Optimal Provision of Distributed 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 regulation service reserves (RSR) by controlling dynamic demand preferences in smart buildings. A major contribution over past dynamic pricing work is that we pioneer the relaxation of static, uniformly distributed utility of demand. In this paper we model explicitly the dynamics of energy service preferences leading to a non-uniform and time varying probability distribution of demand utility. More explicitly, we model active and idle duty cycle appliances in a smart building as a closed queuing system with price-controlled arrival rates into the active appliance queue. Focusing on cooling appliances, we model the utility associated with the transition from idle to active as a non-uniform time varying function. We (i) derive an analytic characterization of the optimal policy and the differential cost function, and (ii) prove optimal policy monotonicity and value function convexity. These properties enable us to propose and implement a smart assisted value iteration (AVI) algorithm and an approximate DP (ADP) that exploits related functional approximations. Numerical results demonstrate the validity of the solution techniques and the computational advantage of the proposed ADP on realistic, large-state-space problems.

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 CO\textsubscript{2} 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 Reserves (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, which have been so far employed for this purpose, 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 of thermostatic appliances [6], [7], [8], decentralized scheduling of vehicle-to-grid integration [9], [10], [11], [12], and the optimal coordination of multiple units in a micro-grid flexible loads [13], [4], [16].

This paper provides a significant extension of past work on duty cycle appliance loads modeled as Markov modulated processes and controlled by dynamic price signals to track real time Independent System Operator (ISO) Regulation Service Signals [7], [17], [18]. Similar control policies have been investigated also in the context of Internet service and mobile telephony bandwidth management [19], [20], [21], [22]. Despite its significant contributions, this extensive literature that is merely sampled above, has relied on the common assumption that potential demand is associated with a practically infinite pool of relatively homogeneous users or appliances whose utility for bandwidth or energy is reasonably represented by a static probability distribution that is practically independent of recent price controls. We depart from this assumption, proposing and solving a stochastic DP that optimizes tradeoffs between tracking ISO RSR requests in the seconds time scale while respecting dynamic energy service preferences of smart building occupants. More broadly construed, we make a contribution to the real time price demand control literature by pioneering the relaxation of the usual assumption that demand for service - or power consumption - is adequately represented in the short term by a static probability distribution of preferences. We explicitly model the dynamic nature of preferences and their short term evolution in response to past control and state trajectories. In other words, we consider demand preferences to be dynamic, and, as such, comprise a component of the system state that evolves dynamically just like all of the other state components. Consider for further elaboration the demand for an alternative mobile service provider’s bandwidth, or the demand for turning on a cooling appliance. Works to this date have assumed routinely that this demand can be reasonably modeled by a static probability distribution of the typical participant’s utility function, or reservation price. They assume that the population of appliances is characterized by a temperature sensitive reservation price for the purchase of a block of energy whose probability distribution across appliances remains unchanged regardless of whether a high or a low price has been broadcasting repeatedly. In other words, they assume that regardless of the recent price control history that has resulted in more or fewer appliances being active, the temperature distribution will remain uniform. It is quite obvious that this assumption is unrealistic and hence restrictive and inaccurate. In fact, the higher the reserves offered by a finite number of appliances, the more restrictive
the assumption of uniform and price control independent temperature probability distribution. If we are interested in offering more than minuscule – and hence insignificant – levels of reserves from the finite number of duty cycle appliances in a smart building, the more important it is to stop relying on this assumption. Evidence on so called ratepayer revolt [23] speaks to the consequences of ignoring energy user preferences.

This paper models specifically the dynamics of building occupant preferences and in particular their stochastic dynamic evolution as a function of the current state and past control actions. Focusing for simplicity of exposition, but without loss of generality, on a multiple cooling appliance smart building load, we use a dynamic probability distribution to represent cooling zone occupant preferences to transition their cooling appliance from an idle to an active state. We account for the fact that a sustained broadcast of high prices that has discouraged turning on idle cooling appliances will increase the likelihood that the temperature of a typical cooling zone is high and thus raise the occupant’s preference to commence cooling by turning on the idle appliance. The opposite is naturally true after a sustained period of broadcasting low prices.

We model dynamically evolving cooling preferences without an explosion in the DP state size. 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 typical 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 algorithm [24] that estimates optimal value function approximation parameters and near optimal policies. To summarize, the contribution of this paper lies in the synergy of

(i) introducing a dynamic sufficient statistic representing the probability distribution of idle cooling appliance preferences,
(ii) proving value and policy function properties that assist accurate value function parametrization, and
(iii) using approximate DP to obtain near optimal policies for realistic size systems.

The rest of the paper proceeds as follows. In Sec[I] we formulate a DP problem by describing the state dynamics, the sufficient statistic of cooling zone preferences, period costs, and the Bellman Equation. Sec[II][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 approximate dynamic programming approach for estimating the value function approximation parameters. Numerical results are provided to illustrate the performance of the two DP solution algorithms. Sec[VI] concludes the paper and proposes future research 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. Moreover, 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 $p(t)$ to all cooling appliances in order to modulate the number of active appliances and hence control the resulting consumption. Appliances respond according to their current preferences for cooling service which depend on their corresponding cooling zone temperature. Deficient ISO RSR signal tracking penalties and occupant utility realizations constitute period costs. The objective is to find a state feedback optimal policy that minimizes the associated infinite horizon discounted cost. Individual cooling zone preferences are modelled by a dynamically evolving probability distribution of idle-appliance-zone temperatures. Assuming a known relationship connecting cooling zone preferences to temperatures, the probability distribution of preferences (or derived utility levels) for cooling appliance activation is derived. We next introduce notation, describe system dynamics, and formulate the period cost function of the relevant stochastic dynamic problem.

A. Notation

$N$: Total number of active or idle appliances.

$i(t)$: Number of active, i.e. connected, appliances at time $t$.

We assume $N_1 \leq i(t) \leq N_2$ for all $t$.

$y(t)$: RSR signal specified by the ISO at time $t$. The ISO specifies $y(t)$ as the output of a pre-specified proportional integral filter of observed Area Control Error (ACE) and System Frequency excursions. This results in a zero time average behavior of $y(t)$ and a well defined Markovian model of $y(t)$ dynamics [25].

$D(t)$: RSR signal $y(t)$’s direction. $D(t) = \text{sgn}(y(t) - y(t - \varepsilon))$.

$T(t)$ or simply $T$: Temperature in the cooling zone of an idle appliance at time $t$.

$p_i(T)$: Probability density function (pdf) of $T$ at time $t$, $T \in [T_{\text{min}}, T_{\text{max}}]$. $T_{\text{min}}$ and $T_{\text{max}}$ are the threshold values of room temperature that determine the appliance cooling zone occupant’s comfort zone. Assuming standard heat transfer relationships, extensive simulation reported below indicates that the controlled system results in a very accurate trapezoidal approximation of $p_i(T)$, shown in Fig. [I] with base $(T_{\text{min}}, 0)$ to $(T_{\text{max}}, 0)$ and top side $(T_{\text{min}}, h)$ to $(\bar{T}, h)$ where $h = 2/(T_{\text{max}} + \bar{T} - 2T_{\text{min}})$. Note that $\bar{T}$ 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.

\( \bar{T}(t) \) or more simply \( \hat{T} \): State parameter characterizing fully \( p_i(T) \) as shown in Fig. 1.

\( \bar{\pi}(i), \bar{\pi}(i) \): Component state to \( i(t) \) that is composed of \( y(i), D(t) \), and \( \bar{T}(i) \). Specifically, \( \bar{\pi}(i) = \{y(i), D(t) = 1, \bar{T}(i) \} \) and \( \bar{\pi}(i) = \{y(i), D(t) = -1, \bar{T}(i) \} \). It will be shown that \( \bar{\pi}(i) \) and \( \bar{\pi}(i) \) are independent of the control history.

