Optimal Tracking Strategies for Uncertain Ensembles of Thermostatically Controlled Loads

Coimbatore Anand, Sribalaji; Baldi, Simone

DOI
10.1109/ICCA51439.2020.9264495

Publication date
2020

Document Version
Accepted author manuscript

Published in
Proceedings of the IEEE 16th International Conference on Control and Automation, ICCA 2020

Citation (APA)
Coimbatore Anand, S., & Baldi, S. (2020). Optimal Tracking Strategies for Uncertain Ensembles of Thermostatically Controlled Loads. In *Proceedings of the IEEE 16th International Conference on Control and Automation, ICCA 2020* (pp. 901-906). IEEE. https://doi.org/10.1109/ICCA51439.2020.9264495

Optimal Tracking Strategies for Uncertain Ensembles of Thermostatically Controlled Loads

Sribalaji Coimbatore Anand and Simone Baldi

***Abstract*—Demand side energy management (DSEM) promises to regulate ensembles of loads to track desired power levels, in response to grid events (demand peaks, emergencies, variable renewable power generation, etc). A large fraction of such loads are Thermostatically Controlled Loads (TCLs) such as refrigerators, electric water heaters, and air conditioners. Such loads exhibit parametric uncertainty and heterogeneity which make power tracking difficult. Adaptive control strategies are explored in this work as a way to achieve power tracking. Effectiveness of such strategies are studied via numerical simulations.**

***Index Terms*—Thermostatically Controlled Loads, Demand side energy management, Adaptive optimal control**

I. Introduction

Traditional power plants have limited ability to adapt to the varying power demands caused by the increasing deployment of renewable energy sources. Researches have put forward the idea of DSEM as a viable way to manage the power grid [1]–[5]. In DSEM a population (ensemble) is required to track a desired power level, in response to grid events such as demand peaks, emergencies, or variable renewable power generation. TCLs have a slack term on their system dynamics which makes it possible to control [6]. Such control algorithms has been studied in the literature. A few of these modeling and control approaches of TCLs are summarized below.

The authors in [7] propose a state-space model relating the offset applied to the temperature set-point of the homogeneous population of TCL (input) to the power consumed by the population (output). An observer based LQR controller is adopted to achieve power tracking. The model in [8] is a bi-linear state-space model relating the offset applied to the temperature set-point of the homogeneous population of TCL (input) to the power consumed by the population (output) and develops a non linear controller. The modeling approach in [9] is similar to [8] except for the fact that the bi-linearity is removed and included as a separate block as a part of a model predictive control algorithm. A 2 dimensional state bin model is proposed in [10] instead of 1-dimensional model as used in the previous works. A heterogeneous group of TCL consisting of smaller groups of homogeneous TCL is considered in [11] where the control is based on a hybrid partial differential equation with numerical stability analysis.

As it can bee seen from this overview, the controllers used in the literature for TCL are model-based. In general a system model can be hard to obtain [12]. This leads to an opportunity to study how model-free adaptive optimal control algorithms apply to TCL. The recent advances in literature in the field of model-free adaptive optimal algorithm is recalled hereafter.

Adaptive optimal control originates from reinforcement learning [13]. The work [14] develops a Policy Iteration (PI) algorithm for an LTI systems. It solves the regulation problem but only using partial knowledge of the system dynamics, and it requires only the knowledge of input matrix; [15] uses the same idea to solve a tracking problem instead of a regulation problem; [16] develops a PI algorithm for a bi-linear system. The work [17] develops a Value Iteration (VI) algorithm for LTI systems; the main advantage is that the algorithm gets rid of the assumption on the partial knowledge of the system; [18] works on the same regulation problem as of [17], but is based on stochastic approximation to develops a VI algorithm. Other related works are [19], [20] and [21].

