Utility Learning Model Predictive Control for Personal Electric Loads

Insoon Yang* Melanie N. Zeilinger* Claire J. Tomlin

Abstract—A personalized control framework that tightly combines online learning of the energy consumer’s utility function and the control of the consumer’s electric loads according to real-time updates of the utility is proposed. This framework is particularly useful to automatically customize the controller of electric loads that directly affect the consumer’s comfort. Because the utility function is identified and predicted online using Gaussian process regression, the controller is capable of immediately setting its objective function to the learned utility function and of adjusting its control action to maximize the new objective. Furthermore, no separate training period to learn the consumer’s utility is needed. The performance of the proposed method is demonstrated by the application to a personalized thermostat controlling indoor temperature.

I. INTRODUCTION

One of the major changes expected for the future electric power grid is a transition towards demand side management, with objectives including (i) the enhancement of energy efficiency of a consumer’s electric loads and (ii) the guiding of a consumer’s electricity usage patterns or flexibility via energy pricing and/or incentive payment schemes [1]. The latter is also called demand response. Increasing controllability of a consumer’s loads with appropriate control schemes is expected to play a key role in achieving these objectives [2]. A number of sophisticated control and scheduling methods have been proposed for various types of personal loads, such as electric vehicles [3], [4], deferrable loads [5], [6] and thermostatically controlled loads [7], [8], [9]. In addition to achieving good energy efficiency or demand response, a main task of the control method is to ensure that each consumer’s comfort does not deteriorate. Personalized control of loads taking into account the consumer’s satisfaction has recently shaped up as a successful concept for addressing this goal. The Nest smart thermostat [10] can be taken as one example.

This work proposes a novel personalized control framework that is suitable for consumer’s electric loads. The proposed method has the capability of learning online the consumer’s utility function that represents her satisfaction (e.g., comfort and energy saving) with the controlled personal loads. More specifically, the consumer reports her satisfaction with the control performance, e.g., in the form of a simple rating, say a number between 1 and 5, through a given interface between the consumer and the controller at any convenient time. For example, the consumer can rate the controller of her air conditioner judging from the amount of energy savings and her comfort with the indoor temperature and report it to the controller via her smartphone. Using the satisfaction data, the proposed method infers the consumer’s utility function that is used to automatically customize the controller for the consumer.

Most existing control and scheduling methods for electric loads do not explicitly consider the utility function of a consumer, or the methods assume that the utility function is given and fixed. Considering objective (i) of demand side management, a consumer may want to frequently tune the controller for her loads to find an optimal tradeoff between energy efficiency and comfort. However, tuning the controller by herself can be cumbersome. To achieve objective (ii), an aggregator provides direct or indirect load control programs. Under the direct load control program, in which an aggregator has the authority to control its consumer’s load, neglecting a consumer’s utility function can make her feel uncomfortable with and inconvenienced by her electric loads controlled for demand side management. Direct load control methods with a fixed utility function may result in a similar effect if the true utility deviates from the fixed utility. The resulting dissatisfaction with the usage of her loads by the aggregator can lead the consumer to drop out of the direct load control program and is therefore undesirable for both parties. Under the indirect load control program, an aggregator or a load serving entity designs pricing or reward schemes to incentivize its consumer to control her loads in a way that is desirable to the aggregator. However, a consumer may not control her loads in an expected way if her true utility deviates over time from the nominal utility function adopted in the design of the pricing or incentive schemes. These concerns motivate the need for a personalized control scheme that takes into account the consumer’s utility function and allows it to vary over time.