\( U(T) \): Utility of idle cooling appliance zone occupant for resumption of cooling when the zone temperature is \( T \). \( U(T) \) is defined as a linear function of \( T \) that is monotonically increasing over \( T \in [T_{\text{min}}, T_{\text{max}}] \).

\( \pi(i) \): Price broadcast by the SBO to all idle cooling appliances at time \( t \).

\( \lambda \): Price detection rate at which an idle appliance considers resuming cooling by comparing its utility \( U(T) \) to \( \pi(i) \).

\( u(t) \): Threshold temperature value obtained by solving \( \pi(y) = U(u(t)) \). Idle appliances that consider to resume cooling at time \( t \) will do so if their utility is no less than \( \pi(i) \), namely if \( U(u(t)) \geq \pi(i) \). Since the mapping between \( \pi(i) \) and \( u(t) \) is linear, and hence bijective, for the rest of the paper we use \( u(t) \) to represent the control policy.

\( \bar{u}(i, \bar{\pi}(i)), \bar{u}(i, \bar{\pi}(i)) \): Optimal control policy for infinite horizon discounted cost dynamic programming problem that is regardless of time \( t \) and depends on state \( i, \bar{\pi}(i) \) or \( \bar{\pi}(i) \).

\( \mu \): Rate at which active, i.e., connected, appliances switch to idle, i.e., the rate at which they disconnect. Note that \( \mu \) is estimated so that \( 1/\mu \) equals the average time it takes a cooling appliance to complete the consumption of an energy packet. In this paper, our building grid operating model uses the concept of energy packet that has been introduced in [8] and [25], 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.

\( \bar{n} \): The constant energy consumption rate that the SBO purchased in the hour ahead market. We define the unit of capacity to represent the average consumption rate of an active appliance. Recalling that average up or down RSR reserves required equals 0, i.e., the average value of \( y(t) \) is guaranteed in advance to equal 0 [25], we conclude that \( \bar{n} \) 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.

\( R \): Maximum up or down RSRs that the SBO agreed to provide in the hour-ahead market. Recall that when \( R \) and \( \bar{n} \) are decided in the hour ahead market, the SBO assumes the responsibility to modulate its actual real time consumption during the hour to track \( \bar{n} + y(t)R \). Imperfect tracking results in tracking penalties. The SBO must select its subsequent real time control decisions to trade off tracking costs against competing cooling zone occupant utility losses.

\( e(t) = y(t) - \bar{n} - y(t)R \): The RSR signal tracking error at time \( t \). We assume here that the average consumption rate of an active cooling appliance is 1, i.e., it provides the unit of measuring capacity.

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

**B. State Dynamics**

Recall that state variables, contain \( y(i), y(t), D(t) \) and \( \bar{T}(i) \). Queues \( y(i) \) and \( N-i(i) \) constitute a closed queueing network where the service rate of one queue determines the arrival rate into the other. Queue \( N-i(i) \) 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 \( p_i(T) \). Queue \( y(i) \) behaves also as an infinite server queue with each server exhibiting a constant service rate \( \mu \). The dynamics of \( y(t) \) and the dependent state variable \( D(t) = \text{sgn}(y(t) - y(t-e)) \) are characterized by transitions taking place in short but constant time intervals, \( \tau(t) \) resulting in \( y(t) \) staying constant, increasing or decreasing by a typical amount of \( \Delta y = \gamma \tau(t) 150 \text{sec} \) [25]. 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 white noise process resulting from imbalance between stochastic demands and supply, \( y(t) \) is then an non-anticipating 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 Markov process that allows us to uniformize the DP problem formulation. To uniformize the DP problem we introduce a control update period of \( \Delta t << \tau \) which assures that during the period \( \Delta t \), 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 \( y(t) \) and \( D(t) \). Statistical analysis on historical PJM data on \( y(t) \) trajectories indicate a week dependence on \( y(t) \) yielding the reasonable approximation:

\[
\begin{align*}
\text{Prob}(y(t + \tau) = y(t) + \Delta y | D(t) = 1) &= 0.8 \\
\text{Prob}(y(t + \tau) = y(t) - \Delta y | D(t) = 1) &= 0.2 \\
\text{Prob}(y(t + \tau) = y(t) - \Delta y | D(t) = -1) &= 0.8 \\
\text{Prob}(y(t + \tau) = y(t) + \Delta y | D(t) = -1) &= 0.2
\end{align*}
\]

Denoting by \( \gamma(t) \) the rate at which \( y(t) \) will jump up by \( \Delta y \) during a control update period when \( D(t) = 1 \) (\( D(t) = -1 \))

\(^1\)This varies across ISOs. In PJM it is either 2 or 4 seconds depending on the type of regulation service offered.
\[
\begin{align*}
\gamma_1^d & = 0.8\Delta_r / \tau_r, \\
\gamma_2^d & = 0.2\Delta_r / \tau_r.
\end{align*}
\]

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

\[
\begin{align*}
\text{Prob}(y(t+1) = y(t) + \Delta_r|D(t) = 1) &= \gamma_1^d \\
\text{Prob}(y(t+1) = y(t) - \Delta_r|D(t) = 1) &= \gamma_2^d \\
\text{Prob}(y(t+1) = y(t) + \Delta_r|D(t) = -1) &= \gamma_1^d \\
\text{Prob}(y(t+1) = y(t) - \Delta_r|D(t) = -1) &= \gamma_2^d.
\end{align*}
\]

2) **Dynamics of \( i(t) \):** The dynamics of \( i(t) \) is governed by the arrival rate 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 \( \lambda \), 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

\[
\begin{align*}
\lambda p_u(t) &= (N - i(t)) \lambda \\
&= \int_{u(t)}^{T_{\text{max}}} p(T) dT,
\end{align*}
\]

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 inverse of the average duration of a cooling cycle. Modelling the cooling cycle duration as an exponential random variable with rate \( \mu \) such that \( 1/\mu \) equals the average cooling cycle duration we have,

\[
d(t) = i(t) \mu.
\]

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

\[
\begin{align*}
p(\tilde{t} = 1) &= a(t) \\
p(\tilde{t} = -1) &= d(t) \\
p(\tilde{t} = 0) &= 1 - a(t) - d(t) + \gamma,
\end{align*}
\]

where \( \gamma = \gamma_1^d + \gamma_2^d = \gamma_1^d + \gamma_2^d \) is the total probability that the ISO RSR signal will change.

Fig. 2 represents the stochastic dynamics of the number of active and idle appliances as a two queue closed queuing network with queue levels summing to \( N \).

3) **Dynamics of \( D(t) \):** Recalling that \( D(t) = \text{sgn}(y(t) - y(t-1)) \), it is clear that the dynamics of \( D(t) \) are fully determined by the dynamics of \( y(t) \). We next argue 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 modeling [9] and [27], 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 demonstrate that they have the ability to track. This is referred to in the PJM manual as the standard T-50 qualifying test [25]. The second is a real time signal downloaded from PJM [28]. We record the temperature levels prevailing across the \( N \) cooling zones when a control trajectory is applied that results in near-perfect tracking of the ISO RSR requests implied by the aforementioned two signals. Simulation results indicate that the time evolution of the probability distribution of cooling zone temperatures conforms to a dynamically changing trapezoid characterized fully by \( \hat{T}(t) \); see Fig. 1.

Fig. 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 blue 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 blue curves taken at each temperature state. The red curve is then approximated by the trapezoid green curve proposed in Fig. 1. \( \hat{T}(t) \) 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.

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
parametrized by a single parameter \( \hat{T} \). The height of the trapezoid is \( h = 2/(T_{\text{max}} + T_{\text{min}}) \).