In this work we look into the application of the above-mentioned adaptive model-free optimal control strategies for TCL. The main contribution is to highlights advantages and disadvantages of the models in literature, and which algorithms can actually be implemented for this relevant problem. The rest of the work is organized as follows: models for homogeneous and heterogeneous populations of TCLs are explained in Sect. III. The control problem is formulated in Sect. II; an output feedback algorithm for the homogeneous model is studied and applied in Sect/ IV, whereas a non-linear adaptive optimal controller is applied Sect. V. Conclusions and discussions are provided in Sect. VI.

Notation: Throughout this article, $\mathbb{R}$ denotes the sets of real numbers. Vertical bars $|.|$ represent the euclidean norm for vectors, or the induced matrix norm for matrices. $\otimes$ indicate Kronecker product. I_n stands an identity matrix of size n. $\nabla f(x)$ represents the gradient of the function $f(x)$. $\underline{a}$ represents a column vector with individual elements equal to a and $\sigma(A)$ represents the spectral radius of matrix A.

II. TCL Population Model

In this section, two models are recalled, one for homogeneous population and one for heterogeneous population of TCLs, respectively. Advantages and limitations are also discussed in the context of adaptive optimal control. For simplicity, we will focus on a cooling scenario.

This work was partially supported by the Fundamental Research Funds for the Central Universities under Grant 4007019109 (RECON-STRUCT), and by the special guiding fund for double first-class under Grant 4007019201. (Corresponding author: Simone Baldi).

S. Coimbatore Anand was with Delft Center for Systems and Control, Delft University of Technology (TU Delft), Delft, Netherlands, and is now with Division of Signals and System, Uppsala University, Sweden.

S. Baldi is with the School of Mathematics, Southeast University, Nanjing 210096, China, and guest with the Delft Center for Systems and Control, TU Delft, 2628 Delft, Netherlands (e-mail: s.baldi@tudelft.nl)

A. Homogeneous Population of TCLs

The thermostatic behaviour of temperature $\theta(t)$ can be described

$$\dot{\theta} = \begin{cases} -\frac{1}{CR}(\theta - \theta_{amb} + PR), & \text{ON State.} \\ -\frac{1}{CR}(\theta - \theta_{amb}), & \text{OFF State.} \end{cases}$$

where TCL switches from OFF to ON State if $\theta > \theta_s + \Delta/2$ and TCL switches from ON to OFF State if $\theta < \theta_s - \Delta/2$. Here C is the thermal capacitance - kWh/°C, R is the thermal resistance - °C/kW, θ_{amb} is the ambient temperature - °C, θ_s is the temperature set-point - °C, Δ is the temperature dead-band - °C, P is the power drawn - kW, δ is the step change applied to the input - °C.

For a homogeneous ensemble of N TCLs, the TCLs have the same parameters (C, R, ...). Let N_c and N_h represent the loads in the ON and OFF states respectively. The probability density functions can be approximated respectively as

$$f_1(\theta) = \frac{CR}{(T_c + T_h)(PR + \theta_{amb} - \theta)} \tag{1}$$

$$f_0(\theta) = \frac{CR}{(T_c + T_h)(\theta_{amb} - \theta)} \tag{2}$$

When a step change is made in the set-point of the TCL, the dead-band changes. The change in average power consumption caused by the set-point step change is calculated by integrating the product of the probability density functions (1) and (2).
***Assumption 1*: [7]**

$$\Delta << (\theta_s - \theta_{amb} + PR)$$
$$\Delta << (\theta_{amb} - \theta_s), \quad \delta << \Delta,$$

