & \phi(\Delta(i+1, \tilde{i}^d)) - \phi(\Delta(i, \tilde{i}^d)) \\ &= [p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d)] \Big|_{u=u(i, \tilde{i}^d)} \\ &\quad - [p_u U_u - \alpha p_u \Delta(i, \tilde{i}^d)] \Big|_{u=u(i-1, \tilde{i}^d)} \\ &\leq [p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d)] \Big|_{u=u(i, \tilde{i}^d)} \\ &\quad - [p_u U_u - \alpha p_u \Delta(i, \tilde{i}^d)] \Big|_{u=u(i, \tilde{i}^d)} \\ &= -\alpha p_{u(i, \tilde{i}^d)} (\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)). \end{aligned} \quad (26)$$

The inequality holds because $\phi(\Delta(i, \tilde{i}^d))$ is evaluated at $u = u(i, \tilde{i}^d)$ rather than at the optimal policy $u = u(i-1, \tilde{i}^d)$, and therefore it results in a higher than the optimal cost which yields an upper bound.

(2) Similarly

$$\begin{aligned} & \phi(\Delta(i+1, \tilde{i}^d)) - \phi(\Delta(i, \tilde{i}^d)) \\ &= [p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d)] \Big|_{u=u(i, \tilde{i}^d)} \\ &\quad - [p_u U_u - \alpha p_u \Delta(i, \tilde{i}^d)] \Big|_{u=u(i-1, \tilde{i}^d)} \\ &\geq [p_u U_u - \alpha p_u \Delta(i+1, \tilde{i}^d)] \Big|_{u=u(i-1, \tilde{i}^d)} \\ &\quad - [p_u U_u - \alpha p_u \Delta(i, \tilde{i}^d)] \Big|_{u=u(i-1, \tilde{i}^d)} \\ &= -\alpha p_{u(i-1, \tilde{i}^d)} (\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)). \end{aligned} \quad (27)$$

This inequality holds also by a similar argument. \square

Lemma 1 and Lemma 2 provide the monotonicity property as well as bounds on $\phi(\cdot)$ function differences between adjacent states. We next use these bounds to prove three monotonicity properties of $\Delta(i, \tilde{i}^d)$ with respect to state space parameter changes in RSR signal value y , current aggregated demand i , and tracking error e . Properties of $\Delta(i, \tilde{i}^d)$ will be used to prove the main Theorem on the structure of the optimal policy at the end of the section.

A. Monotonicity of $\Delta(i, \tilde{i}^d)$ for Key State Space Parameters

We first discuss the monotonicity of $\Delta(i, \tilde{i}^d)$ for a fixed ISO RSR signal y . In this case \tilde{i}^d will be fixed and i, e will change in the same direction. $\Delta(i, \tilde{i}^d)$ represents the optimal value difference between two adjacent states having only one consumption difference. Proposition 3 provides properties of $\Delta(i+1, \tilde{i}^d)$ when state space variable \tilde{i}^d is fixed while i varies.

Proposition 3 The following properties hold for a fixed y :

- (1) $\Delta(i+1, \tilde{i}^d) \geq \Delta(i, \tilde{i}^d) + \epsilon^l$, where $\epsilon^l = \frac{2\kappa}{1-\alpha[1-2(\lambda+\mu)-v]}$ with $v = \lambda(N - N_1)$.

- (2) $\Delta(i, \tilde{i}^d) + \epsilon^u \geq \Delta(i+1, \tilde{i}^d)$, where $\epsilon^u = \frac{2\kappa}{1-\alpha[1-2(\lambda+\mu)]}$.

Proof. See Appendix A.

Remark 4 Since both ϵ^l and ϵ^u are positive, we have $\Delta(i+1, \tilde{i}^d) > \Delta(i, \tilde{i}^d)$. The optimal value function $J(i, \tilde{i}^d)$ exhibits convex-like behavior for a given \tilde{i}^d , in the sense that

$$J(i+1, \tilde{i}^d) + J(i-1, \tilde{i}^d) > 2J(i, \tilde{i}^d). \quad (28)$$

This convexity property can be used to design approximate DP (ADP) algorithms with convex functional approximation. We explore this possibility in Sec. V.

We next discuss the monotonicity of $\Delta(i, \tilde{i}^d)$ for a fixed tracking error e with \tilde{i}^d and i changing accordingly. Since one of our objectives is to accurately track the ISO RSR signal, it is reasonable to speculate that the SBO would use the same optimal policy for states $\{i, \tilde{i}^d\}$ and $\{i+1, \tilde{i}^d + \Delta y\}$ that have the same e , and therefore it is reasonable to have $\Delta(i+1, \tilde{i}^d + \Delta y) = \Delta(i+1, \tilde{i}^d + \Delta y)$. However, this speculation ignores the fact that the expected consumer arrival rate and the expected period utility reward will be different if same policy is used since the queuing system is closed and the total number of appliances is finite (different current consumption i infers different number of counterpart idle appliances). We formally investigate the properties of $\Delta(i, \tilde{i}^d)$ for a given e and state the monotonicity properties as follows.

Proposition 4 The following properties hold for a fixed e :

- (1) $\Delta(i, \tilde{i}^d) \geq \Delta(i+1, \tilde{i}^d + \Delta y)$.
- (2) $\Delta(i+1, \tilde{i}^d + \Delta y) + \bar{\epsilon}^u \geq \Delta(i, \tilde{i}^d)$ where $\bar{\epsilon}^u = \frac{\alpha(\lambda+\mu)}{1-\alpha[1-(\lambda+\mu)]} \epsilon^u$.

The proof of Proposition 4 is similar to the proof of Proposition 3. We omit the proof due to page limitation.

In the end, we derive a last property when the aggregated consumption i is fixed while the ISO RSR signal y and tracking error $e = i - \bar{i} - yR$ change in the same direction.

Proposition 5 The following properties hold for a fixed i :

- (1) $\Delta(i, \tilde{i}^d) \geq \Delta(i, \tilde{i}^d + \Delta y) + \epsilon^l$.
- (2) $\Delta(i, \tilde{i}^d + \Delta y) + \epsilon^u + \bar{\epsilon}^u \geq \Delta(i, \tilde{i}^d)$.

Proof. We have

$$\Delta(i, \tilde{i}^d) \geq \Delta(i+1, \tilde{i}^d + \Delta y) \geq \Delta(i, \tilde{i}^d + \Delta y) + \varepsilon^l, \quad (29)$$

where the first (second) inequality is the direct result of Proposition 4 (3). And similarly the second part of the proposition holds. \square

The above proposition completes our discussion of the properties of the differential cost function $\Delta(i, \tilde{i}^d)$. These properties and the relation between $\Delta(i, \tilde{i}^d)$ and $u(i, \tilde{i}^d)$ result in the following useful properties of the optimal policy in the next section.