In the proposed framework, the consumer’s utility is learned online from her satisfaction data and the control objective is changed accordingly. The approach is suitable even for the situation in which the consumer’s behavior affects the system controlled by her loads. For example, opening a window affects the indoor temperature controlled by the consumer’s air conditioner. To achieve online learning of the consumer’s utility and behavior in this setting, we propose a system manager, which is interfaced with the consumer and the electric load. The system manager consists of a
behavior learning module, a controller and a utility learning module. Intuitively speaking, the behavior and utility learning modules infer the consumer’s behavior and utility from the state measurements and the satisfaction data, respectively, and deliver the inferred values to the controller that uses them to compute personalized control actions. More specifically, the behavior learning module estimates and predicts online the effect of the consumer’s behavior, which is not known a priori or cannot be modeled explicitly, on the system state controlled by her loads, e.g., indoor temperature. A Gaussian process (GP) model is used to learn the effect from online state measurements and construct a stochastic prediction [11], [12], [13]. The prediction of the consumer’s behavior enhances the performance of the model predictive control (MPC) [14], [15], [16], [17], which is used as a controller in the proposed framework. The objective of the MPC controller is chosen as the consumer’s utility function, which is inferred by the utility learning module. The identification of the consumer’s utility is performed online by solving a convex optimization problem whenever the consumer reports a new satisfaction data to the system manager. Furthermore, the module provides an option to learn the user’s utility online as a time-varying function using GP regression and therefore to generate an estimate of the utility at all sampling times, i.e., also when no satisfaction data is available. The MPC controller immediately sets the updated consumer’s utility as its objective function and thereby modifies the resulting control law according to her preferences. We call this framework the utility learning model predictive control.

The most distinctive feature of the proposed framework is that the consumer’s utility is learned online during closed-loop control. Therefore, (i) the controller is capable of immediately updating its objective function as the identified or predicted consumer’s utility and of controlling the system to maximize the new objective, and (ii) no separate training period to learn the consumer’s utility is needed. These are advantages of the proposed framework, which tightly integrates the utility learning and system control, over offline utility learning approaches (e.g., [18], [19]).

The rest of the paper is organized as follows: the setup of the utility learning MPC is proposed in Section II. We then introduce the methods for learning the consumer’s behavior and her utility online and for combining them with the MPC in Section III. The performance of the proposed method is demonstrated with numerical tests for personalized thermostats in Section IV.

II. THE SETUP

We begin by describing the setting of the utility learning model predictive control framework, which consists of the consumer, the load (and the system controlled by the load), and the system manager.

A. System Model

Consider an electric load of a consumer, e.g., a water heating system, an air conditioner or a CO\textsubscript{2} controller. Let $x_{\tau} \in \mathbb{R}^{n}$ be the system state, e.g., the water temperature, the indoor temperature, or the indoor CO\textsubscript{2} level, at time $\tau \Delta t$, $\tau = 0, 1, \cdots$ for some positive time step $\Delta t$. We let $u_{\tau} \in \mathbb{R}^{m}$ be the control input, e.g., the power consumption by the load, and assume that the system can be represented or at least well approximated by linear system dynamics of the form

$$x_{\tau+1} = Ax_{\tau} + Bu_{\tau} + z_{\tau} + w_{\tau},$$

where $w_{\tau} \in \mathbb{R}^{n}$ models the exogenous (or environmental) uncertainty (with a known probability distribution) that affects the system. Here, $z_{\tau} \in \mathbb{R}^{n}$ represents an unknown effect of the consumer’s behavior on the system dynamics. The consumer’s behavior is assumed to be independent of the system state but to be dependent on time. This effect can make the system model arbitrarily wrong if we do not have a good knowledge (or estimation) of the consumer’s behavior. Neglecting this effect can cause a significant model mismatch and deteriorate the closed-loop performance of the MPC controller.

B. Consumer’s Utility Function

The consumer has multiple objectives to control her load. Most commonly, she wants to maximize her utility associated with energy savings and, at the same time, maximize her utility associated with her comfort/convenience. To mathematically model the $i$th objective of the consumer, we choose a basis utility function, $\beta_{i} : \mathbb{R}^{n} \times \mathbb{R}^{m} \rightarrow \mathbb{R}$, for each $i = 1, \cdots, K$. Then, the total utility function of the consumer (at time $\tau \Delta t$) can be modeled as the weighted sum of the basis utility functions, i.e.,

$$J(x_{\tau}, u_{\tau}; \theta) := \sum_{i=1}^{K} \theta^{i} \beta_{i}(x_{\tau}, u_{\tau}),$$

where $\theta^{i} \in \mathbb{R}$ is the weight of the $i$th basis. The vector of weights is denoted by $\theta \in \mathbb{R}^{K}$. Here, the basis functions are assumed to be concave. This assumption is valid in many practical problems.