Under ***Assumption 1***, the linear transfer function relating the step change in set point δ and the average power consumption P_{tot} can be approximated as

$$T(s) = \frac{P_{tot}(s)}{\delta/s} = -\frac{N}{\eta R} + \frac{A_\Delta \omega s}{s^2 + \omega^2}. \tag{3}$$

where

$$A_\Delta = \frac{5\sqrt{15}C(\theta_{amb} - \theta_+)(PR - \theta_{amb} + \theta_+)}{\eta(P^2R^2 + 3PR(\theta_{amb} - \theta_+) - 3(\theta_{amb} - \theta_+)^2)^{3/2}} \frac{(3PR - \theta_{amb} + \theta_+)N}{T_{c0} + T_{h0}},$$

$$\omega = \frac{2\sqrt{15}C(\theta_{amb} - \theta_+)(PR - \theta_{amb} + \theta_+)}{CR\Delta\sqrt{(P^2R^2 + 3PR(\theta_{amb} - \theta_+) - 3(\theta_{amb} - \theta_+)^2)}},$$

The corresponding state space representation of the transfer function (3) is

$$\dot{x} = \underbrace{\begin{bmatrix} -2\sigma & -\omega \\ \frac{\sigma^2+\omega^2}{\omega} & 0 \end{bmatrix}}_{\mathbf{A}} x + \underbrace{\begin{bmatrix} \omega A_\Delta \\ 0 \end{bmatrix}}_{\mathbf{B}} u \tag{4}$$

$$y = \underbrace{\begin{bmatrix} -1 & 0 \end{bmatrix}}_{\mathbf{C}} x + \underbrace{-\frac{N}{\eta R}}_{\mathbf{D}} u \tag{5}$$

An open problem in literature is that a physical interpretation of the states for the system (4)-(5) cannot be found. Hence, we conclude that state-feedback approaches are not appropriate for (4)-(5), and alternative can be proposed in the following ways. *(i)* Adopt an OutPut FeedBack (OPFB) algorithm since the output $y(t)$ for the system is measurable. *(ii)* Adopt a different system representation where the states are measurable, like the one presented below.

B. Heterogeneous Population of TCLs

Consider a heterogeneous population of N TCLs. The probability of TCLs going from θ_{start} to θ_{end} is $P(\theta_{end}|\theta_{start}) = P(a_i)$ where $a_i = \frac{\theta_a - \theta_{end} - m_t\theta_g}{\theta_a - \theta_{start} - m_t\theta_g}$. Similarly, the probability of the TCL going from $\theta_m < \theta_{start} < \theta_{m+1}$ to $\theta_n < \theta_{end} < \theta_{n+1}$ is

$$P(\theta_n < \theta_{end} < \theta_{n+1} | \theta_m < \theta_{start} < \theta_{m+1}) = \int_{\theta_m}^{\theta_{m+1}} \int_{a_1}^{a_2} p(a)\, da\, d\theta_{start} \tag{6}$$

where $a_1 = \frac{\theta_a - \theta_1 - m_t\theta_g}{\theta_a - \theta_{start} - m_t\theta_g}$ $a_2 = \frac{\theta_a - \theta_2 - m_t\theta_g}{\theta_a - \theta_{start} - m_t\theta_g}$. Here, $\theta_1 = \theta_{n/n+1}$ and $\theta_2 = \theta_{n+1/n}$ when the TCL is traversing from low/high to high/low temperature, $\theta_g = RP$ is the ON temperate gain of the TCL and m is a boolean variable $1/0$ defining the ON/OFF state of the TCL respectively.

Since this probability depends on the temperature gains, the parameter heterogeneity is inbuilt in the parameters R and C. Let us divide the temperature dead band of the TCL into N_2 state bins When (6) is evaluated for every starting and ending bins, the system matrix $A \in \mathbb{R}^{2N_2 \times 2N_2}$ can be analytically derived (not reported for lack of space, cf. [22] for details), or identified from data. Hence in this model,

- The state $x \in \mathbb{R}^{2N_2}$ is measurable and represents the number of TCL in each temperature bins. Therefore, state-feedback approaches can be adopted for this model.
- The control input $u \in \mathbb{R}^{N_2}$ represents the number of TCLs to be switched in a specific bin from ON/OFF to OFF/ON respectively. The matrix B can be hence constructed as in (7).
- The output y represents the aggregate power of TCLs. The matrix C can be hence constructed as in (7).

$$B = \begin{bmatrix} -1 & \dots & 0 \\ \vdots & \dots & \vdots \\ \vdots & \dots & -1 \\ 0 & \dots & 1 \\ \vdots & \dots & \vdots \\ 1 & \dots & 0 \end{bmatrix} \quad C^T = P \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{7}$$

III. Problem formulation

In view of the previously presented models, let the system dynamics of a population of TCLs be represented by the following LTI state-space form

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) \tag{8}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$.