B. Monotonicity Properties of the Optimal Policy $u(i, \tilde{i}^d)$

Based on Propositions 3, 4, and 5, we present the following theorem as the main result illustrating the monotonicity properties of the optimal policy $u(i, \tilde{i}^d)$.

Theorem 1 The following properties hold for the state feedback optimal policy $u(i, \tilde{i}^d)$ for all $\{i, \tilde{i}^d\}$:

(1) For the same RSR signal, the optimal price policy is a monotonically non-decreasing function of i . Namely $u(i+1, \tilde{i}^d) \geq u(i, \tilde{i}^d)$.

(2) For the same tracking error, the optimal price policy is a monotonically non-increasing function of i . Namely $u(i, \tilde{i}^d) \geq u(i+1, \tilde{i}^d + \Delta y)$.

(3) For the same consumption level, the optimal price policy is a monotonically non-increasing function of \tilde{i}^d . Namely $u(i, \tilde{i}^d) \geq u(i, \tilde{i}^d + \Delta y)$.

Proof. The proof is straightforward. From Proposition 3, 4, and 5 we have

$$\Delta(i, \tilde{i}^d) \geq \Delta(i+1, \tilde{i}^d + \Delta y) \geq \Delta(i, \tilde{i}^d + \Delta y).$$

From Proposition 2 the optimal control is a non-decreasing function of the $\Delta(i, \tilde{i}^d)$, therefore

$$u(i, \tilde{i}^d) \geq u(i+1, \tilde{i}^d + \Delta y) \geq u(i, \tilde{i}^d + \Delta y),$$

and the three statements above are true. \square

Remark 5 The policy monotonicity properties are valid for both $D(t) = 1$ and $D(t) = -1$. Bang-bang optimal control will be used when greater imbalance exists between the aggregated consumption and ISO regulation signal level. When the state is at a high value of i and a low value of \tilde{i}^d , the ISO would broadcast highest price signal. When the state has a small value of i and a large \tilde{i}^d , the SBO would broadcast a lower price signal. Otherwise, the SBO would broadcast the optimal price in between.

Remark 6 The three monotonicity properties can be interpreted as follows: we set a low price for states with a higher RSR signal, and a high price for states with a higher aggregate consumption. When the tracking error is the same, a high current consumption suggests a smaller number of disconnected appliances, which means we have a smaller number of appliances that are able to respond to the SBO's signal. In such cases, a lower price is chosen to achieve a larger percentage of disconnected appliances.

C. Monotonicity of Price Sensitivity w.r.t. N

We investigate the change in the optimal price sensitivity with respect to the number of connected appliances i , as the total number of appliances N and the maximum Reserve obligation R increase both in a constant proportion. Consider a SBO who provides regulation reserves R equal to a fixed proportion q of the duty cycle appliances N ³, namely $R = qN$. Note that q can be determined by cooling appliance user preferences, appliance technical specifications, etc [32]. As N and R increase at the same rate, the effective penalty parameter $\kappa = K/R^2$ will decrease. In addition, the uniformized policy update interval Δ_t and discount factor α will change as N increases following the relationships:

$$\Delta_t = \frac{1}{N \max(\lambda, \mu) + \gamma} \approx \frac{1}{N \max(\lambda, \mu)}, \quad (30)$$

and

$$\alpha = \frac{1}{1 + r\Delta_t} \quad (31)$$

where r is the prevailing discount rate. Substituting (30) into (31) we can write

$$\alpha = \frac{N \max(\lambda, \mu)}{r + N \max(\lambda, \mu)}. \quad (32)$$

Observing that both the discount rate and the policy update period increase as N increases, we show by Proposition 6 that the change in the value function differentials $\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)$ and $\Delta(i, \tilde{i}^d) - \Delta(i, \tilde{i}^d + \Delta y)$ approaches zero as N approaches infinity.

Proposition 6 ε'' and ε^l , defined in Proposition 3, and $\bar{\varepsilon}''$, defined in Proposition 4, will decrease as N increases, and moreover for $N \rightarrow \infty$, their asymptotic limit is 0.

Proof. Using explicitly Δ_t which for notational simplicity was selected as the time unit and set equal to 1 in the proof of Proposition 3, we can write

$$\varepsilon'' = \frac{2\kappa\Delta_t}{1 - \alpha[1 - 2(\lambda + \mu)\Delta_t]}. \quad (33)$$

Substituting into (33) the effective discount factor α and the relation $\Delta_t \approx 1/(N \max(\lambda + \mu))$ we obtain

$$\varepsilon''(N) = \frac{\frac{2K/(qN)^2}{N \max(\lambda, \mu)}}{1 - \frac{N \max(\lambda, \mu)}{r + N \max(\lambda, \mu)} [1 - 2(\lambda + \mu) \frac{1}{N \max(\lambda, \mu)}]} \quad (34)$$

which in turn simplifies to $\varepsilon''(N) = \frac{2K/(qN)^2}{r + 2(\lambda + \mu)}$ verifying that ε'' decreases as N increases. It can be similarly shown that ε^l also decreases as N increases. Finally $\bar{\varepsilon}''$ is shown below to equal a positive multiple of ε''

$$\begin{aligned} \bar{\varepsilon}'' &= \frac{\alpha(\lambda + \mu)\Delta_t}{1 - \alpha[1 - (\lambda + \mu)\Delta_t]} \varepsilon'', \\ &= \frac{\frac{N \max(\lambda, \mu)}{r + N \max(\lambda, \mu)} \Delta_t}{1 - \frac{N \max(\lambda, \mu)}{r + N \max(\lambda, \mu)} [1 - (\lambda + \mu) \frac{1}{N \max(\lambda, \mu)}]} \varepsilon'', \\ &= \frac{1}{r + \lambda + \mu} \varepsilon''. \end{aligned} \quad (35)$$

³Using the average consumption rate of an active appliance as the unit that measures capacity, N equals the maximal consumption rate of the cooling appliances under the SBO's control.

We can now conclude that all three parameters $\bar{\epsilon}^u$, ϵ^u and ϵ^l will approach zero as N goes to infinity. \square

Proposition 6 describes the asymptotic impact of building size described by N on $\Delta(i, \tilde{i}^d)$ and the optimal price policy $u(i, \tilde{i}^d)$. According to Propositions 3 and 5, the difference between differential cost functions for fixed \tilde{i}^d and i respectively is bounded by

$$\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d) \in [\epsilon^l, \epsilon^u] \quad (36)$$

and

$$\Delta(i, \tilde{i}^d) - \Delta(i, \tilde{i}^d + \Delta y) \in [\epsilon^l, \epsilon^u + \bar{\epsilon}^u]. \quad (37)$$

which by Proposition 6 implies that these differences go to 0.

Using the expression for the optimal policy proven in Proposition 2, we can conclude that $u(i, \tilde{i}^d)$, $u(i+1, \tilde{i}^d)$, $u(i, \tilde{i}^d)$ and $u(i, \tilde{i}^d + \Delta y)$ get closer together as N increases, and as a result the optimal policy function becomes flatter with respect to its arguments.