Example 1. Let $x_{\tau}$ and $u_{\tau}$ be the indoor temperature and the power consumption by the consumer’s air conditioner at time $\tau \Delta t$, respectively. Let $c_{r}$ (dollar/kWh) be the energy price; then, the consumer’s utility function with respect to energy costs can be modeled as the negative value of energy
costs, i.e., \( \beta_1(x, u) := -c_x u \Delta t \). Assuming that the user feels most comfortable at the temperature \( \hat{x} \), the consumer's utility function with respect to her comfort can be formulated as \( \beta_2(x, u) := -(x - \hat{x})^2 \Delta t \). This is a modified version of the comfort metric defined in terms of temperature deviation proposed in the ANSI/ASHRAE standards [20]. Note that these two basis utility functions are concave. These basis functions are employed in the numerical tests in Section IV.

The task is then to learn the weights to identify and infer the consumer’s utility. Learning the weights online from appropriate data is essential to align the system manager’s objective with the consumer’s actual objective. Given the basis weights, the system manager synthesizes a controller that maximizes the utility function in a receding horizon approach as will be shown in Section III.

C. Information Flow

The information flow between the consumer, the system controlled by the load, and the system manager is depicted in Fig. 1. The consumer’s behavior affects the system state (e.g., indoor temperature) and this effect is represented by \( z \) as proposed. The system is also perturbed by the exogenous disturbance or noise, \( w \), and the current state information, \( x \), is delivered to all three modules of the system manager: (i) the behavior learning module; (ii) the controller; and (iii) the utility learning module.

In the behavior learning module, the consumer’s effect, \( z \), is learned by Gaussian process (GP) regression. The inferred value of \( z \) is denoted as \( \hat{z} \). Weakly periodic behavior of the consumer is informative when designing a kernel of the GP model with good performance.

Judging from the system state and control during \( (\tau_{j-1}, \tau_j) \) (and/or even before the interval), the consumer can report her satisfaction, \( s_j \in \mathbb{R} \), to the system manager, where \( \tau_{j-1} \) is the time at which satisfaction was last reported and \( \tau_j \) is the current time as illustrated in Fig. 2. Therefore, \( \{\tau_j\}_{j=1,...} \) is a subsequence of the MPC sampling time steps \( \tau = 0, 1, \cdots \). Note that the reporting time can be arbitrarily chosen by the consumer. For example, the user rates the performance of her smart thermostat as a number, \( s_{j_i} \), between 1 and 5 based on the amount of energy saving and her comfort with the indoor temperature during \( (\tau_{j-1}, \tau_j) \) and reports it to the thermostat via her smartphone. The utility learning module then identifies the basis weights given the satisfaction data \( \{s_j, s_{j-1}, \cdots, s_{M+1}\} \), where \( M \) can be chosen by the consumer. The identified basis weight vector is denoted as \( \theta^* \). The module also provides an option to learn the basis weight vector as a function in time using GP regression. The inferred value of \( \theta \) is denoted as \( \hat{\theta} \). The utility identification and learning procedures will be discussed in Section III-C.

The controller of the system manager receives the information on the system state, \( x \), the inferred behavioral effect, \( \hat{z} \), and the identified (or inferred) basis weights, \( \theta^* \) (or \( \hat{\theta} \)). It then generates a control signal \( u \) using an MPC approach that maximizes the updated user’s utility.

III. UTILITY LEARNING MODEL PREDICTIVE CONTROL

We now propose the utility learning model predictive control method under the setting described in the previous section. The proposed method builds on the three main building blocks: the controller, and the behavior and utility learning modules. The details for each of these components are discussed in the following.

A. Model Predictive Control with Learned Consumer’s Utility Function and Behavior

A model predictive control scheme is proposed, which allows to tightly couple the three blocks and compute a control input minimizing the cost (or maximizing the utility) obtained from the utility learning module, subject to the imposed system constraints and the prediction of the model dynamics inferred in the behavior learning module. The system (1) is stochastic due to the uncertain consumer’s effect as well as the exogenous uncertainty. To take into account the stochastic dynamics, a stochastic MPC problem is formulated:

\[
\begin{align*}
\min_{u_t \in \mathbb{R}^m} \quad & - \sum_{i=\tau}^{\tau+N-1} \mathbb{E}[J(x_t, u_t; \theta^*)] \\
\text{subject to} \quad & x_{\tau+i} = x_{\tau+i}^* \\
& x_{\tau+i+1} = Ax_{\tau+i} + Bu_{\tau+i} + \hat{z}_{\tau+i} + w_{\tau+i}, \\
& u_{\tau+i} \in \mathcal{U}, \quad i = 0, \cdots, N - 1,
\end{align*}
\]

where \( x_{\tau+j}^* \) is the system state measured/estimated at time step \( \tau \), \( u_{\tau+j} := \{u_i\}_{i=\tau}^{\tau+N-1} \) is the sequence of control inputs and \( N \) is the prediction horizon. \( \mathcal{U} \) is a compact set, defining the set of admissible control values. It should be noted that state constraints are not explicitly incorporated in the MPC problem. In the application context of electric loads, it is assumed that state constraints are introduced as soft constraints and included as a basis function. Note that with a slight abuse of notation we use \( \theta^* \) in the MPC cost as the weight provided by the utility learning module for both updating strategies proposed in Section III-C.

By using the mean and variance of \( \hat{z}_{\tau+i} \) over the prediction horizon provided by the behavior learning module (see Section III-B) and the given probability distribution of \( w_{\tau+i} \), the stochastic MPC problem can be reduced to a deterministic formulation. In the case of a quadratic stage cost on the states and inputs, for instance, the stochastic MPC problem corresponds to employing the expected values \( \mathbb{E}[\hat{z}_{\tau+i}] \), \( \mathbb{E}[w_{\tau+i}] \) in the dynamics. This can be easily seen from \( \mathbb{E}[x_t^\top Q x_t] = \text{tr}(Q \text{Var}[x_t]) + \mathbb{E}[x_t] \top Q \mathbb{E}[x_t] \) and the fact that \( \text{Var}[x_t] = \text{Var}[\hat{z}_t] + \text{Var}[w_t] \) is known or provided by the prediction and hence a given constant at the time of computation. Let \( u_{\tau+i}^* \) be the solution to the
optimization problem at time step $\tau$. The MPC control law, $\kappa: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, is defined in a receding horizon fashion by applying the first control input, i.e., $\kappa(x, \tau) = u^*_\tau$.

B. Learning The Consumer’s Behavior

The effect of the consumer’s behavior is learned from data collected during online control. We make use of a Gaussian process model, providing a general non-parametric modeling framework and a posterior uncertainty description. Informally, a Gaussian process (GP) can be thought of as describing a probability distribution over functions. Due to the fact that human behavior generally follows a daily routine, the consumer’s effect can be assumed to exhibit dynamics with periodic characteristics, which may, however, change over multiple period lengths. A GP model for capturing ‘locally periodic’ dynamical effects has been recently proposed and incorporated in an MPC controller in [13].

A function $g$ is here said to be ‘locally periodic’ provided that $g(t) \equiv g(t + n\omega)$ if $n\omega \ll l$ but $g(t) \not\equiv g(t + n\omega)$ if $n\omega \gg l$, where $\omega$ is the period length, $l$ is a measure of locality and $n = 1, 2, \ldots$. This approach can be directly employed to learn the consumer’s effect online as described in the following. We only discuss the main steps of GP regression, more details about Gaussian processes can be found in [11], see also [12] for a more general overview of kernel methods for system identification.

Consider the case of learning a scalar function $g_i : \mathbb{R} \to \mathbb{R}$ for $z^i_t = g_i(\tau \Delta t)$, where $z^i_t$ denotes the $i$th element of the vector $z_t$. A Gaussian process model is defined by a mean function $\mu : \mathbb{R} \to \mathbb{R}$ and a covariance function $k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, which together form the GP prior. The choice of this prior is an important modeling assumption as it determines the properties of the functions considered for $z^i_t$ and in turn affects convergence and extrapolation properties. In the particular case of time varying functions, extrapolation is crucial because future function values have to be predicted from past observations.