***Assumption 2*:** The system (8) is controllable and observable. Let y_d represent a constant reference trajectory. The objective is to find a stabilizing control input $u(t)$ such that

$$\lim_{t\to\infty} y(t) - y_d \to 0$$

and minimizing the cost/value function

$$V(t) = \int_t^\infty e^{-\gamma(\tau-t)}\Big(x(t)^T Q x(t) + u(t)^T R u(t)\Big)dt. \quad (9)$$

In case of output feedback, the first term of (9) becomes $y(t)^T Q y(t)$. Without loss of generality, the cost function can be taken of the form $V(t) \triangleq x(t)^T P x(t)$. The next section aims at solving the above mentioned problem for a homogeneous population of TCL.

IV. Output feedback for homogeneous population of TCLs

To keep the presentation self-contained, in the following section, we will recall the OPFB algorithm from [20]. Let us assume that a state feedback control of the form $u = Kx(t)$ is applied to the system (4)-(5). The solution $x(t)$ becomes

$$x(t) = e^{(t-t_0)(A+BK)}x(t_0) \quad (10)$$

The solution $y(t)$ in terms of $x(t)$ can be written as

$$y(t - i\Delta t) = Ce^{-i\Delta t(A+BK)}x(t)$$

Suppose that there are N_1 output measurements available, using the above representation, a stacked fictious state $\bar{y}_t$ can be constructed as

$$\underbrace{\begin{bmatrix} y(t) \\ y(t-\Delta t) \\ \vdots \\ y(t-(N_1-1)\Delta t)\end{bmatrix}}_{\bar{y}_t} = \underbrace{\begin{bmatrix} C \\ Ce^{-\Delta t(A+BK)} \\ \vdots \\ Ce^{-(N-1)\Delta t(A+BK)}\end{bmatrix}}_{G} x(t)$$

$$\implies \bar{y}_t = Gx(t). \quad (11)$$

The idea here is to learn the value function $V(t)$ in term of the output measurements $\bar{y}_t$. Using (11), the quadratic value function $V(t)$ whose solution is to be found can be rewritten as

$$V(t) = x(t)^T P x(t) = \bar{y}_t^T G_{N_1}^T P G_{N_1} \bar{y}_t \quad (12)$$

where $G_{N_1} = (G^T G)^{-1} G^T$. Define $\bar{P} = G_{N_1}^T P G_{N_1}$. Using (10)-(12), the bellman equation equivalent of (9) becomes

$$e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T \bar{P}\bar{y}_{t+\Delta t} - \bar{y}_t^T \bar{P}\bar{y}_t = -\int_t^{t+\Delta t} e^{-\gamma(\tau-t)}\bar{y}_t^T \bar{Q}_i \bar{y}_t d\tau - 2\int_t^{t+\Delta t} e^{-\gamma(\tau-t)} w^T R \bar{K}^{i+1}\bar{y}_t d\tau \quad (13)$$

where $\bar{Q} = \begin{bmatrix}\underline{1} & 0 & 0 & 0\end{bmatrix}^T Q \begin{bmatrix}\underline{1} & 0 & 0 & 0\end{bmatrix}$ and w is the probing noise. This equation does not require the system state measurements and results in Algorithm 1.

Algorithm 1: VI algorithm for OPFB

Result: Riccati solution $\bar{P}, \bar{K}$

1 **Input**: An initial stabilizing control policy u^0

2 **Initialization**: Set $i \leftarrow 0$ and $t \leftarrow 0$

3 **Online data collection**: Apply the control policy $u = u^0 + e$ (where e is a probing/exploration noise) and collect the system output and input information.