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

two vector 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 \( \hat{T} \) on \( y(t) \) with the following linear function

\[
\hat{T}(t) = \alpha_0 + \alpha_1 y(t) + \omega,
\]

where \( \alpha_0 \) corresponds to the value of \( \hat{T}(t) \) when the building’s energy consumption level is \( \bar{n} \), \( \alpha_0 < 0 \), and \( \omega \) is a zero mean symmetrically distributed error. These results do 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 \( T_{\text{min}} \) 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 \( \omega \) 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 \( \hat{T} \) (blue 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, which is consistent with our assumption in proposing (3).

C. Period Cost

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

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

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

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

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

\[
U_u = \frac{\int_{u(t)}^{T_{\text{max}}} U(T)p(T)dT}{\int_{u(t)}^{T_{\text{max}}} p(T)dT}. \tag{6}
\]
Noting that the probability that an idle appliance will decide to resume cooling is $a(t)$, the expected realized utility rate is

$$a(t) U_a = (N - i(t)) \bar{\lambda} p_u U_u,$$

$$= (N - i(t)) \bar{\lambda} \int_{u(t)}^{T_{\max}} P(T) p(T) dT,$$

$$= (N - i(t)) \bar{\lambda} \int_{u(t)}^{T_{\max}} U(T) p(T) dT. \tag{7}$$

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

$$c(i(t), y(t), u(t)) = \kappa i(t) - \bar{n} - y(t) R^2 - \int_{u(t)}^{T_{\max}} U(T) p(T) dT. \tag{8}$$

**D. Bellman Equation**

The state variables can be grouped according to their dependence on $u(t)$: $i(t)$ depends explicitly on $u(t)$. $\hat{\beta}(t)$ 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 therefore consider all state variables, other than $i(t)$, to have dynamics that do not depend on $u(t)$.

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

$$J(i, \hat{\beta}) = \min_{u \in [\hat{T}_{\min}, \hat{T}_{\max}]} \{g(i, \hat{\beta}) - a(t) U_u + \alpha_a a(t) J(i+1, \hat{\beta} + d) + d(T - T_{\max}) + \gamma_f J(i, \hat{\beta} + \Delta y) + \gamma_f J(i, \hat{\beta} - \Delta y) + (1 - a(t) - d(t)) J(i, \hat{\beta})\}. \tag{9}$$

$J(i, \hat{\beta})$ is the value function satisfying the Bellman equation, with $\alpha$ denoting the discount factor. For notational simplicity we denote by $\hat{\beta} + \Delta y$ 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 remaining state variables remain unchanged. Similarly we denote by $\hat{\beta} - \Delta y$ 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 remaining state variables remain unchanged. The superscripts $u$ ($d$) stand for upwards (downwards) RSR signals. $J(i, \hat{\beta})$ 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 the utility enjoyed by activating an idle appliance and allowing it to begin a cooling cycle

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

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

$$a(t) U_u = (N - i(t)) \bar{\lambda} \int_{u(t)}^{T_{\max}} U(T) p(T) dT,$$

$$= (N - i(t)) \bar{\lambda} \int_{u(t)}^{T_{\max}} b(T - T_{\min}) \frac{1}{T_{\max} - T_{\min}} dT, \tag{11}$$

$$= (N - i(t)) \bar{\lambda} \max_{u} \frac{b(T_{\max} - u)(T_{\max} + u - 2T_{\min})}{2T_{\max} - T_{\min}}$$

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 a random parameter characterizing individual consumer’s utility function choice, and $\varepsilon$ is a random noise built on top of the function – as long as two properties hold for the utility function: (i) we need $\varepsilon$ 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 of $U_u$ in (5), we have

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

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

$$\frac{d}{du} a(t) U_u = \frac{d}{du} \max_{u(t)} \int_{u(t)}^{T_{\max}} U(T) p(T) dT,$$

$$= - \max_{u(t)} \int_{u(t)}^{T_{\max}} U(T) p(T) dT \frac{b(T_{\max} - u)(T_{\max} + u - 2T_{\min})}{2T_{\max} - T_{\min}}, \tag{13}$$

Since $E_{\varepsilon}[U(u, p, \varepsilon)]$ increases with $u$, we have $\frac{d}{du} a(t) U_u < 0$. 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

$$p(T) = \begin{cases} \frac{T_{\max} + \hat{T}}{2T_{\max}}, & T \leq \hat{T}, \\ \frac{2(T_{\max} - \hat{T})}{2T_{\max}}, & T \geq \hat{T}. \end{cases}$$
For example, if we use a linear utility function as in (10), it yields the following expected period utility rate

\[ a(t)U_u = \begin{cases} \left[N - i(t)\right]\frac{2h(C_1 - \frac{1}{2}u^2 + T_{\min}u)}{T_{\max} + T - 2T_{\min}}, & u \leq T, \\
\left[N - i(t)\right]\frac{2h(C_2 - \frac{1}{2}u^2 + T_{\min} + T_{\max})}{(T - T_{\max})(T_{\max} + T - 2T_{\min})}, & u \geq T, \end{cases} \]

(14)

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

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

(15)

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

Under the static uniform probability distribution \( p(T) \), the optimal policy can be easily obtained since we can set the 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 the dynamic \( p(T) \) assumption. 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 \( J(i, \bar{p}) \) w.r.t. the active appliance state variable \( i(t) \) as

\[ \Delta(i + 1, \bar{p}) = J(i + 1, \bar{p}) - J(i, \bar{p}). \]

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

\[ u^*(i, \bar{p}) = \arg\min_u g(i, \bar{p}) - \lambda(N - i) p_u + \alpha \left\{ i u J(i - 1, \bar{p}) + \frac{1}{2} \left( \sigma^2 J(i, \bar{p}) + \lambda(N - i) p_u \right) \right\} + \bar{p} \left\{ i u J(i + 1, \bar{p}) + \frac{1}{2} \left( \sigma^2 J(i, \bar{p}) - \lambda(N - i) p_u \right) \right\} + \left\{ i - (i + 1) \frac{1}{2} \right\} \left\{ J(i, \bar{p}) - \lambda(N - i) p_u \right\}, \]

(16)

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

\[ f(u, \Delta(i + 1, \bar{p})) = p_u - \alpha p_d \Delta(i + 1, \bar{p}), \]

(17)

we can write that the optimal policy must satisfy

\[ \max_{u \in \left[\max(\hat{T}, T_{\min} + \frac{T_{\max}}{2}), T_{\max}\right]} f(u, \Delta(i + 1, \bar{p})). \]

(18)

Proposition 1 Under the conventional assumption that consumers’ utility preference is statically uniformly distributed \( (T = T_{\max}) \), \( f(u, \Delta(i + 1, \bar{p})) \) 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 \( u^*(i, \bar{p}) \) satisfying \( \frac{d}{du} f(u, \Delta(i + 1, \bar{p})) \bigg|_{u = u^*(i, \bar{p})} = 0 \).

Proposition 1 is straightforward because the first term in \( f(u, \Delta(i, \bar{p})) \) is quadratic and the second term is a linear function of \( u \) for \( T = T_{\max} \). When \( \hat{T} < T_{\max} \) with \( p(T) \) no longer uniform but trapezoid, \( f(u, \Delta(i + 1, \bar{p})) \) stops possessing

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 \( E[U(T, b, \varepsilon)] \), the optimal policy that solves (18) is described by the following relations

\[ u^*(i, \bar{p}) = \begin{cases} T_{\max}, & \text{if } \alpha \Delta(i + 1, \bar{p}) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)] \\
T_{\min}, & \text{if } \alpha \Delta(i + 1, \bar{p}) \leq 0 \\
\Gamma^{-1}(\alpha \Delta(i + 1, \bar{p})), & \text{otherwise}, \end{cases} \]

(19)

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

Proof. Let \( f(u, \Delta(i + 1, \bar{p})) = p_u - \alpha \Delta(i + 1, \bar{p}) \), for the three conditions in (19) we claim the following:

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

2. When \( \alpha \Delta(i + 1, \bar{p}) \leq 0 \), both \( p_u, \alpha \Delta(i + 1, \bar{p}) \) and \( p_u \) are monotonically decreasing functions of \( u \). Therefore \( f(u, \Delta(i + 1, \bar{p})) \) reaches its maximum at \( u^*(i, \bar{p}) = T_{\min} \).