A prior covariance function focusing probability mass on periodic functions has been recently presented in [13], which is composed of a square exponential kernel $k_{SE}(t, t')$ and a periodic kernel $k_{P}(t, t')$ for $t, t' \in \mathbb{R}$:

$$k(t, t') = \sigma^2 \cdot k_{SE}(t, t') \cdot k_{P}(t, t'),$$  \hspace{1cm} (3)

where $k_{SE}(t, t') = \exp\left(-\frac{(t - t')^2}{2l^2_{SE}}\right)$ with length scale $l_{SE}$, $k_{P}(t, t') = \exp\left(-2\sin^2\left(\frac{\pi}{l_{P}}(t - t')\right)/l^2_{P}\right)$ with length scale $l_{P}$ and period length $l_{P}$, and $\sigma^2$ is the signal variance. Intuitively speaking, this kernel considers two input times similar if they are similar under both the square exponential and the periodic kernel. The square exponential kernel admits the function to be not strictly periodic and for $l_{SE} \gg \omega$ to vary over a longer time scale. As shown in [13], this kernel has the significant advantage that it offers good extrapolation properties for functions with this property. A function drawn from a GP with this kernel is shown for the application example in Fig. 6, a more detailed discussion and illustration of the different parameters can be found in [13]. The role of the mean function is less critical. Since no prior hypothesis about the mean of the consumer behavior is available, the mean function is here chosen to zero.

After choosing the prior hypothesis class, GP regression is performed by inferring the function from observed data. Samples of the consumer effect $z_t$ are obtained from the dynamical model in (1) by measuring the state $x_t$ at two consecutive sampling times. We denote by $\hat{z}_{T}^j \in \mathbb{R}^D$ the vector of observations at time points $T \in \mathbb{R}^D$. If data points are observed with Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$, i.e., $z^i_t = z^i_{t'} + \epsilon_i$, $i = 1, \ldots, n$, then, given the Gaussian process prior, the posterior distribution is also Gaussian [11]. For a given test point $t = j \Delta t$, the resulting predictive distribution with mean $\hat{z}^j_T$ and variance $\mathbb{V}(z^j_T)$ is given by:

$$\hat{z}^j_T = \mu(t) + K(t, T)^\top (K(T, T) + \sigma^2 I)^{-1}(\hat{z}_T - \bar{\mu}(T)),$$  \hspace{1cm} (4a)

$$\mathbb{V}(z^j_T) = K(t, T) - K(t, T)^\top (K(T, T) + \sigma^2 I)^{-1}K(T, t),$$  \hspace{1cm} (4b)

where $[K(T, T')_{mn} := k(T_m, T'_n), \bar{\mu}(T)]_{m} := \mu(T_m)$. The posterior mean and variance completely characterize the inferred value, $\hat{z}_T^j$, of $z^j$. By means of (4), a prediction of the mean and variance of $\hat{z}_T^j$, $j = \tau, \ldots, \tau + N - 1$, is constructed for $i = 1, \ldots, n$ and is incorporated in the MPC problem.

The covariance function has several free hyperparameters, which may be difficult to choose in practice. While in the case of human behavior the period length can be chosen from intuition as $24h$, choosing good values for the remaining hyperparameters is important to provide a good model. In GP regression, hyperparameters are often inferred from the available training data. In the considered case of online learning, inference is performed recurrently during closed-loop operation, after a new batch of $L$ data points has been collected, where $L$ is a tuning parameter. Different techniques for hyperparameter inference are available, see [11] for a detailed discussion, or [13] for a customized approach for periodic kernels.

C. Learning The Consumer’s Utility

As proposed in Section II-C, the consumer can provide the satisfaction data to the system manager at any sampling time. We propose two techniques for utility learning in the following. The first approach is based on an immediate identification of the weights as soon as new satisfaction data is available, after which the weight is kept constant until the next update. The second approach additionally learns time varying functions for the basis weights, which are used to predict the weights in the future and are updated with every new satisfaction data point. As shown in Fig. 2, let $s_j \in \mathbb{R}$ be the consumer’s satisfaction with the system operation during $[t_{j-1}, t_j] = (\tau_{j-1} \Delta t, \tau_j \Delta t)$. We also denote

$$d_j := \sum_{\tau = \tau_{j-1} + 1}^{\tau_j} (\beta_1(x^*_\tau, u^*_\tau), \ldots, \beta_K(x^*_\tau, u^*_\tau)) / (\tau_j - \tau_{j-1} - 1).$$  \hspace{1cm} (5)

In other words, $d_j \in \mathbb{R}^K$ represents the averaged basis function with the MPC control over $[t_{j-1}, t_j]$. 