4 **Policy evaluation**: Solve for $\bar{P}_i$ and $\bar{K}_i$ from (13)

5 **Stopping criterion**: Update $i \leftarrow i+1$ and $t \leftarrow t + \Delta t$, and go to **Step 3**, until

$$||\bar{P}_i - \bar{P}_{i-1}|| \leq \epsilon$$

where $\epsilon > 0$ is sufficiently small predefined threshold.

6 **Actual control policy improvement**: Terminate the exploration noise e and $u = u_0$ as the control input. Apply the control policy $u = \bar{K}_i \bar{y}_t$.

N_1	5	6	7
K^T	$\begin{bmatrix}0.6902\\0.1111\\-0.4557\\0.0644\\-0.4114\end{bmatrix}$	$\begin{bmatrix}0.4406\\0.2315\\-0.0301\\-0.2762\\0.0416\\-0.4091\end{bmatrix}$	$\begin{bmatrix}0.3066\\0.2163\\0.0963\\-0.0538\\-0.1907\\0.0297\\-0.4066\end{bmatrix}$
Computation time [s]	113	111	111

Table I: Performance comparison for varying N

A. Results and discussion

Consider a homogeneous population of TCLs described in (4) - (5). The system matrices are obtained from [23]. As stated before, since the states of this system are immeasurable, Algorithm 1 is applied with $\gamma = 0.1, Q = 0.1$ and $\Delta t = 0.1$. The TCLs are required to track as step change from 15.5 to 20 kW. The number of stored data in the history is 3 i.e: $\bar{y}_t = [y(t)\ y(t-\Delta t)\ y(t-2\Delta t)\ r(t)]$. A probing noise of the form $\sum_{\omega=1}^{100} sin(\omega t)$ is applied. Since $\sigma(A) < 0$, the algorithm is initialized with $\bar{K}_0 = 0$. The tracking performance with different input costs (R) is shown in Fig. 1. As the input cost increases, the input magnitude decreases, but the settling time increases as well. The figure highlights that the results corresponding to $R = 1$ and $R = 3$ are quite realistic. We then fix $R = 1$ and $Q = 0.1$, and further study the effect of increasing the memory variables N_1: the results are reported in Table I: the table reports that increasing N_1 may not necessarily lead to higher computational time. In the next section, we study the model free adaptive control strategies for the heterogenous model described in (II-B).

V. Nonlinear state feedback for heterogeneous population of TCLs

Although the heterogeneous model (II-B) is linear, state-feedback control algorithms cannot be applied, since the

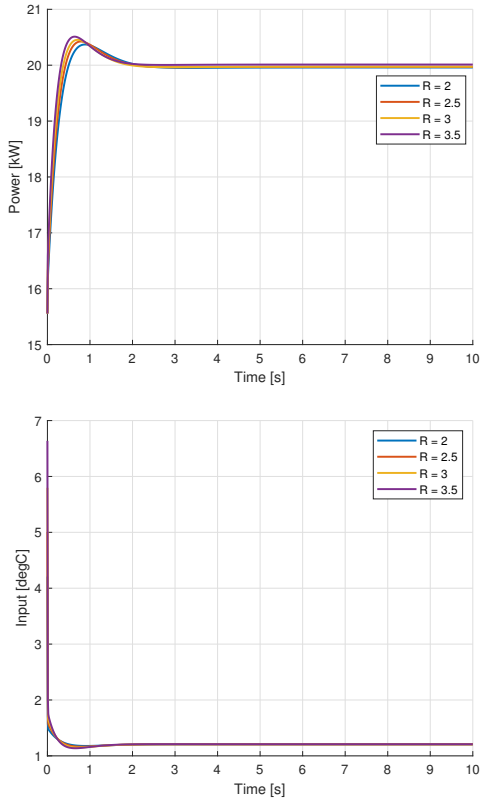

Figure 1: Output and input trajectories with varying costs

system is not controllable ($\sum_i x_i$ is constant). This issue can be solved by defining a desired state x_{set} corresponding to the desired output. In other words, the set-point power is represented in terms of distribution of TCL across state bins. Because some heterogeneous TCLs models in literature are bilinear, let us address a nonlinear controller.