3. When \( \alpha \Delta(i + 1, \bar{p}) \in (0, E_\varepsilon[U(T_{\max}, b, \varepsilon)]) \), we take derivative of \( f(u, \Delta(i + 1, \bar{p})) \) as \( f(u, \Delta(i + 1, \bar{p})) \) is continuously differentiable on \( (0, T_{\max}) \).

\[ \frac{d}{du} \left[ \Gamma^{-1}(\alpha \Delta(i + 1, \bar{p})) \right] = \frac{-p_u[U_u - \alpha \Delta(i + 1, \bar{p})]}{\alpha \Delta(i + 1, \bar{p})}, \]

(20)

Denote by \( u^*(i, \bar{p}) \) the optimal control that minimizes \( f(u, \Delta(i + 1, \bar{p})) \). A necessary condition is to have \( u^*(i, \bar{p}) \) be a local maximum of \( f(u, \Delta(i + 1, \bar{p})) \). Therefore it satisfies the first order condition

\[ p(u)[E_\varepsilon[U(u, T, \varepsilon)] - \alpha \Delta(i + 1, \bar{p})]|_{u = u^*(i, \bar{p})} = 0. \]

(21)

According to the proof in 1), \( p(u) = 0 \) if \( u^*(i, \bar{p}) = T_{\max} \) if and only if \( \alpha \Delta(i + 1, \bar{p}) \geq E_\varepsilon[U(T_{\max}, b, \varepsilon)] \). Therefore in this case \( p(u) \neq 0 \). The only solution to satisfy (21) is

\[ E_\varepsilon[U(u^*(i, \bar{p}), b, \varepsilon)] - \alpha \Delta(i + 1, \bar{p}) = 0. \]

(22)

Based on the definition of the inverse function, we have

\[ u^*(i, \bar{p}) = \Gamma^{-1}(\alpha \Delta(i + 1, \bar{p})). \]

(23)
For second order condition, it can be verified that \( \frac{d^2}{du^2} f(u, \Delta(i + 1, \vec{p}^i)) \big|_{u=u^*(i, \vec{p}^i)} < 0 \). Therefore \( u^*(i, \vec{p}^i) \) is a local maximum. Moreover, given \( f(u, \Delta(i + 1, \vec{p}^i)) \) is first order differentiable, \( \frac{d}{du} f(u, \Delta(i + 1, \vec{p}^i)) \) is continuous and has only one critical point inside the allowable control set. Thus the local maximum is the global maximum for \( u \in [T_{\min}, T_{\max}] \), namely \( u^*(i, \vec{p}^i) = \Delta^{-1}(\alpha \Delta(i + 1, \vec{p}^i)) \).

**Remark 1** The optimal policy characterization between \( u^*(i, \vec{p}^i) \) and \( \Delta(i + 1, \vec{p}^i) \) does not rely on \( p(T) \), namely it holds for broader possible realizations of consumers’ real time preferences. This is because \( u^* \) 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), as long as the solution of \( u^* \) is unique. Furthermore, it holds for a non-homogeneous utility function incorporating individual consumers’ utility preferences \( b \), and uncertainty \( \varepsilon \), as long as the parameters are independent of the current temperature.

**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 \( \Delta(i + 1, \vec{p}^i) \). In the extreme case when \( b \to \infty \), we have \( u = T_{\min} \) namely the lowest price is broadcast to guarantee the largest possible utility reward; when \( b \to 0 \), the optimal controller is bang-bang depending on the sign of \( \Delta(i + 1, \vec{p}^i) \) indicating that consumers become extremely elastic.

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

**IV. Properties of the Optimal Policy**

Proposition 2 expresses the optimal policy \( u^*(i, \vec{p}^i) \) as a function of \( \Delta(i + 1, \vec{p}^i) \). To study the properties of \( u^*(i, \vec{p}^i) \), we focus on the structure of \( \Delta(i + 1, \vec{p}^i) \). In this section we derive key properties of \( \Delta(i + 1, \vec{p}^i) \) in terms of the changes in state space variables that lead to desirable structures for \( u^*(i, \vec{p}^i) \). There are three state variables that affect \( u^*(i, \vec{p}^i) \), the aggregate consumption over all active appliances \( \iota(t) \), 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 \( \Delta(i + 1, \vec{p}^i) \) 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:

**Lemma 1** If we denote

\[
\phi(\Delta(i + 1, \vec{p}^i)) = \max_{u \in [T_{\min}, T_{\max}]} p_u u - \alpha p_u \Delta(i + 1, \vec{p}^i),
\]

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

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

\[
\frac{d \phi(\Delta(i + 1, \vec{p}^i))}{d \Delta(i + 1, \vec{p}^i)} = \frac{dp_{u^*(i, \vec{p}^i)}}{du} |_{u = u^*(i, \vec{p}^i)} \frac{\Delta^{-1}(\alpha \Delta(i + 1, \vec{p}^i)) - \Delta^{-1}(\alpha \Delta(i + 1, \vec{p}^i))}{d \Delta(i + 1, \vec{p}^i)}
\]

\[
= \left[ \frac{E[U(u^*(i, \vec{p}^i), b, \varepsilon)] - \alpha \Delta(i + 1, \vec{p}^i)}{p_u u^*(i, \vec{p}^i)} \right] \frac{d \Delta^{-1}(\alpha \Delta(i + 1, \vec{p}^i))}{d \Delta(i + 1, \vec{p}^i)}
\]

\[
= -\alpha p_{u^*(i, \vec{p}^i)} \leq 0.
\]

Therefore \( \phi(\Delta(i + 1, \vec{p}^i)) \) is a monotonically non-increasing function of \( \Delta(i + 1, \vec{p}^i) \).

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

**Lemma 2** \( \phi(\Delta(i + 1, \vec{p}^i)) - \phi(\Delta(i, \vec{p}^i)) \) has the following upper and lower bound:

(1) \( \phi(\Delta(i + 1, \vec{p}^i)) - \phi(\Delta(i, \vec{p}^i)) \leq -\alpha p_{u^*(i, \vec{p}^i)} (\Delta(i + 1, \vec{p}^i) - \Delta(i, \vec{p}^i)) \).

(2) \( \phi(\Delta(i + 1, \vec{p}^i)) - \phi(\Delta(i, \vec{p}^i)) \geq -\alpha p_{u^*(i - 1, \vec{p}^i)} (\Delta(i + 1, \vec{p}^i) - \Delta(i, \vec{p}^i)) \).

**Proof.** (1) The proof is a straightforward sequence of steps

\[
\phi(\Delta(i + 1, \vec{p}^i)) - \phi(\Delta(i, \vec{p}^i)) = \left[ p_u u - \alpha p_u \Delta(i + 1, \vec{p}^i) \right] |_{u = u^*(i, \vec{p}^i)}
\]

\[
- \left[ p_u u - \alpha p_u \Delta(i, \vec{p}^i) \right] |_{u = u^*(i - 1, \vec{p}^i)}
\]

\[
\leq \left[ p_u u - \alpha p_u \Delta(i + 1, \vec{p}^i) \right] |_{u = u^*(i, \vec{p}^i)}
\]

\[
- \left[ p_u u - \alpha p_u \Delta(i, \vec{p}^i) \right] |_{u = u^*(i, \vec{p}^i)}
\]

\[
= -\alpha p_{u^*(i, \vec{p}^i)} (\Delta(i + 1, \vec{p}^i) - \Delta(i, \vec{p}^i)).
\]

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

(2) Similarly

\[
\phi(\Delta(i + 1, \tilde{p}^d)) - \phi(\Delta(i, \tilde{p}^d)) = [p_i u_i - \alpha p_i \Delta(i + 1, \tilde{p}^d)]_{u = u^i(i, \tilde{p}^d)} \geq [p_i u_i - \alpha p_i \Delta(i + 1, \tilde{p}^d)]_{u = u^i(i - 1, \tilde{p}^d)} \geq -\alpha p_i \Delta(i + 1, \tilde{p}^d) - \Delta(i, \tilde{p}^d).
\]

(26)

This inequality holds also by a similar argument.