V. NUMERICAL SOLUTION ALGORITHMS

The analytical results presented so far are not merely exercises in analysis that capture abstract properties of the DP optimality conditions. Most notably, the optimal policy structure of Proposition 2 and the monotonicity and second derivative related properties proven in Propositions 3 to 5 are valuable ammunition that enable design and implementation of efficient and scalable numerical solution algorithms. This section demonstrates the value of the analytical results in doing just that and provides elaborative computational results.

A. Value Iteration Based Approaches

We first propose and implement two numerical DP solution algorithms, the first for benchmarking and comparison purposes using the conventional value iteration (CVI) approach [27], and the second by leveraging the optimal policy structure proven in Proposition 2 of Section IV 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 (20) because it is optimal for a given value function resembling policy iteration algorithms. Our AVI algorithm recognizes that the state space is discrete while the policy space is continuous. It benefits from (1) the analytic characterization of the optimal policy in terms of the current iteration estimate of the value function thus avoiding both state space discretization and exhaustive search for the optimal policy, and (2) avoidance of the sub-optimality gap introduced by the policy space discretization.

Numerical results from the CVI and AVI algorithms are shown in Fig. 7. In the upper sub-figure, the parameter values used were $N = 200$, $\bar{n} = 100$, $R = 20$, $\lambda = 2$, $\mu = 0.5$. We choose a linear utility function as in (10) with $b = 20$. We find that the CVI algorithm yields policies selected from the discretized allowable policy set and the AVI algorithm provides a smooth and continuous policy. The observed price monotonicity are consistent with properties derived in Theorem 1. The comparison between the two sub-figures demonstrates the price sensitivity when we increase N and R to the

same proportion. Note that the rate at which the optimal price increases from $u = T_{\min}$ to $u = T_{\max}$ decreases, unsurprisingly, by a factor of 2. Another interesting observation is that when N increases, different curves for ISO signals y get closer to each other for a fixed i . This is consistent with our analysis of the monotonicity of price sensitivity in Sec. IV-C.

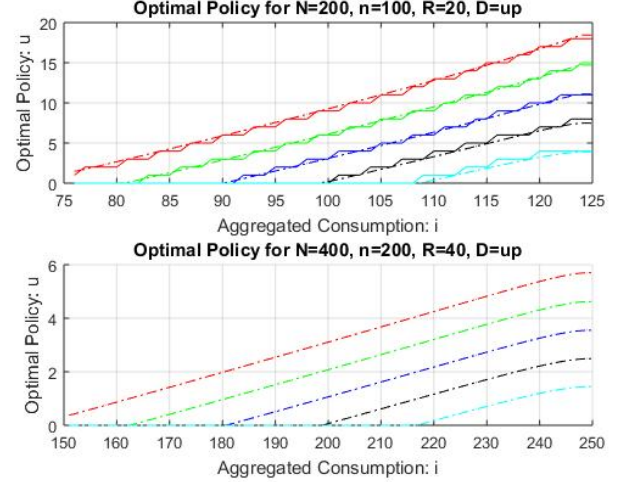


Fig. 7. Optimal policies obtained from the CVI and AVI Algorithms. The solid (dashed) lines are simulation results of CVI (AVI) algorithm. The red, green, dark black, black, and light black curves correspond to RSR signal level at -0.8 , -0.4 , 0.0 , 0.4 , 0.8 , respectively. In CVI, the policy space is discretized into 20 possible prices corresponding to the temperature threshold from 1 to 20. In AVI, the policy space is continuous. Monotonicity properties in Theorem 1 are verified. Equal $Y = yR$ policy functions demonstrate that the slope of the price between $u = T_{\min}$ to $u = T_{\max}$ decreases as N and R increase. The vertical distance among price policy lines also decreases.

B. Functional Approximate DP Approach

We proceed to propose a numerical solution algorithm based on an analytic functional approximation of the value function $J(i, y, D)$. This algorithm leverages the properties of value function first and second differences derived in Sec. IV. In particular we use Proposition 3 which shows that $\Delta(i+1, \tilde{i}^d) \geq \Delta(i, \tilde{i}^d)$. Given the discrete state space of our problem, this property is equivalent to convexity of $J(i, \tilde{i}^d)$ in the number of active appliances i for a given pair of ISO signal's value and direction.

In addition, we note that from Fig. 7 that the increased rate of the optimal policy for a fixed RSR signal is approximately a constant value k . Therefore we approximately have

$$u(i+1, \tilde{i}^d) - u(i, \tilde{i}^d) = k, \quad (38)$$

which is equivalent to have $\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)$ being some constant. Since $\Delta(i, \tilde{i}^d)$ is the differential of the value function with respect to i , it means that the second order differential of the value function with respect to i is approximately constant, namely $\frac{\partial^2}{\partial i^2} J(i, y, D)$ is approximately constant. Similarly, from Fig. 7 we note that the rate of policy's vertically changes is constant for varying RSR signals, therefore the $\frac{\partial}{\partial \tilde{i}^d} \Delta(i, \tilde{i}^d)$ is approximately constant, namely $\frac{\partial}{\partial \tilde{i}^d} \frac{\partial}{\partial i} J(i, \tilde{i}^d)$ is constant.

These properties motivate an approximation of $J(i, \tilde{i}^d)$ by $\hat{J}_d(i, y)$ that is quadratic in $i - yR$. In fact, we treat $D(t)$ as

a binary argument and propose the following basis function approximation when $D(t) = -1$.

$$\hat{f}_d(i, y) = r_{11}(i - yR)^2 + r_{12}(i - yR) + r_{13} \quad (39)$$

with $r_{11} > 0$ to guarantee convexity. In addition, the differential of the value function with respect to i , namely $\Delta(i, \bar{i})$, is

$$\frac{\partial}{\partial i} \hat{f}_d(i, y) = \Delta(i, \bar{i}) = 2r_{11}(i - yR) + r_{12}i. \quad (40)$$

Proposition 4 proved that $\Delta(i, \bar{i}^d) \geq \Delta(i + 1, \bar{i}^d + \Delta y)$ which suggests that $\frac{\partial}{\partial i} \hat{f}_d(i, y)$ monotonically decreases as a function of i for a fixed $e = i - yR$, hence $r_{12} < 0$ in (40).