At time step $\tau_j$, the system manager considers the following data:
\[
\{(d_{j-M+1}, s_{j-M+1}), \ldots, (d_j, s_j)\},
\] (6)
where the parameter $M$ determines how many data points in the past are taken into account, which can be chosen by the consumer. If the consumer wants the system manager to consider her most recent previous satisfaction data and the current satisfaction data, for example, then she should choose $M = 2$.

1) **Online identification of the consumer’s utility:** From the data, the system manager estimates the basis weights $\theta^*$ at time step $\tau_j$ by solving the following optimization problem:
\[
\theta^*_{\tau_j} = \arg \min_{\theta \in \mathbb{R}^K} \sum_{i=j-M+1}^j (d_i^\top \theta - s_i)^2 + R(\theta) \tag{7}
\]
subject to $\theta_L \leq \theta \leq \theta_U$, where the lower and upper bounds $\theta_L$ and $\theta_U$ of the basis weights, respectively, can be chosen by the consumer. Here, $R : \mathbb{R}^K \to \mathbb{R}$ is a strictly convex function, which can be interpreted as a regularizer to guarantee the uniqueness of the solution. Let $D := (d_{j-M+1}, \ldots, d_j)^\top \in \mathbb{R}^{M \times K}$, and $s := (s_{j-M+1}, \ldots, s_j) \in \mathbb{R}^{M}$. Then, the cost function can be rewritten as
\[
\sum_{i=j-M+1}^j (d_i^\top \theta - s_i)^2 + R(\theta) = \| D \theta - s \|^2 + R(\theta).
\]
If rank($D$) $\geq K$, then the optimization problem (7) has a unique solution even if $R = 0$. However, the rank condition is not guaranteed, even when $M \geq K$, because of the possibility that the linearly independent data in (6) are less than $K$. Therefore, an appropriate strictly convex function $R$ needs to be chosen to guarantee the uniqueness of the solution to (7). In the examples in Section IV, we choose $R$ as a ‘smoothness-inducing function’. More specifically, set
\[
R(\theta) = \lambda \| \theta - \theta^*_{\tau_{j-1}} \|^2,
\]
where $\theta^*_{\tau_{j-1}}$ indicates the basis weight vector identified at the last utility learning time, i.e., $\tau_{j-1} \Delta t$, and $\lambda$ is a positive constant. This regularization function encourages the smoothness in the variation of the basis weights, i.e., it discourages abrupt jumps in the objective (utility) function in the MPC module. It is important to note that the utility identification is performed online. Therefore, the proposed method is ideal when the consumer’s utility function changes over time and these changes are critical in controlling the load.

2) **Online learning of the consumer’s utility:** Using the identified weights, a time-varying function for $\theta$ can be learned online, which allows to predict the weight vector for all time points, i.e., even when no satisfaction data is available. A GP model as described in Section III-B is employed, where the observed data points in this case are given by the identified weight vectors, $\{\theta^*_{\tau_1}, \ldots, \theta^*_{\tau_j}\}$,

\[\text{defined in (7). Due to the overall periodicity of human behavior, the consumer’s preference can again be assumed to vary periodically and the kernel in (3) is employed. All weights are initially assumed to be equivalently important by choosing a constant mean equal to one, i.e. } \mu(t) = 1, \forall t. \text{ Let } T_j := (t_1, \ldots, t_j) \in \mathbb{R}^j, \text{ where } t_l := \tau_l \Delta t, l = 1, 2, \ldots. \text{ For } i = 1, \ldots, K, \text{ let } \theta^i_{\tau_j} \in \mathbb{R}^j \text{ denote the vector of observations for the } i\text{th weight at times } T_j, \text{ i.e., } \theta^i_{\tau_l} := (\theta^i_{\tau_1}, \ldots, \theta^i_{\tau_l}). \text{ We also assume that } \theta^i_{\tau_l} = \theta^i_{\tau_1} + c^i, l = 1, 2, \ldots, \text{ where } c^i \sim N(0, \sigma^2_i). \text{ An estimation of the weight vector at time } t = (\tau_j + l) \Delta t \in \mathbb{R}, l = 0, 1, 2, \ldots \text{ is obtained by using the mean predictive equation for the inferred value } \hat{\theta}^i_{\tau_{j+l}}:
\]
\[
\hat{\theta}^i_{\tau_{j+l}} = K_t(t, T_j)^\top (K_t(T_t, T_j) + \sigma^2_i I)^{-1} (\hat{\theta}^i_{T_j} - \hat{\mu}(T_j)) + \mu(t),
\]
(8)
where $k_i(\cdot, \cdot)$ is the covariance function in (3), $\mu(t)$ is the mean function of $\theta^i$, $[K_t(T_t, T_j)]_{mn} := k_i(t_m, t_n')$ and $\hat{\mu}(T_j) := \mu(t_m)$. Equation (8) is used to generate the predictions, $\hat{\theta}^i_{\tau_{j+l}}$’s, and the GP model is updated whenever a new satisfaction data is reported. At each sampling time, $\tau_j + l$, we use the basis weight in the MPC objective function (2) as $\hat{\theta}^i_{\tau_{j+l}}$ for the entire prediction horizon, provided that $\tau_j + l < \tau_{j+1}$, as shown in Fig. 3.