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 \( \Delta(i, \tilde{p}^d) \) with respect to state space parameter changes in RSR signal value \( y \), current aggregated demand \( i \), and tracking error \( e \). Properties of \( \Delta(i, \tilde{p}^d) \) will be used to prove the main Theorem on the structure of the optimal policy at the end of the section.

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

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

**Proposition 3** The following properties hold for a fixed \( \tilde{p}^d \):

1. \( \Delta(i + 1, \tilde{p}^d) \geq \Delta(i, \tilde{p}^d) + e^d \), where \( e^d = \frac{2\varepsilon}{1 - \alpha(1 - 2\varepsilon + 2\mu)u^d} \) with \( u = \lambda(N - N_1) \).
2. \( \Delta(i, \tilde{p}^d) + e^u \geq \Delta(i + 1, \tilde{p}^d) \), where \( e^u = \frac{2\varepsilon}{1 - \alpha(1 - 2\varepsilon + 2\mu)u} \).

**Proof.** See Appendix A.

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

\[
J(i + 1, \tilde{p}^d) + J(i - 1, \tilde{p}^d) > 2J(i, \tilde{p}^d).
\]

(27)

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 \( \Delta(i, \tilde{p}^d) \) for a fixed tracking error \( e \) with \( \tilde{p}^d \) 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 \( \{i, \tilde{p}^d\} \) and \( \{i + 1, \tilde{p}^d + \Delta y\} \) that have the same \( e \), and therefore it is reasonable to have \( \Delta(i, \tilde{p}^d) = \Delta(i + 1, \tilde{p}^d + \Delta y) \). 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 numbers of consuming appliances, \( i \), implies different numbers of idle appliances, \( N - i \)). We formally investigate the properties of \( \Delta(i, \tilde{p}^d) \) for a given \( e \) and state the monotonicity properties as follows.

**Proposition 4** The following properties hold for a fixed \( e \):

1. \( \Delta(i, \tilde{p}^d) \geq \Delta(i + 1, \tilde{p}^d + \Delta y) \).
2. \( \Delta(i + 1, \tilde{p}^d + \Delta y) + \varepsilon^u \geq \Delta(i, \tilde{p}^d) \) where \( \varepsilon^u = \frac{1}{1 - \alpha(1 - 2\varepsilon + 2\mu)} \varepsilon^u \).

**Proof.** See Appendix B.

In the end, we derive a last property wherein the aggregated consumption \( i \) is fixed while the ISO RSR signal \( y \) and tracking error \( e = i - i - y \) change in the same direction.

**Proposition 5** The following properties hold for a fixed \( i \):

1. \( \Delta(i, \tilde{p}^d) \geq \Delta(i, \tilde{p}^d + \Delta y) + e^d \).
2. \( \Delta(i, \tilde{p}^d + \Delta y) + e^u \geq \Delta(i, \tilde{p}^d) \).

**Proof.** We have

\[
\Delta(i, \tilde{p}^d) \geq \Delta(i + 1, \tilde{p}^d + \Delta y) \geq \Delta(i, \tilde{p}^d + \Delta y) + e^d,
\]

(28)

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

The above proposition completes our discussion of the properties of the differential cost function \( \Delta(i, \tilde{p}^d) \). These properties and the relation between \( \Delta(i + 1, \tilde{p}^d) \) and \( u^*(i, \tilde{p}^d) \) result in the following useful properties of the optimal policy in the next section.

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

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

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

1. For the same RSR signal, the optimal price policy is a monotonically non-decreasing function of \( i \). Namely \( u^*(i + 1, \tilde{p}^d) \geq u^*(i, \tilde{p}^d) \).
2. For the same tracking error, the optimal price policy is a monotonically non-increasing function of \( i \). Namely \( u^*(i, \tilde{p}^d) \geq u^*(i + 1, \tilde{p}^d + \Delta y) \).
3. For the same consumption level, the optimal price policy is a monotonically non-increasing function of \( \tilde{p}^d \). Namely \( u^*(i, \tilde{p}^d) \geq u^*(i, \tilde{p}^d + \Delta y) \).

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

\[
\Delta(i + 1, \tilde{p}^d) \geq \Delta(i + 2, \tilde{p}^d + \Delta y) \geq \Delta(i + 1, \tilde{p}^d + \Delta y).
\]

From Proposition 2 the optimal control \( u^*(i, \tilde{p}^d) \) is a non-decreasing function of the \( \Delta(i + 1, \tilde{p}^d) \), therefore

\[
u^*(i, \tilde{p}^d) \geq u^*(i, \tilde{p}^d + \Delta y) \geq u^*(i, \tilde{p}^d + \Delta y),
\]

and the three statements above are true.

**Remark 5** The policy monotonicity properties are valid for both \( D(i) = 1 \) and \( D(i) = -1 \). A state partition of the optimal policy can be drawn based on Theorem 1; see Fig. [V]. 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 \( \tilde{p}^d \), the ISO 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
increases following the relationships: 

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

(29)

and

$$
\alpha = \frac{1}{1 + r \Delta},
$$

(30)

where \( r \) is the prevailing discount rate. Substituting (29) into (30) we can write

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

(31)

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 \( \Delta(i + 1, \tilde{y}^d) - \Delta(i, \tilde{y}^d) \) and \( \Delta(i, \tilde{y}^d) - \Delta(i, \tilde{y}^d + \Delta y) \) approaches zero as \( N \) approaches infinity.

**Proposition 6** \( \epsilon^u \) and \( \epsilon^l \), defined in Proposition 3, and \( \epsilon^u \), defined in Proposition 4, will decrease as \( N \) increases, and moreover for \( N \to \infty \), their asymptotic limit is 0.

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

$$
\epsilon^u = \frac{2 \kappa \Delta}{1 - \alpha [1 - 2(\lambda + \mu) \Delta]}.
$$

(32)

Substituting into (32) the effective discount factor \( \alpha \) and the relation \( \Delta \approx 1/(N \max(\lambda, \mu)) \) we obtain

$$
\epsilon^u(N) = \frac{2K/(qN)^2}{1 - N \max(\lambda, \mu)[1 - 2(\lambda + \mu) \frac{1}{N \max(\lambda, \mu)}]}
$$

(33)

which in turn simplifies to \( \epsilon^u(N) = 2K/(qN)^2 \) verifying that \( \epsilon^u \) decreases as \( N \) increases. It can be similarly shown that \( \epsilon^l \) also decreases as \( N \) increases. Finally \( \epsilon^u \), as defined in Proposition 4, is shown below to equal a positive multiple of \( \epsilon^u \)

$$
\tilde{\epsilon}^u = \frac{\alpha (\lambda + \mu) \Delta}{1 - \alpha [1 - 2(\lambda + \mu) \frac{1}{N \max(\lambda, \mu)}]} \epsilon^u,
$$

(34)

$$
= \frac{1}{r + \lambda + \mu} \epsilon^u.
$$

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

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

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

(35)

and

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

(36)

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 \( u^u(i, \tilde{y}^d) \), \( u^*(i, \tilde{y}^d) \) and \( u^u(i, \tilde{y}^d + \Delta y) \) 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 resources 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 [24], 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 [19] 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. 8. 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 \( u = T_{\text{min}} \) to \( u = T_{\text{max}} \) 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.