We generalize the approximation in (39) by differentiating the parameters depending on the discrete value of the direction D . Thus we define for $D = -1$

$$\hat{f}_d(i, y, \mathbf{r}_d) = r_1 i^2 + r_2 i + r_3 y^2 + r_4 y + r_5 i y + r_6 \quad (41)$$

and the corresponding function $\hat{f}_u(i, y, \mathbf{r}_u)$ for $D = 1$. The value function $J(i, y, D, \mathbf{r})$ is then approximated by:

$$\hat{f}(i, y, D, \mathbf{r}) = \mathbb{1}_{\{D=1\}} \hat{f}_u(i, y, \mathbf{r}_u) + \mathbb{1}_{\{D=-1\}} \hat{f}_d(i, y, \mathbf{r}_d). \quad (42)$$

The two components of the expression in (42) approximate state features associated with increasing or decreasing ISO signals. The vector \mathbf{r} is a vector of twelve parameters six from \mathbf{r}_d and \mathbf{r}_u each. Written in matrix form, (42) is equivalent to the following

$$\hat{f} = \Phi \mathbf{r}, \quad (43)$$

where Φ is a $|N| \times |y| \times |D|$ by 12 matrix with rows being the feature vector for each state. The functional approximation is therefore transformed into the problem of solving the projected Bellman equation

$$\Phi \mathbf{r} = \Pi T(\Phi \mathbf{r}), \quad (44)$$

where T is the operator of the form $TJ = g + \alpha PJ$ with state transition matrix P . Π is the projection operator onto the set spanned by the basis functions $S = \{\Phi \mathbf{r} | \mathbf{r} \in \mathbb{R}^{12}\}$. It is shown in [27] that the solution to the above projection problem is given by

$$\mathbf{r}^* = C^{-1} d, \quad (45)$$

where $C = \Phi' \Xi (I - \alpha P) \Phi$, $d = \Phi' \Xi g$, and Ξ is the matrix with diagonal elements being the steady state probability distribution of the states. It is further shown that (45) can be solved in an iterative form by the projected value iteration (PVI) algorithm

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \gamma G_k (C_k \mathbf{r}_k - d_k), \quad (46)$$

where C_k and d_k are given by

$$C_k = \frac{1}{k+1} \sum_{t=0}^k \phi(i_t, y_t, D_t) (\phi(i_t, y_t, D_t) - \alpha \phi(i_{t+1}, y_{t+1}, D_{t+1})),$$

$$d_k = \frac{1}{k+1} \sum_{t=0}^k \phi(i_t, y_t, D_t) g(i_t, y_t, D_t, u(t)). \quad (47)$$

To choose γ and G_k , it is proposed to have $\gamma = 1$ and

$$G_k = \left(\frac{1}{k+1} \sum_{t=0}^k \phi(i_t, y_t, D_t) \phi(i_t, y_t, D_t)' \right)^{-1}. \quad (48)$$

Based on the above approach that finds a good approximation of the value function for a *fixed* policy, we successfully use the following algorithm to construct a good approximation of the value function as well as the *optimal* policy based on sample trajectories obtained by Monte Carlo simulation. The algorithm contains the following four steps:

ADP Algorithm

Step 1. Initialization $\mathbf{r} = 0$.

Step 2. Initialization $\mathbf{r}_{old} = \mathbf{r}$, $\mathbf{r}_0 = \mathbf{r}$, $k = 0$, $\{i_k, y_k, D_k\}$.

Step 3. Generate

Optimal policy $u_{\mathbf{r}_{old}}(i_k, y_k, D_k)$

Next state $\{i_{k+1}, y_{k+1}, D_{k+1}\}$

Period cost $g(i_k, y_k, D_k, u_{\mathbf{r}_{old}}(i_k, y_k, D_k))$

Update

C_k, d_k , and \mathbf{r}_{k+1} based on (46)–(48)

If $k \geq k_{min}$ and $\|\mathbf{r}_{k+1} - \mathbf{r}_k\|_2 < \bar{\epsilon}$

$\mathbf{r} = \mathbf{r}_{k+1}$, go to **Step 4**.

Else

$k = k + 1$, go to **Step 3**.

Step 4. If $\|J(i, y, D, \mathbf{r}) - J(i, y, D, \mathbf{r}_{old})\|_\infty < \tau$

return $\mathbf{r}^* = \mathbf{r}$. Algorithm ends.

Else go to **Step 2**.

The algorithm starts with an initial guess of the parameters, $\mathbf{r} = 0$. Step 2 initializes the iteration count, parameters \mathbf{r}_0 , and state variables $\{i_0, y_0, D_0\}$. Step 3 iteratively updates the value function for the fixed policy $u_{\mathbf{r}_{old}}$ using the PVI algorithm described above. Step 3 is repeated for at least a minimum number of iterations, $k \geq k_{min}$, and stops when the change in \mathbf{r} meets a desired tolerance, $\|\mathbf{r}_{k+1} - \mathbf{r}_k\|_2 < \bar{\epsilon}$. Steps 4 compares the value function parametrized by \mathbf{r}_{old} and the updated \mathbf{r} obtained by step 3. If the infinity norm of the vector is less than the threshold τ , then the value function converges and the algorithm returns the optimal parameter $\mathbf{r}^* = \mathbf{r}$. Else, it returns to step 2 for a new iteration.

C. Comparison between Value Iteration Algorithms and the ADP

We compare the computational performance of the CVI, AVI and Functional ADP algorithms for different state space size problems in Table I. Based on the optimal condition derived in Proposition 2, the AVI algorithm is effective in reducing the computational time by approximately 90% since the optimal policy per state and per iteration is calculated on the fly based on the current value function. However, it is not fast for large problems up to 400,000 states since it needs more than 2 hours to solve for the optimal policy. Considering that the RSR is bid and served for every one hour, the AVI algorithm may not be practical for real time implementation, especially when the RSR provision capacity of the energy provider is huge. However, the functional approximation based ADP algorithm further reduces the computational time by more than 90% from the AVI algorithm. In fact, in the inner loop described in Algorithm 1, the number of states visited in Monte Carlo simulation is approximately 10% of all the

states. As for the outer loop, it also needs few iterations for the convergence of the value function compared to the AVI algorithm. Therefore the total computational time is reduced for more than 90%.

TABLE I
COMPUTATIONAL PERFORMANCES OF THE THREE ALGORITHMS

Problem Size ($ N * y * D $)	50*21*2	500*41*2	5000*41*2
CVI Computation Time (sec)	200.35	6761.2	66082
AVI Computation Time (sec)	20.14	596.65	9489.5
ADP Computation Time (sec)	2.69	34.2	514.8

We examine the convergence result of the proposed algorithm. The coefficient vector \mathbf{r} is composed of 12 parameters, six for each direction $D = 1, -1$. Fig. 8 shows the convergence result of the six parameters for direction $D = -1$ proposed by the ADP algorithm. We find all parameters would converge after 10 iterations. The convergence of \mathbf{r} also indicates the convergence of the value function \hat{J} .