**D. Algorithm**

We summarize the description of the proposed utility learning model predictive control in Algorithm 1. The behavior learning module generates the forecast of the consumer’s effect at $N-1$ future time points by performing the Gaussian process regression discussed in Section III-B. The prediction of the consumer’s behavior is used in the MPC (2) to compute the control signal in a receding horizon fashion. As soon as the consumer reports her satisfaction, the utility learning module kicks in and solves the problem (7) to identify the consumer’s utility. The identified basis weights are used in the MPC from the next time step until a new satisfaction data point is provided if the prediction of the weights is not employed. If the basis weight vector is learned as a time-varying function by GP prediction (optional), on the other hand, the predicted basis weight vector, $\hat{\theta}_c$, which
Algorithm 1: Utility learning model predictive control

1 Initialization:
2 \( \theta^* \leftarrow \theta_0; \)
3 \( L, N, M \) selected by the consumer;
4 for \( \tau = 0, 1, \cdots \) do
5 \( x^\tau \leftarrow \text{current measured/estimated state} \)
6 Behavior learning module:
7 if \( \text{mod}(\tau, L) = 0 \) then
8 \( \text{Perform hyperparameter estimation;} \)
9 end
10 \( \{\hat{x}_\tau, \cdots, \hat{x}_{\tau+N-1}\} \leftarrow \text{prediction of the consumer's effect from GP model (4);} \)
11 Controller:
12 \( u^\tau_+ := \{u^\tau_1\}_{i=\tau}^{\tau+N-1} \leftarrow \text{solution of the MPC (2);} \)
13 \( u^\tau_\star \leftarrow \hat{u}^\tau_\star; \)
14 Utility learning module:
15 if the consumer reports the \( j \)th satisfaction, \( s_j \), at time \( \tau \Delta t \) then
16 \( \tau_j \leftarrow \tau; \)
17 \( d_j \leftarrow \text{averaged basis function data (5);} \)
18 \( \theta^\tau_\star \leftarrow \text{solution of the utility identification (7);} \)
19 end
20 (optional) \( \hat{\theta}_\tau \leftarrow \text{prediction of the basis weights from GP model (8);} \)
21 end

is the mean of the inferred value \( \hat{\theta}_\tau \), computed at each sampling time step (line 21) and used as the weights in the MPC objective function (2) as shown in Fig. 3.

IV. APPLICATION TO PERSONALIZED THERMOSTATS

We consider the application of the proposed utility learning MPC scheme to the personalized thermostat that controls the consumer’s room temperature. In this scenario, the thermostat plays the role of the system manager.

A. Indoor Temperature Dynamics and Utility Functions

Let \( x_\tau \in \mathbb{R} \) be the indoor room temperature at time \( \tau \Delta t \), \( \tau = 0, 1, \cdots \). We set \( u_\tau \in \mathbb{R} \) to be the ratio between the duration in which the air conditioner (AC) is ON and the period \( \Delta t \). By definition, \( 0 \leq u_\tau \leq 1 \), i.e., \( \mathcal{U} = [0, 1] \). Frequent ON/OFF switching of the AC is not desirable because (i) it can result in physical damage to the AC; and (ii) each switching ON of the AC generates a transient spike of power consumption, which is higher than steady state power consumption [21]. To avoid frequent switching, the time step is chosen as 20 minutes, i.e., \( \Delta t = 1/3 \) h in our numerical tests. Given the (forecasted) outdoor temperature \( \Theta_\tau \in \mathbb{R} \), \( \tau = 0, 1, \cdots \), the indoor temperature dynamics can be modeled as