Consider a non-linear system of the form

$$\dot{x} = f(x) + g(x)u, \quad x(0) = x_0 \tag{14}$$

$$J(x,u) = \int_0^\infty \underbrace{q(x(t)) + u(t)^T R(x)u(t)}_{r(x(t),u(t))} dt, \tag{15}$$

where $f(\cdot)$ and $g(\cdot)$ are Lipschitz continuous functions. The objective is to find a control input $u(\cdot)$ that minimizes the cost function. The system (14) can be related to the model developed in Section II-B as $f(x)$ representing $A(x - x_{set})$ and $g(x)$ representing B. The system (14) can be rewritten in the form

$$\dot{x} = f(x) + g(x)u_i(x) + g(x)v_i$$

where $v_i = u_0 - u_i + e$, u_0 is the initial control input e is the exploration noise.

***Assumption 3*:** The system (14) is Input to State Stable (ISS) when e is considered as input.

The solution of the value function (15), under ***Assumption 3*,** along the trajectory of (V) and integrating on the interval $[t, t + \Delta t]$ yields

$$\nabla V(x(\cdot)) = -\int_t^{t+\Delta t} [q(x) + u_i^T R(x)u_i + 2u_{i+1}^T R(x)v_i]d\tau \tag{16}$$

By approximation theory, $V(\cdot)$ and $u(\cdot)$ can be approximated by basis function

$$\hat{V}_i(x) = \sum_{j=1}^{N_1} \hat{c}_{i,j}\phi_j(x) \quad \hat{u}_{i+1}(x) = \sum_{j=1}^{N_2} \hat{w}_{i,j}\psi_j(x)$$

where $\hat{c}$ and $\hat{w}$ are weights to be determined. Hence (16) becomes [24]

$$\begin{aligned}\sum_{j=1}^{N_1} \hat{c}_{i,j}[\phi_j(x(t_{k+1})) - \phi_j(x(t_k))] \\ = -\int_{t_k}^{t_{k+1}} [q(x) + \hat{u}_i^T R(x)\hat{u}_i dt \\ - \int_{t_k}^{t_{k+1}} 2\sum_{j=1}^{N_2} \hat{w}_{i,j}\psi_j^T(x)R(x)\hat{v}_i dt + e_{i,k}\end{aligned} \tag{17}$$

The solution $\hat{c}$ and $\hat{w}$ can be found by minimizing $e_{i.k}$ in a least squares sense. The equation (17) does not depend on the system dynamics but only on the state and input measurements. This brings us to the online adaptive Algorithm-2.

Algorithm 2: VI for non-linear non-affine systems

Result: Weights of the basis functions $\hat{w}, \hat{c}$

1 **Input**: A initial stabilizing control policy u_0

2 **Initialization**: Determine the set $\Omega \in \mathbb{R}^n$ for approximating the states $x(t)$. Set $i \leftarrow 0$

3 **Online data collection**: Apply the control policy $u = u^0 + e$ and collect the system state and input information.

4 **Policy evaluation and improvement**: Solve for $\hat{w}$ and $\hat{c}$ from (17).

5 **Stopping criterion**: Let $i \leftarrow i + 1$, and go to **Step 3**, until

$$\sum_{j=1}^{N_1} |\hat{c}_{i,j} - \hat{c}_{i-1,j}|^2 \leq \epsilon$$

where $\epsilon > 0$ is sufficiently small predefined threshold.