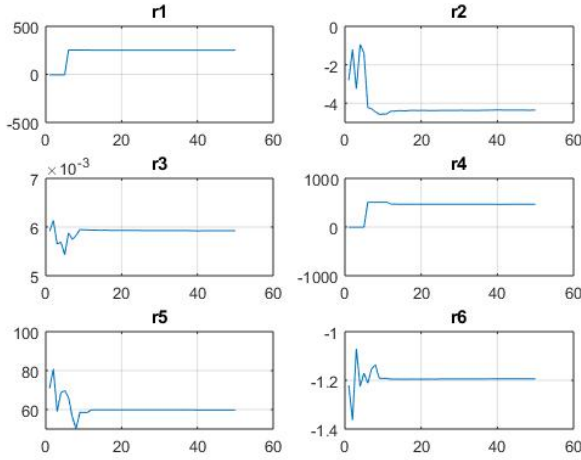


Fig. 8. Convergence of the feature parameters in (41). The coefficients of the six basis converge after 10 iterations. Convergence of the feature parameters indicates the convergence of the value function approximation \hat{J} .

Fig. 9 compares the converged value function and the corresponding optimal policy generated by the AVI and the functional ADP. Left column figures are plots of the value function generated by the AVI and the functional ADP algorithm corresponding to ISO RSR signal direction $D = 1$. The functional ADP algorithm learns the convex structure of the value function accurately. The error between $\hat{J}(i, y, D, \mathbf{r})$ and $J(i, y, D)$ is relatively small. Right column figures compare the optimal policies of Proposition 2 that are generated by the value functions based on the functional ADP and the AVI algorithm. The functional ADP algorithm performs well relative to the AVI algorithm that derives the true optimal policy.

D. RSR Provision Quality with Dynamic Preference

We validate the tracking performance of the ADP algorithm optimal control with RSR signals from PJM. To illustrate the improvement gained by having dynamic preference, we

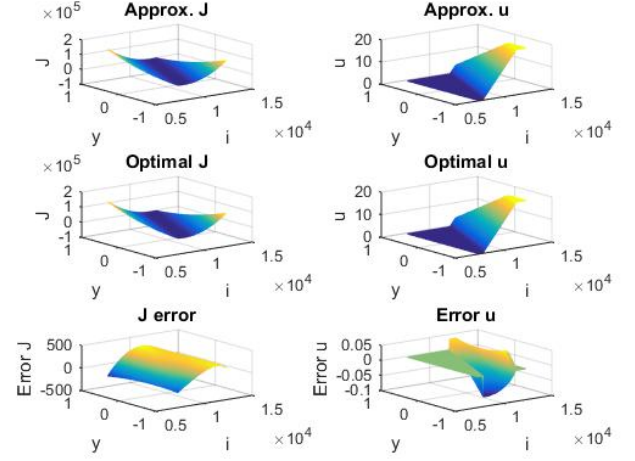


Fig. 9. Left column figures compare at the top the value functions obtained by the functional approximation DP algorithm and the AVI algorithm, and plot the error at the bottom. Right column figures plot the corresponding optimal policy and the error amongst them. Results indicates that the functional approximation DP algorithm captures the properties proven in section IV.

compare results to RSR provision where the optimal policy is generated based on the assumption of *time invariant* consumers' preference, namely to assume that consumers' preferences distribution is statically uniformly distributed regardless of time. In this simulation we assume that there are 20,000 appliances within the aggregator each consuming 1kW of energy. The aggregator has mean consumption of 10,000 kW and provide 2,000 kW up (down) reserve to the market, namely 10% of its total capacity. Fig. 10 shows the comparison of the tracking performance between the two assumptions. The red curve is the reference PJM signal at a frequency of 4 seconds. The black curves are the tracking performance (upper figure) and the tracking error (lower figure) that result from the dynamic preferences optimal control. The green curves are results from the static preferences optimal control. In general, both approaches are effective in providing demand side RSR with tracking error controlled within 5%. Nonetheless, the tracking provided by dynamic preference assumption achieves higher accuracy. The accuracy can be seen based on the statistics in Table II. The mean tracking error is decreased from 86MW to 79MW if dynamic preference is applied, namely a mean error reduction of 8%. The min and max tracking error spikes are also decrease by 12% and 17%, respectively. As to the aggregator, the sum of squared tracking penalty is decreased significantly by 9.4%.

Converting tracking error to the system operation perspective, with more accurate dynamic preference modeling, the system will have 8% less total reserve generation and 17% less peak reserve generation from traditional thermal units. These will provide significant operation values since these capacity can otherwise be committed as energy, rather than reserve, and thus possibly lower the overall energy costs across the system either in the day ahead or real time market clearing.

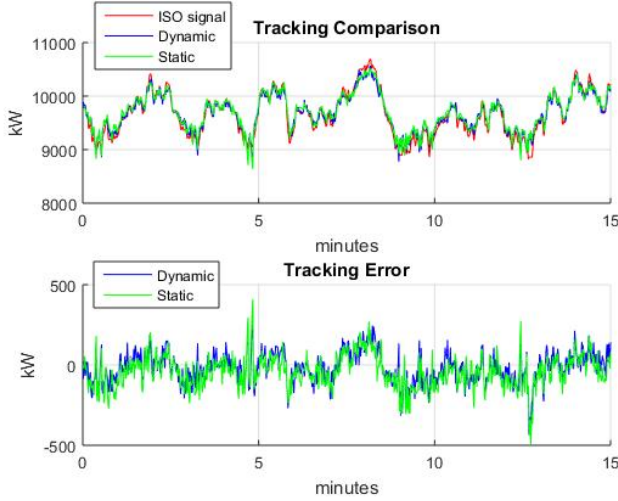


Fig. 10. Comparison of the optimal control generated based on the assumption of dynamic preference vs. static uniform preference. In both assumptions, the optimal control can track the trend of RSR signal. Dynamic preferences optimal control tracks with better statistical performances of the tracking error (mean, standard deviation, max, and min) than the uniform preference optimal control.

TABLE II
DYNAMIC VS. STATIC PREFERENCE RSR PROVISION QUALITY

Statistics	Dynamic Preference	Static Preference
Mean Error (kW)	79.18	86.26
Std.Dev. Error (kW)	62.31	68.25
Min Error (kW)	-437	-492
Max Error (kW)	342	410
Sum of Squared Penalty	913330	1008400

VI. CONCLUSION

This paper relaxes a common, but unrealistic assumption in the dynamic pricing literature, which, for the sake of simplifying the analysis of the resulting problem formulation, claims that it is reasonable to approximate the preferences of market participants with a static, usually uniform, distribution that is independent of control history. We show that a dynamically evolving trapezoidal 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 believe that dynamic, control driven evolution of preferences can be modeled and analyzed in more general contexts. Under preference dynamics modeling, we derive an analytic expressions of the optimal policy, monotonicity properties of the value function, and the behavior of the optimal policy. We also prove the existence of policy sensitivity bounds and their asymptotic convergence as the number of the duty cycle appliances increase. The aforementioned analytical results prove invaluable in guiding us to design and implement efficient and scalable numerical solution algorithms. Simulation based on PJM signals shows the advantage of RSR provision under dynamic preference modeling over static modeling assumptions. In future work we will consider to prove the convergence of the ADP algorithm, to embed additional sufficient statistics into the dynamic preference, and

to model imperfect state observation where the utility function or the preferences cannot be observed.