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 \( \Delta(i+1,\bar{\mathcal{P}}) > \Delta(i,\bar{\mathcal{P}}) \). Given the discrete state space of our problem, this property is equivalent to convexity of \( J(i,\bar{\mathcal{P}}) \) 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. 8 that the increase rate of the optimal policy for a fixed RSR signal is approximately a constant value \( k \). Therefore we approximately have

\[
u^*(i+1,\bar{\mathcal{P}}) - u^*(i,\bar{\mathcal{P}}) = k, \tag{37}\]

which is equivalent to have \( \Delta(i+1,\bar{\mathcal{P}}) - \Delta(i,\bar{\mathcal{P}}) \) being some constant. Since \( \Delta(i,\bar{\mathcal{P}}) \) 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. 8 we note that the rate of policy’s vertical changes is constant for varying RSR signals, therefore the \( \frac{\partial}{\partial y^d} \Delta(i,\bar{\mathcal{P}}) \) is approximately constant, namely \( \frac{\partial}{\partial y^d} J(i,\bar{\mathcal{P}}) \) is constant.

These properties motivate an approximation of \( J(i,\bar{\mathcal{P}}) \) by \( \hat{J}_d(i,y) \) that is quadratic in \( i - yR \). In fact, we treat \( D(i) \) as a binary argument and propose the following basis function approximation when \( D(i) = -1 \).

\[
\hat{J}_d(i,y) = r_{11}(i - yR)^2 + r_{12}(i - yR) + r_{13} \tag{38}
\]

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

\[
\frac{\partial}{\partial i} \hat{J}_d(i,y) = \Delta(i,\bar{\mathcal{I}}) = 2r_{11}(i - yR) + r_{12}i. \tag{39}
\]

Proposition 4 proved that \( \Delta(i,\bar{\mathcal{P}}) \geq \Delta(i+1,\bar{\mathcal{P}}) + \Delta y \) which suggests that \( \frac{\partial}{\partial y^d} \hat{J}_d(i,y) \) monotonically decreases as a function of \( i \) for a fixed \( e = i - yR \), hence \( r_{12} < 0 \) in (39).

We generalize the approximation in (38) by differentiating the parameters depending on the discrete value of the direction \( D \). Thus we define for \( D = -1 \)

\[
\hat{J}_d(i,y,\mathbf{r}_d) = r_{11}i^2 + r_{12}i + r_{3}y^2 + r_{4}y + r_{5}iy + r_{6} \tag{40}
\]

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

\[
\hat{J}(i,y,D,r) = \mathbb{1}_{\{D=1\}} \hat{J}_d(i,y,\mathbf{r}_d) + \mathbb{1}_{\{D=-1\}} \hat{J}_d(i,y,\mathbf{r}_u). \tag{41}
\]

The two components of the expression in (41) approximate state features associated with increasing or decreasing ISO signals. The vector \( r \) is a vector of twelve parameters six from \( \mathbf{r}_d \) and \( \mathbf{r}_u \) each. Written in matrix form, (41) is equivalent to the following

\[
\hat{J} = \Phi r. \tag{42}
\]
where $\Phi$ 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

$$\Phi r = \Pi T (\Phi r),$$

(43)

where $T$ is the operator of the form $TJ = g + \alpha PJ$. $P$ is the state transition matrix that can be gotten based on the state dynamics matrix derived in Sec IV-B. $\Pi$ is the projection operator onto the set spanned by the basis functions $S = \{ \Phi r | r \in \mathbb{R}^{12} \}$. It is shown in [24] that the solution to the above projection problem is given by

$$r^* = C^{-1} d,$$

(44)

where $C = \Phi' \Xi (I - \alpha P) \Phi$, $d = \Phi' \Xi g$, and $\Xi$ is the matrix with diagonal elements being the steady state probability distribution of the states. It is further shown that (44) can be solved in an iterative form by the projected value iteration (PVI) algorithm

$$r_{k+1} = r_k - \gamma G_k (C_k r_k - d_k),$$

(45)

where $C_k$ and $d_k$ are given by

$$C_k = \frac{1}{k+1} \sum_{i=0}^{k} \phi(i, y_i, D_i) (\phi(i, y_i, D_i) - \alpha \phi(i+1, y_{i+1}, D_{i+1})), $$

$$d_k = \frac{1}{k+1} \sum_{i=0}^{k} \phi(i, y_i, D_i) g(i, y_i, D_i, u(t)).$$

(46)

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

$$G_k = \left( \frac{1}{k+1} \sum_{i=0}^{k} \phi(i, y_i, D_i) \phi(i, y_i, D_i) \right)^{-1}.$$ 

(47)

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** $r = 0$. 

**Step 2. Initialization** $r_{old} = r$, $r_0 = r$, $k = 0$, $\{i_k, y_k, D_k\}$.

**Step 3. Generate**

- Optimal policy $u_{r_{old}}(i_k, y_k, D_k)$
- Next state $\{i_{k+1}, y_{k+1}, D_{k+1}\}$
- Period cost $g(i_k, y_k, D_k, u_{r_{old}}(i_k, y_k, D_k))$

**Update**

- $C_k$, $d_k$, and $r_{old}$ based on (45)–(47)

- If $k \geq k_{min}$ and $\|r_{k+1} - r_k\| < \epsilon$
  - $r = r_{k+1}$, go to **Step 4**.
- Else
  - $k = k + 1$, go to **Step 3**.

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

- return $r^* = r$. Algorithm ends.
- Else go to **Step 2**.

The algorithm starts with an initial guess of the parameters, $r = 0$. Step 2 initializes the iteration count, parameters $r_0$, and state variables $\{i_0, y_0, D_0\}$. Step 3 iteratively updates the value function for the fixed policy $u_{r_{old}}$ 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 $r$ meets a desired tolerance, $\|r_{k+1} - r_k\| < \epsilon$. Steps 4 compares the value function parametrized by $r_{old}$ and the updated $r$ obtained by step 3. If the infinity norm of the vector is less than the threshold $\tau$, then the value function converges and the algorithm returns the optimal parameter $r^* = r$. 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 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 by more than 90%.

**TABLE I**

| Problem Size $([N] \times [y] \times [D])$ | CVI Computation Time (sec) | AVI Computation Time (sec) | ADP Computation Time (sec) |
|-------------------------------------------|-----------------------------|-----------------------------|-----------------------------|
| $500^21^2$                                | 320.35                      | 761.62                      | 596.65                      |
| $500^41^2$                                | 66082                       | 9849.5                      | 514.8                       |

We examine the convergence result of the proposed algorithm. The coefficient vector $r$ is composed of 12 parameters, six for each direction $D = 1, -1$. Fig. 10 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 $r$ also indicates the convergence of the value function $J$.

Fig. 10 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 $J(i, y, D, r)$ 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 exhibiting a negligible discrepancy error. The solution accuracy can also be seen from Fig. 11 in which we draw the value function and the policy comparisons for five different \( y' \)s, which form curves along the \( y \) axis in Fig. 10.

**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 (i) analytic expressions characterizing the optimal policy, and (ii) a range of monotonicity properties that capture first and second derivative-type properties of the value function, and describe further 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 parameters – or equivalently the Smart Building size – increases. The aforementioned analytical results prove invaluable in guiding us to design and implement efficient and scalable numerical solution algorithms. In particular, they guide us to propose an analytical value function approximation DP algorithm that surpasses in performance and scalability conventional value-iteration-based DP solutions. In future work we will consider imperfect state observation dynamic programming formulations where the utility function of consumers or the dynamic time varying preference distribution cannot be observed by the SBO.