6 **Actual control policy improvement**: Terminate the exploration noise e and $u = u_0$ as the control input. Once $x(t) \in \hat{\Omega}_i$, apply the control policy $u = \hat{u}_{i+1}$.

A. Results and Discussion

With the TCl parameters from [23], a linear model as described in Section-II-B is developed. Algorithm 2 is applied to this system. The number of state bins considered is 4, number of TCLs considered is 40 and a probing noise of

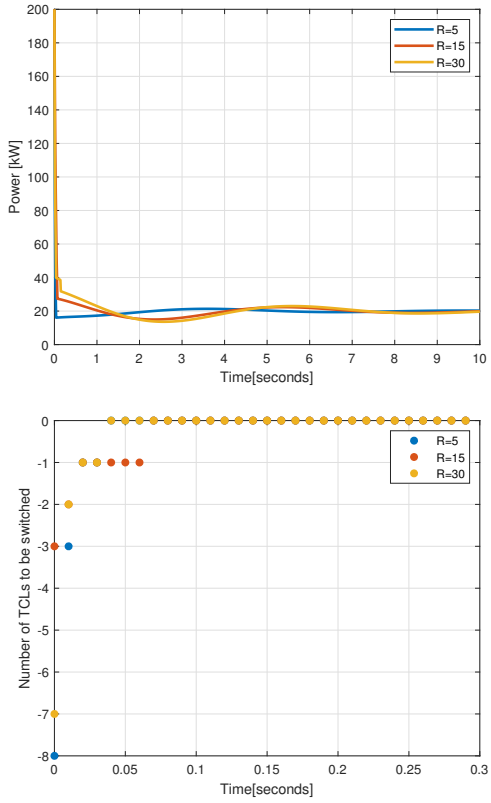

Figure 2: Output and input trajectories with varying costs

Number of state bins N_2	4	6	8	10
Computational time [ms]	8.50	33.78	43.28	80.07

Table II: Computational complexity for varying state bins

the form $\sum_{\omega=1}^{100} sin(\omega t)$ is applied. Since the system has the property $\sigma(A) < 0$, the algorithm is initialized with $u_0 = 0$. The tracking performance with increasing input cost (R) is shown in Fig. 2.

In general, the accuracy of the system representation increases with increasing number of state bins. Hence a study is made with increasing state bins, and the resulting convergence time is reported in Table II. As expected, the computational time increases with the number of state bins.

VI. Conclusions and Future Work

We have studied the problem of the adaptive optimal control problem for TCLs. Both homogeneous and heterogeneous populations of TCLs have been considered and an appropriate problem statement is formulated. The interest in adaptive optimal control was motivated by the difficulty in getting accurate system parameters. Overall this works proves the feasibility of adaptive optimal control for TCLs, although some open problems are still open for future research: a) obtain a physical interpretation for the state of the homogeneous population of TLCs; b) reduce the requirements for persistency of excitation in the learning phase [25], [26]; c) address the inevitable presence of external disturbances and explicit constraints on input.

Another open problem in literature is the following: in case of TCL systems, the input matrix B is easy to be known (upon prefiltering of the control input as explained in [27]), which means that PI algorithms can be sought, since PI converges faster than VI. Hence a PI algorithm for a partially unknown system can sought for OPFB control. Let us consider the state space representation (8). The conventional solution for (9), with complete system knowledge, can be found by solving

$$(A-0.5\gamma I)^T P + P(A-0.5\gamma I) - PBR^{-1}B^T P = -C^T QC \tag{18}$$

The solution P can be found for a partially unknown system (only matrix B is known) using the state measurements online by solving recursively [15]:

$$x(t)^T P^i x(t) - e^{-\gamma\Delta t} x(t+\Delta t)^T P^i x(t+\Delta t) = \frac{1}{2}\int_t^{t+\Delta t} e^{-\gamma(\tau-t)}\Big[x(t)^T C^T QCx(t) + u_i^T Ru_i\Big]d\tau.$$

Using (10) and (12) in the above equation results in

$$\bar{y}_t^T \bar{P}_i \bar{y}_t - e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T \bar{P}_i \bar{y}_{t+\Delta t} = \frac{1}{2}\int_t^{t+\Delta t} e^{-\gamma(\tau-t)}\Big[\bar{y}_t^T Q\bar{y}_t + u_i^T Ru_i\Big]d\tau \tag{19}$$

which is independent of the states and the system matrices. Hence, (19) can be used to propose **Algorithm 3**.