APPENDIX A PROOF OF PROPOSITION 3

Proof. (1) For a sequence of $J_0(i, \tilde{i}^d), \dots, J_k(i, \tilde{i}^d)$ generated by value iteration, we have $\lim_{k \rightarrow \infty} J_k(i, \tilde{i}^d) = J(i, \tilde{i}^d)$ based on the value iteration convergence property. We define the differential of the value function at the k^{th} iteration as

$$\Delta_k(i, \tilde{i}^d) = J_k(i, \tilde{i}^d) - J_k(i-1, \tilde{i}^d).$$

It follows that $\lim_{k \rightarrow \infty} \Delta_k(i, \tilde{i}^d) = \Delta(i, \tilde{i}^d)$.

We assume $\Delta_k(i+1, \tilde{i}^d) \geq \Delta_k(i, \tilde{i}^d) + \varepsilon_k^l$ and $\Delta_k(i+1, \tilde{i}^u) \geq \Delta_k(i, \tilde{i}^u) + \varepsilon_k^l$ with $\varepsilon_k^l = 0$ for all $\{i, \tilde{i}^d\}$, $\{i, \tilde{i}^u\}$ and $k=0$, which holds trivially when at the initial iteration the value function is taken to equal zero. At iteration $k+1$, $J_{k+1}(i, \tilde{i}^d)$ can be written using the Bellman equation as

$$J_{k+1}(i, \tilde{i}^d) = g(i, \tilde{i}^d) - \lambda(N-i)\phi(\Delta_k(i+1, \tilde{i}^d)) + \alpha \{ [1 - (\gamma_1^d + \gamma_2^d)] J_k(i, \tilde{i}^d) - \mu i \Delta_k(i, \tilde{i}^d) + (\gamma_1^d J_k(i, \tilde{i}^u + \Delta y) + \gamma_2^d J_k(i, \tilde{i}^d - \Delta y)) \}, \quad (49)$$

Starting with the definition of $\Delta_k(i+1, \tilde{i}^d)$, we can write

$$\begin{aligned} & \Delta_{k+1}(i+1, \tilde{i}^d) \\ &= J_{k+1}(i+1, \tilde{i}^d) - J_{k+1}(i, \tilde{i}^d), \\ &= [g(i+1, \tilde{i}^d) - g(i, \tilde{i}^d)] + \alpha \{ [1 - \gamma_1^d - \gamma_2^d] \Delta_k(i+1, \tilde{i}^d) - \mu(i+1) \Delta_k(i+1, \tilde{i}^d) + \mu i \Delta_k(i, \tilde{i}^d) + \gamma_1^d \Delta_k(i+1, \tilde{i}^u + \Delta y) + \gamma_2^d \Delta_k(i+1, \tilde{i}^d - \Delta y) \} \\ & \quad - \lambda(N-i-1)\phi(\Delta_k(i+2, \tilde{i}^d)) + \lambda(N-i)\phi(\Delta_k(i+1, \tilde{i}^d)), \end{aligned}$$

This can be used to derive the change in $\Delta_{k+1}(i, \tilde{i}^d)$ when i increases by one,

$$\begin{aligned} & \Delta_{k+1}(i+1, \tilde{i}^d) - \Delta_{k+1}(i, \tilde{i}^d) \\ &= g(i+1, \tilde{i}^d) - 2g(i, \tilde{i}^d) + g(i-1, \tilde{i}^d) + \alpha \{ (1 - \gamma_1^d - \gamma_2^d) [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)] - \mu(i+1) [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)] + \mu(i-1) [\Delta_k(i, \tilde{i}^d) - \Delta_k(i-1, \tilde{i}^d)] + \gamma_1^d [\Delta_k(i+1, \tilde{i}^u + \Delta y) - \Delta_k(i, \tilde{i}^u + \Delta y)] + \gamma_2^d [\Delta_k(i+1, \tilde{i}^d - \Delta y) - \Delta_k(i, \tilde{i}^d - \Delta y)] \} - \lambda(N-i-1) [\phi(\Delta_k(i+2, \tilde{i}^d)) - \phi(\Delta_k(i+1, \tilde{i}^d))] + \lambda(N-i+1) [\phi(\Delta_k(i+1, \tilde{i}^d)) - \phi(\Delta_k(i, \tilde{i}^d))]. \quad (50) \end{aligned}$$

From Lemma 2,

$$\phi(\Delta_k(i+2, \tilde{i}^d)) - \phi(\Delta_k(i+1, \tilde{i}^d)) \leq -\alpha p_{u(i+1, \tilde{i}^d)} [\Delta(i+2, \tilde{i}^d) - \Delta(i+1, \tilde{i}^d)],$$

$$\phi(\Delta_k(i+1, \tilde{i}^d)) - \phi(\Delta_k(i, \tilde{i}^d)) \geq -\alpha p_{u(i-1, \tilde{i}^d)} [\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)].$$

We substitute the above two inequalities into (50),

$$\begin{aligned}
& \Delta_{k+1}(i+1, \tilde{i}^d) - \Delta_{k+1}(i, \tilde{i}^d) \\
\geq & g(i+1, \tilde{i}^d) - 2g(i, \tilde{i}) + g(i-1, \tilde{i}^d) \\
& \alpha \{ [1 - \gamma_1^d - \gamma_2^d - \mu(i+1)] [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)] \\
& + \mu(i-1) [\Delta_k(i, \tilde{i}^d) - \Delta_k(i-1, \tilde{i}^d)] \\
& + \gamma_1^d [\Delta_k(i+1, \tilde{i}^u + \Delta y) - \Delta_k(i, \tilde{i}^u + \Delta y)] \\
& + \gamma_2^d [\Delta_k(i+1, \tilde{i}^d - \Delta y) - \Delta_k(i, \tilde{i}^d - \Delta y)] \} \\
& - \lambda(N-i+1) p_{u(i-1, \tilde{i}^d)} [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)] \\
& + \lambda(N-i-1) p_{u(i+1, \tilde{i}^d)} [\Delta_k(i+2, \tilde{i}^d) - \Delta_k(i+1, \tilde{i}^d)] \\
\geq & 2\kappa + \alpha[1 - 2(\lambda + \mu) - \lambda(N-i)(p_{u(i-1, \tilde{i}^d)} - p_{u(i+1, \tilde{i}^d)})] \varepsilon_k^l \\
\geq & 2\kappa + \alpha[1 - 2(\lambda + \mu) - \lambda(N-N_1)] \varepsilon_k^l
\end{aligned} \tag{51}$$

The second inequality holds since the following four terms are greater or equal than ε_k^l based on the assumption at iteration k : $\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)$, $\Delta_k(i, \tilde{i}^d) - \Delta_k(i-1, \tilde{i}^d)$, $\Delta_k(i+1, \tilde{i}^u + \Delta y) - \Delta_k(i, \tilde{i}^u + \Delta y)$, $\Delta_k(i+1, \tilde{i}^d - \Delta y) - \Delta_k(i, \tilde{i}^d - \Delta y)$.