**APPENDIX A**

**PROOF OF PROPOSITION 3**

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

\[
\Delta_k (i, \bar{p}^d) = J_k (i, \bar{p}^d) - J_k (i - 1, \bar{p}^d).
\]

It follows that \( \lim_{k \to \infty} \Delta_k (i, \bar{p}^d) = \Delta (i, \bar{p}^d) \).
We assume $\Delta_k(i + 1, \vec{p}^d) \geq \Delta_k(i, \vec{p}^d) + \epsilon_k^t$ and $\Delta_k(i + 1, \vec{p}^u) \geq \Delta_k(i, \vec{p}^u) + \epsilon_k^u$ with $\epsilon_k^t = 0$ for all $\{i, \vec{p}^d\}$, $\{i, \vec{p}^u\}$ and $k = 0$, which holds trivially when at the initial iteration the value function is taken to equal zero. At iteration $k + 1$, $J_{k+1}(i, \vec{p}^d)$ can be written using the Bellman equation as

$$J_{k+1}(i, \vec{p}^d) = g(i, \vec{p}^d) - \lambda(N - i)\phi(\Delta_k(i + 1, \vec{p}^d)) + \alpha \{(1 - \gamma_1^k - \gamma_2^k)J_k(i, \vec{p}^d) + \mu i_k \Delta_k(i, \vec{p}^d) + \gamma_1^k J_k(i, \vec{p}^d + \Delta\nu) + \gamma_2^k J_k(i, \vec{p}^d - \Delta\nu)\}.$$  

(48)

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

$$\Delta_k(i + 1, \vec{p}^d) = J_{k+1}(i + 1, \vec{p}^d) - J_{k+1}(i, \vec{p}^d),$$

$$= [g(i + 1, \vec{p}^d) - g(i, \vec{p}^d)] + \alpha \{(1 - \gamma_1^k - \gamma_2^k)\Delta_k(i + 1, \vec{p}^d) - \mu i_k \Delta_k(i, \vec{p}^d) + \gamma_1^k \Delta_k(i, \vec{p}^d + \Delta\nu) + \gamma_2^k \Delta_k(i, \vec{p}^d - \Delta\nu)\} - \lambda(N - i - 1)\phi(\Delta_k(i + 2, \vec{p}^d)) + \lambda(N - i)\phi(\Delta_k(i + 1, \vec{p}^d)).$$

(49)

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

$$\Delta_k(i + 1, \vec{p}^d) - \Delta_k(i, \vec{p}^d) = g(i + 1, \vec{p}^d) - 2g(i, \vec{p}^d) + g(-i, \vec{p}^d) + \alpha \{(1 - \gamma_1^k - \gamma_2^k)\Delta_k(i + 1, \vec{p}^d) - \mu i_k \Delta_k(i, \vec{p}^d) + \gamma_1^k \Delta_k(i, \vec{p}^d + \Delta\nu) + \gamma_2^k \Delta_k(i, \vec{p}^d - \Delta\nu)\} - \lambda(N - i - 1)\phi(\Delta_k(i + 2, \vec{p}^d)) + \lambda(N - i)\phi(\Delta_k(i + 1, \vec{p}^d)).$$

(50)

Substituting the above two inequalities into (49), we get

$$\Delta_k(i + 1, \vec{p}^d) - \Delta_k(i, \vec{p}^d) \geq g(i + 1, \vec{p}^d) - 2g(i, \vec{p}^d) + g(-i, \vec{p}^d) + \alpha \{(1 - \gamma_1^k - \gamma_2^k)\Delta_k(i + 1, \vec{p}^d) - \mu i_k \Delta_k(i, \vec{p}^d) + \gamma_1^k \Delta_k(i, \vec{p}^d + \Delta\nu) + \gamma_2^k \Delta_k(i, \vec{p}^d - \Delta\nu)\} - \lambda(N - i - 1)\phi(\Delta_k(i + 2, \vec{p}^d)) + \lambda(N - i)\phi(\Delta_k(i + 1, \vec{p}^d)).$$

(51)

By mathematical induction, it is easy to show that

$$\Delta_k(i + 1, \vec{p}^d) - \Delta_k(i, \vec{p}^d) \geq \epsilon_{k+1}^t, \Delta_k(i + 1, \vec{p}^u) - \Delta_k(i, \vec{p}^u) \geq \epsilon_{k+1}^u$$

holds for all $k$ in the infinite series $\epsilon_k^t$ generated recursively by

$$\epsilon_{k+1}^t = 2\kappa + \alpha(1 - 2(\lambda + \mu))\epsilon_k^t.$$  

(52)

Since $\epsilon_0^t = 0$, $\epsilon_k^t$ must converge for $k \to \infty$. In fact, it converges to $\epsilon^t$ with

$$\epsilon^t = \lim_{k \to \infty} \epsilon_k^t = \frac{2\kappa}{1 - \alpha(1 - 2(\lambda + \mu))}.$$  

(53)

And we can hence conclude that

$$\Delta(i + 1, \vec{p}^d) - \Delta(i, \vec{p}^d) \leq \lim_{k \to \infty} \left[\Delta_k(i + 1, \vec{p}^d) - \Delta_k(i, \vec{p}^d)\right].$$  

(54)

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

$$-\phi(\Delta(i + 2, \vec{p}^d)) - \phi(\Delta(i + 1, \vec{p}^d)) \leq \alpha p_{\nu^u(i, \vec{p}^d)}\Delta(i + 2, \vec{p}^d) - \Delta(i + 1, \vec{p}^d),$$

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

Substituting the above two inequalities into (49), we get

$$\Delta_k(i + 1, \vec{p}^d) - \Delta_k(i, \vec{p}^d) \leq g(i + 1, \vec{p}^d) - 2g(i, \vec{p}^d) + g(-i, \vec{p}^d) + \alpha \{(1 - \gamma_1^k - \gamma_2^k)\Delta_k(i + 1, \vec{p}^d) - \mu i_k \Delta_k(i, \vec{p}^d) + \gamma_1^k \Delta_k(i, \vec{p}^d + \Delta\nu) + \gamma_2^k \Delta_k(i, \vec{p}^d - \Delta\nu)\} - \lambda(N - i - 1)\phi(\Delta_k(i + 2, \vec{p}^d)) + \lambda(N - i)\phi(\Delta_k(i + 1, \vec{p}^d)).$$

(55)

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

Denote $\nu = (N - N_1)$ and $\epsilon_{k+1}^u = 2\kappa + \alpha(1 - 2(\lambda + \mu))\epsilon_k^u$. It can be now seen that

$$\Delta_{k+1}(i + 1, \vec{p}^d) - \Delta_{k+1}(i, \vec{p}^d) \geq \epsilon_{k+1}^u$$

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

$$\Delta_{k+1}(i + 1, \vec{p}^u) - \Delta_{k+1}(i, \vec{p}^u) \geq \epsilon_{k+1}^u.$$  

(56)

By mathematical induction, we conclude that

$$\Delta_k(i + 1, \vec{p}^d) - \Delta_k(i, \vec{p}^d) \leq \epsilon_k^u, \Delta_k(i + 1, \vec{p}^u) - \Delta_k(i, \vec{p}^u) \leq \epsilon_k^u$$

holds for all $k$ in the infinite series $\epsilon_k^u$ generated recursively by

$$\epsilon_{k+1}^u = 2\kappa + \alpha(1 - 2(\lambda + \mu))\epsilon_k^u.$$  

(57)

Since $\epsilon_0^u = 0$, $\epsilon_k^u$ must converge as $k \to \infty$. In fact, it converges to $\epsilon^u$ with

$$\epsilon^u = \lim_{k \to \infty} \epsilon_k^u = \frac{2\kappa}{1 - \alpha(1 - 2(\lambda + \mu))}.$$  

(58)
Hence we have
\[
\Delta(i+1, \tilde{p}^d) - \Delta(i, \tilde{p}^d) = \lim_{k \to \infty} \Delta_k(i+1, \tilde{p}^d) - \Delta_k(i, \tilde{p}^d),
\]
and this concludes the proof of Proposition 3.