\[
x_{\tau+1} = x_\tau + [\alpha (\Theta_\tau - x_\tau) - \kappa u_\tau + z^\tau + w^\tau] \Delta t,
\]

where \( \alpha (\Theta_\tau - x_\tau) \) and \( w^\tau \in \mathbb{R} \) model the temperature fluctuation due to the heat transfer from outside and the effect of forecast error, respectively. Here, \( z^\tau \in \mathbb{R} \) represents an unknown effect from the consumer’s behavior on temperature, which will be learned. As explained in [22], \( \alpha := R/C \), where \( R \) is the conductance between the outdoor air and indoor air and \( C \) is the conductance between the indoor air and the thermal mass. Furthermore, the positive constant \( \kappa \) depends on the efficiency of the AC.

We assume that the consumer’s objective for the AC operation is twofold: (i) to maximize the energy cost savings and (ii) to minimize the discomfort level that is affected by the indoor temperature. Therefore, we use the basis utility functions in Example 1.

The profiles of the time-of-use energy price \( c_\tau \) and the outdoor temperature \( \Theta_\tau \) used in the simulation are shown in Fig. 4. We assume that the forecast error \( w_\tau \) is distributed with \( \mathcal{N}(0, 0.03^2) \) and that the consumer reports her satisfaction every two hours. We choose \( \alpha = 0.4834, \kappa = 2.5, \bar{x} = 22, \lambda = 10^{-10}, \theta_0^1 = \theta_0^2 = 0.004, \theta_0^4 = 2 \) and \( \theta_0^1 = (1, 1) \). The hyperparameter estimation for the GP is performed every five steps, i.e., \( L = 5 \) using the GPML toolbox [23]. The MPC prediction horizon, \( N \), is chosen as 20 and the number of data used in the utility learning module is set to \( M = 2 \).

B. Numerical Tests

We consider a consumer who prefers to pre-cool: the consumer wants to cool down the room when the energy price is low (before 12pm) and then use her air conditioner less when the energy price is high (after 12pm) as shown in Fig. 4. Suppose that the consumer reports her satisfaction with the following ‘scheme’:

\[
s_j := \begin{cases} 
\sum_{t=j-1+1}^{t=j+1} \frac{\beta_t (x^*_t, u^*_t)}{\tau_j - \tau_{j-1} - 1} & \text{if } 3 < \tau_j \Delta t < 12 \\
\sum_{t=j-1+1}^{t=j+1} \frac{\beta_t (x^*_t, u^*_t)}{\tau_j - \tau_{j-1} - 1} & \text{otherwise}.
\end{cases}
\]

In other words, the consumer cares only about the indoor temperature from 3am to 12pm. After 12pm, she wants to save energy costs as much as possible. This consumer’s preference change is correctly captured by the identified basis weights: as shown in Figs. 5 (b) and (c), \( \theta^1 \) is low and \( \theta^2 \) is high from 10am to 2pm. There is a small delay of the change in the basis weights because \( M \) is set to be 2, i.e., the utility learning module takes into account the satisfaction data from the two previous reporting intervals. Figs. 5 (b) and (c) also show the basis weight vector learned using GP. As a result of the consumer’s utility function, which is learned online,
The predicted effect of the consumer’s behavior by the GP regression is shown in Fig. 6. The true effect is shown together with the noisy samples and the mean prediction of the consumer effect over the prediction horizon obtained from the GP model at every sampling time. After having collected (noisy) data of the human behavior for only about 12h, the GP prediction already provides a very good estimate of the consumer effect.

V. CONCLUSION

We have proposed a novel personalized control framework that achieves the tight integration of online learning of the consumer’s utility and the control of the consumer’s electric loads taking into account real-time updates of the utility. The proposed method is suitable to automatically customize the controller when the consumer’s utility associated with the controlled system changes over time. We believe that this framework has great potential for the personalized control of electric loads to ensure the consumer’s comfort.

REFERENCES