Denote $v = \lambda(N-N_1)$ and $\varepsilon_{k+1}^l = 2\kappa + \alpha[1 - 2(\lambda + \mu) - v] \varepsilon_k^l$. It can be now seen that

$$\Delta_{k+1}(i+1, \tilde{i}^d) - \Delta_{k+1}(i, \tilde{i}^d) \geq \varepsilon_{k+1}^l \tag{52}$$

holds for all $\{i, \tilde{i}^d\}$ at iteration $k+1$. Similarly we can prove

$$\Delta_{k+1}(i+1, \tilde{i}^u) - \Delta_{k+1}(i, \tilde{i}^u) \geq \varepsilon_{k+1}^l. \tag{53}$$

By mathematical induction, it is easy to show that

$$\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d) \geq \varepsilon_k^l, \Delta_k(i+1, \tilde{i}^u) - \Delta_k(i, \tilde{i}^u) \geq \varepsilon_k^l \tag{54}$$

holds for all k in the infinite series ε_k^l generated recursively by

$$\varepsilon_{k+1}^l = 2\kappa + \alpha[1 - 2(\lambda + \mu) - v] \varepsilon_k^l. \tag{55}$$

Since $\varepsilon_0^l = 0$, ε_k^l must converge for $k \rightarrow \infty$. In fact, it converges to ε^l with

$$\varepsilon^l = \lim_{k \rightarrow \infty} \varepsilon_k^l = \frac{2\kappa}{1 - \alpha[1 - 2(\lambda + \mu) - v]} \tag{56}$$

And we can hence conclude that

$$\begin{aligned}
& \Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d) \\
& = \lim_{k \rightarrow \infty} [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)], \\
& \geq \lim_{k \rightarrow \infty} \varepsilon_k^l = \varepsilon^l.
\end{aligned}$$

(2) Assuming $\Delta(i, \tilde{i}^d) + \varepsilon_k^u \geq \Delta(i+1, \tilde{i}^d)$ holds with $\varepsilon_k^u = 0$ for all $\{i, \tilde{i}^d\}$ and $k = 0$, Lemma 2 implies

$$-(\phi(\Delta(i+2, \tilde{i}^d)) - \phi(\Delta(i+1, \tilde{i}^d))) \leq \alpha p_{u(i, \tilde{i}^d)} [\Delta(i+2, \tilde{i}^d) - \Delta(i+1, \tilde{i}^d)],$$

$$\phi(\Delta(i+1, \tilde{i}^d)) - \phi(\Delta(i, \tilde{i}^d)) \leq -\alpha p_{u(i, \tilde{i}^d)} (\Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d)).$$

Substituting the above two inequalities into (50) we get

$$\begin{aligned}
& \Delta_{k+1}(i+1, \tilde{i}^d) - \Delta_{k+1}(i, \tilde{i}^d) \\
\leq & g(i+1, \tilde{i}^d) - 2g(i, \tilde{i}^d) + g(i-1, \tilde{i}^d) \\
& + \alpha \{ (1 - \gamma_1^d - \gamma_2^d - \mu(i+1) - \lambda(N-i+1) p_{u(i, \tilde{i}^d)}) \\
& [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)] \\
& + \mu(i-1) [\Delta_k(i, \tilde{i}^d) - \Delta_k(i-1, \tilde{i}^d)] \\
& + \gamma_1^d [\Delta_k(i+1, \tilde{i}^u + \Delta y) - \Delta_k(i, \tilde{i}^u + \Delta y)] \\
& + \gamma_2^d [\Delta_k(i+1, \tilde{i}^d - \Delta y) - \Delta_k(i, \tilde{i}^d - \Delta y)] \} \\
& + \lambda(N-i-1) \alpha p_{u(i, \tilde{i}^d)} \varepsilon_k^u \\
\leq & 2\kappa + \alpha[1 - 2(\lambda + \mu)] \varepsilon_k^u
\end{aligned}$$

The second inequality holds since the following four terms are smaller or equal than ε_k^u based on our assumption at iteration k : $\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)$, $\Delta_k(i, \tilde{i}^d) - \Delta_k(i-1, \tilde{i}^d)$, $\Delta_k(i+1, \tilde{i}^d + \Delta y) - \Delta_k(i, \tilde{i}^d + \Delta y)$, $\Delta_k(i+1, \tilde{i}^d - \Delta y) - \Delta_k(i, \tilde{i}^d - \Delta y)$. Denoting $\varepsilon_{k+1}^u = 2\kappa + \alpha[1 - 2(\lambda + \mu)] \varepsilon_k^u$, it can be seen that

$$\Delta_{k+1}(i+1, \tilde{i}^d) - \Delta_{k+1}(i, \tilde{i}^d) \leq \varepsilon_{k+1}^u \tag{57}$$

for all $\{i, \tilde{i}^d\}$ at iteration $k+1$. Similarly we can prove

$$\Delta_{k+1}(i+1, \tilde{i}^u) - \Delta_{k+1}(i, \tilde{i}^u) \leq \varepsilon_{k+1}^u \tag{58}$$

By mathematical induction, we conclude that

$$\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d) \leq \varepsilon_k^u, \Delta_k(i+1, \tilde{i}^u) - \Delta_k(i, \tilde{i}^u) \leq \varepsilon_k^u \tag{59}$$

holds for all k in the infinite series ε_k^u generated recursively by

$$\varepsilon_{k+1}^u = 2\kappa + \alpha[1 - 2(\lambda + \mu)] \varepsilon_k^u. \tag{60}$$

Since $\varepsilon_0^u = 0$, ε_k^u must converge as $k \rightarrow \infty$. Indeed it converges to ε^u with

$$\varepsilon^u = \lim_{k \rightarrow \infty} \varepsilon_k^u = \frac{2\kappa}{1 - \alpha[1 - 2(\lambda + \mu)]}. \tag{61}$$

Hence we have

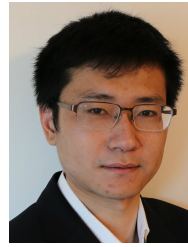
$$\begin{aligned}
& \Delta(i+1, \tilde{i}^d) - \Delta(i, \tilde{i}^d) \\
& = \lim_{k \rightarrow \infty} [\Delta_k(i+1, \tilde{i}^d) - \Delta_k(i, \tilde{i}^d)], \\
& \leq \lim_{k \rightarrow \infty} \varepsilon_k^u = \varepsilon^u.
\end{aligned}$$