**APPENDIX B**

**PROOF OF PROPOSITION 4**

Proof. (1) We prove this by induction. Assuming that \(\Delta_k(i, \tilde{p}^d) \geq \Delta_k(i+1, \tilde{p}^d + \Delta y)\) and \(\Delta_k(i, \tilde{p}^d) \geq \Delta_k(i+1, \tilde{p}^d + \Delta y)\) for \(k = 0\), following the same procedure as in Proposition 3 we have
\[
\Delta_{k+1}(i, \tilde{p}^d) - \Delta_{k+1}(i+1, \tilde{p}^d + \Delta y) = \alpha \left\{ \left[\Delta_k(i, \tilde{p}^d) - \Delta_k(i+1, \tilde{p}^d + \Delta y)\right] - \mu(N - i + 1 - \mu) \right\}.
\]

Therefore (61) becomes
\[
\Delta_{k+1}(i, \tilde{p}^d) - \Delta_{k+1}(i+1, \tilde{p}^d + \Delta y) \geq \alpha \left[\Delta_k(i, \tilde{p}^d) - \Delta_k(i+1, \tilde{p}^d + \Delta y)\right] - \mu(N - i + 1 - \mu),
\]
and the desired result holds.

(2) We assume \(\Delta_k(i+1, \tilde{p}^d + \Delta y) + \tilde{e}_k^u \geq \Delta_k(i, \tilde{p}^d)\) and \(\Delta_k(i+1, \tilde{p}^d + \Delta y) + \tilde{e}_k^u \geq \Delta_k(i, \tilde{p}^d)\) hold with \(\epsilon_k^u = 0\) for all \(\{i, \tilde{p}^d\}\) and \(k = 0\). According to Lemma 2
\[
\phi(\Delta_k(i, \tilde{p}^d)) - \phi(\Delta_k(i+1, \tilde{p}^d + \Delta y)) \leq \alpha \phi(\Delta_k(i, \tilde{p}^d)) - \phi(\Delta_k(i+1, \tilde{p}^d + \Delta y)) \leq -\alpha \phi(\Delta_k(i, \tilde{p}^d)) - \phi(\Delta_k(i+1, \tilde{p}^d + \Delta y)).
\]

Substituting (63) into (61) we obtain
\[
\Delta_{k+1}(i, \tilde{p}^d) - \Delta_{k+1}(i+1, \tilde{p}^d + \Delta y) \leq \alpha \left\{ \left[1 - \gamma_k' - \gamma_k'' \right] \left[\Delta_k(i, \tilde{p}^d) - \Delta_k(i+1, \tilde{p}^d + \Delta y)\right] + \mu(N - i + 1 - \mu) \right\}.
\]

Since \(\Delta_k(i+1, \tilde{p}^d + \Delta y) + \tilde{e}_k^u \geq \Delta_k(i, \tilde{p}^d)\) and \(\Delta_k(i, \tilde{p}^d) + \epsilon_k^u \geq \Delta_k(i+1, \tilde{p}^d)\), we have
\[
\Delta_k(i+1, \tilde{p}^d + \Delta y) + \tilde{e}_k^u \geq \Delta_k(i+1, \tilde{p}^d + \Delta y) \geq \Delta_k(i, \tilde{p}^d) \geq \Delta_k(i+1, \tilde{p}^d) + \epsilon_k^u \geq \Delta_k(i+1, \tilde{p}^d + \Delta y)
\]
for all \(\{i, \tilde{p}^d\}\). Substituting (65) into (64) we get
\[
\Delta_{k+1}(i, \tilde{p}^d) - \Delta_{k+1}(i+1, \tilde{p}^d + \Delta y) \leq \alpha \left[1 - (\lambda + \mu)\right] \tilde{e}_k + \alpha (\lambda + \mu) \epsilon_k.
\]

Defining the recursive series
\[
\tilde{e}_{k+1}^u = \alpha \left[1 - (\lambda + \mu)\right] \tilde{e}_k + \alpha (\lambda + \mu) \epsilon_k,
\]
(67) it follows that \(\Delta_{k+1}(i, \tilde{p}^d) - \Delta_{k+1}(i+1, \tilde{p}^d + \Delta y) \leq \tilde{e}_{k+1}^u\) holds for all \(k\). We can also verify that the infinite series \(\tilde{e}_k^u\) converges to \(\bar{e}^u\) with \(\tilde{e}_k^u = \lim_{k \to \infty} \tilde{e}_k^u = \frac{\alpha (\lambda + \mu)}{1 - (\lambda + \mu)} \bar{e}^u\) where \(\bar{e}^u\) is defined as in Proposition 3. Based on the convergence property of \(\Delta(i, \tilde{p}^d)\), we can conclude that \(\Delta(i, \tilde{p}^d) - \Delta(i+1, \tilde{p}^d + \Delta y) \leq \bar{e}^u\). The two parts discussed above completes the proof of Proposition 4. □

**REFERENCES**

[1] Yuri V. Makarov, Clyde Loutian, Jia 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 nebraska system. In Power and Energy Society General Meeting. 2012 IEEE, pages 1–7, 2012.

[3] D.A. Halamanay, 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, 3(3):321–328, 2011.

[4] Donald J Hammerstrom, Jerry Brous, David P Chasson, Gale R Horst, Robert Kajfetz, Preston Michie, Terry V Oliver, Teresa A Carlson, Conrad Eustis, Olof M Jarvegren, et al. Pacific northwest grid wide 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 thermostat 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] 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.

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

[14] Lingwen Gan, Ufuk Topcu, and Steven Low. Optimal decentralized protocol for electric vehicle charging. In Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, pages 5798–5804. IEEE, 2011.
[15] M.D. Ilic, Le Xie, and Jhi-Young Ioo. Efficient coordination of wind power and price-responsive demand – part i: Theoretical foundations. Power Systems, IEEE Transactions on, 26(4):1875–1884, Nov 2011.

[16] 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.

[17] 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.

[18] 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.

[19] Frank P Kelly. Charging and rate control for elastic traffic. European Transactions on Telecommunications, 8(1):33–37, 1997.

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

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

[22] 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.

[23] Ratepayer revolt in Pacific Northwest. https://news.google.com/newspapers/nid=1916&date=19820305&id=lJJJAAAIABAJ&sjid=nAYNAAAAIBAJ&pg=3879,958338&hl=en

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

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

[26] Bowen Zhang and John Baillieul. Control and communication protocols based on packetized direct load control in smart building microgrids. Proceedings of the IEEE, 104(4):837–857, 2016.

[27] 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.

[28] PJM regulation signal: http://pjm.com/markets-and-operations/ancillary-services/mkt-based-regulation.aspx.

[29] 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 (Member, IEEE) received the Ph.D. degree in Systems Engineering from Boston University, Boston, MA, USA in 2015, and the B.E. degree in Control Science and Engineering from Zhejiang University in 2010. He is now an operations research analyst at Cambridge Energy Solutions. 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. He has worked on the qualitative theory of electric energy systems, and 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. He is currently Distinguished Professor of Engineering at Boston University.

Dr. Baillieul is a Fellow of the International Federation of Automatic Control (IFAC) and the Society for Industrial and Applied Mathematics (SIAM).

John Baillieul (Life Fellow, IEEE) received the Ph.D. in applied mathematics from Harvard University, Cambridge, MA, USA, in 1975, where he was a student of Roger W. Brockett.

He has published research dealing with robotics, the control of mechanical systems, and more generally with mathematical system theory. 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. He has worked on the qualitative theory of electric energy systems, and 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. He is currently Distinguished Professor of Engineering at Boston University.