And this concludes the proof of Proposition 3. \square

REFERENCES

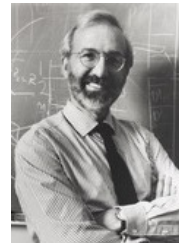
- [1] Yuri V. Makarov, Clyde Loutan, Jian Ma, and Phillip de Mello. Operational impacts of wind generation on california power systems. *IEEE Trans. Power Syst.*, 24(2):1039–1050, May 2009.
- [2] Jian Ma, Shuai Lu, P.V. Etingov, and Y.V. Makarov. Evaluating the impact of solar generation on balancing requirements in southern nevada system. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–7, 2012.
- [3] D.A. Halamay, T.K.A. Brekken, A. Simmons, and S. McArthur. Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation. *Sustainable Energy, IEEE Transactions on*, 2(3):321–328, 2011.
- [4] Donald J Hammerstrom, Jerry Brous, David P Chassin, Gale R Horst, Robert Kajfasz, Preston Michie, Terry V Oliver, Teresa A Carlon, Conrad Eustis, Olof M Jarvegren, et al. Pacific northwest gridwise testbed demonstration projects; part ii. grid friendly appliance project. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2007.
- [5] Demand Response and Advanced Metering, FERC staff report, October 2013: <http://www.pjm.com/~media/documents>.
- [6] Ioannis Ch. Paschalidis, Binbin Li, and Michael C. Caramanis. Demand-side management for regulation service provisioning through internal pricing. *IEEE Trans. Power Systems*, 27(3):1531–1539, August 2012.
- [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] 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.
- [9] 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.
- [10] 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.

- [11] 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.
- [12] Duncan S Callaway. Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. *Energy Conversion and Management*, 50(5):1389–1400, 2009.
- [13] S. Meyn, P. Barooah, A Busic, and J. Ehren. Ancillary service to the grid from deferrable loads: The case for intelligent pool pumps in florida. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 6946–6953, Dec 2013.
- [14] M.D. Ilic, Le Xie, and Jhi-Young Joo. Efficient coordination of wind power and price-responsive demand – part i: Theoretical foundations. *Power Systems, IEEE Transactions on*, 26(4):1875–1884, Nov 2011.
- [15] M. Alizadeh, Tsung-Hui Chang, and A Scaglione. Grid integration of distributed renewables through coordinated demand response. In *Decision and Control (CDC), 2012 IEEE 51st Conference on*, pages 3666–3671, Dec 2012.
- [16] Antonio J Conejo, Juan M Morales, and Luis Baringo. Real-time demand response model. *IEEE Transactions on Smart Grid*, 1(3):236–242, 2010.
- [17] L. P. Qian, Y. J. A. Zhang, J. Huang, and Y. Wu. Demand response management via real-time electricity price control in smart grids. *IEEE Journal on Selected Areas in Communications*, 31(7):1268–1280, July 2013.
- [18] Mardavij Roozbehani, Munther A Dahleh, and Sanjoy K Mitter. Volatility of power grids under real-time pricing. *IEEE Transactions on Power Systems*, 27(4):1926–1940, 2012.
- [19] DJ Hammerstrom, R Ambrosio, J Brous, TA Carlon, DP Chassin, JG DeStee, RT Guttromson, GR Horst, OM Järvegren, R Kajfasz, et al. Pacific northwest gridwise testbed demonstration projects; part i. olympic peninsula project. *Part I. Olympic Peninsula Project*, 210, 2007.
- [20] E. Bilgin, M.C. Caramanis, I.C. Paschalidis, and C.G. Cassandras. Provision of regulation service by smart buildings. *Smart Grid, IEEE Transactions on*, PP(99):1–11, 2015.
- [21] 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.
- [22] Frank P Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(1):33–37, 1997.
- [23] 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.
- [24] Ioannis Ch. Paschalidis and John N. Tsitsiklis. Congestion-dependent pricing of network services. *IEEE/ACM Trans. Networking*, 8(2):171–184, April 2000.
- [25] Huseyin Mutlu, Murat Alanyali, , and David Starobinski. Spot pricing of secondary spectrum access in wireless cellular networks. *IEEE/ACM Trans. Networking*, 17(6):1794–1804, December 2009.
- [26] Ratepayer revolt in Pacific Northwest: <http://news.google.com/news-papers?nid=1908&dat=19820303&id=j0IrAAAIBAJ&sjid=TdQEA AAAIBAJ&pg=3606,5327555>.
- [27] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2007.
- [28] PJM Balancing Operations: <http://pjm.com/markets-and-operations/ancillary-services>.
- [29] Erol Gelenbe. Energy packet networks: adaptive energy management for the cloud. In *Proceedings of the 2nd International Workshop on Cloud Computing Platforms*, page 1. ACM, 2012.
- [30] 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.
- [31] PJM market and operations: <http://pjm.com/markets-and-operations/ancillary-services>.
- [32] Bowen Zhang and John Baillieul. A two level feedback system design to provide regulation reserve. In *the 52nd IEEE Conference on Decision and Control*, pages 4322–4328, 2013.



Bowen Zhang (M'15) received the B.E. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2010 and the M.S. and Ph.D degree in Division of Systems Engineering from Boston University, Boston, MA, in 2013 and 2015, respectively. He is currently an Operations Research Analyst at Cambridge Energy Solutions in Cambridge, MA and a visiting scholar to the Department of Mechanical Engineering at Boston University in Boston, MA.

His research focuses on control and optimization with applications to electric power system, demand side management, and electricity markets.



Michael C. Caramanis (M'81-SM'13) received the B.S. degree in chemical engineering from Stanford University, Stanford, CA, in 1971 and the M.S. and Ph.D. degrees in engineering with minors in microeconomics and economic development from Harvard University, Cambridge, MA, in 1972 and 1976, respectively.

He is a Professor at Boston University, Boston, MA, with appointments in the Department of Mechanical Engineering and the Division of Systems Engineering. He served at the Greek National Energy council (1976-1979), the MIT Energy Laboratory (1979-1982), and was chair of the Greek Regulatory Authority for Energy (2005-2008). His research focuses on complex stochastic dynamic systems with applications in manufacturing systems and dynamic power markets.



John Baillieul (M'83-SM'89-F'93) performs research that deals with mathematical system theory in all of its manifestations and applications. His early work dealt with connections between optimal control theory and what came to be called sub-Riemannian geometry. Other early work treated the control of nonlinear systems modeled by homogeneous polynomial differential equations. Such systems describe, for example, the controlled dynamics of a rigid body. His main controllability theorem applied the concept of finiteness embodied in the Hilbert basis theorem

to develop a controllability condition that could be verified by checking the rank of an explicit finite dimensional operator. Over the past decade, he has turned his attention to the study of information based control. Together with K. Li, he has explored source coding of feedback signals that are designed to provide optimally robust performance in the face of time-varying feedback channel capacity. He has also collaborated with W. S. Wong on research aimed at understanding information complexity in terms of the energy required for two (or more) agents to communicate through the dynamics of a control system. Dr. Baillieul is a Fellow of IFAC and a Fellow of SIAM.