Algorithm 3: PI algorithm for OPFB

Result: Riccati solution $\bar{P}$

1 **Input**: A initial stabilizing control policy

2 **Initialization**: Set $i \leftarrow 0$ and $t \leftarrow 0$

3 **Online data collection**: Apply the control policy $u = u^i + e$ (where e is a probing/exploration noise and collect the system output and input information.

4 **Policy evaluation**: Solve for $\bar{P}_i$ from (19)

5 **Policy improvement**: Apply the control policy $u^i = -R^{-1}B^T G_2 \bar{P}_i \bar{y}_t$

6 **Stopping criterion**: Let $i \leftarrow i+1$ and $t \leftarrow t+\Delta t$, and go to **Step 3**, until

$$||\bar{P}_i - \bar{P}_{i-1}|| \leq \epsilon$$

where $\epsilon > 0$ is sufficiently small predefined threshold.

Lemma VI.1. *The equation* (19) *converge to a sub-optimal positive definite solution of* (18).

Proof. Dividing (19) by Δt and taling a limit results in

$$\lim_{\Delta t\to 0} \frac{\bar{y}_t^T \bar{P}\bar{y}_t - e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T \bar{P}\bar{y}_{t+\Delta t}}{\Delta t} = \lim_{\Delta t\to 0} \frac{\int_t^{t+\Delta t} e^{-\gamma(\tau-t)}\Big[\bar{y}_t^T Q\bar{y}_t + u^T Ru\Big]d\tau}{\Delta t}$$

$$\lim_{\Delta t\to 0}\frac{\int_t^{t+\Delta t} e^{-\gamma(\tau-t)}\Big[\bar{y}_t^T Q\bar{y}_t + u^T Ru\Big]d\tau}{\Delta t} = \bar{y}_t^T Q\bar{y}_t + u^T Ru = x(t)^T C^T QCx(t) + u^T Ru$$

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{\bar{y}_t^T\bar{P}\bar{y}_t - e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T\bar{P}\bar{y}_{t+\Delta t}}{\Delta t}\\ &= \lim_{\Delta t\to 0}\Big(-\gamma e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T\bar{P}\bar{y}_{t+\Delta t}\\ &\qquad + e^{-\gamma\Delta t}\dot{\bar{y}}_{t+\Delta t}^T\bar{P}\bar{y}_{t+\Delta t} + e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T\bar{P}\dot{\bar{y}}_{t+\Delta t}\Big)\\ &= -\gamma\bar{y}_t^T\bar{P}\bar{y}_t + \dot{\bar{y}}_t^T\bar{P}\bar{y}_t + \bar{y}_t^T\bar{P}\dot{\bar{y}}_t\end{aligned}$$

Differentiating (11) results in

$$\dot{\bar{y}}_t = G\dot{x}(t) = GAx(t) + GBu(t)$$

Using this in the previous equation gives

$$\lim_{\Delta t\to 0}\frac{\bar{y}_t^T\bar{P}\bar{y}_t - e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T\bar{P}\bar{y}_{t+\Delta t}}{\Delta t} = x(t)^T(A^TP + PA - \gamma P)x(t)$$

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{\int_t^{t+\Delta t} e^{-\gamma(\tau-t)}\Big[\bar{y}_t^T Q\bar{y}_t + u^T Ru\Big]d\tau}{\Delta t}\\ &\qquad + \lim_{\Delta t\to 0}\frac{\bar{y}_t^T\bar{P}\bar{y}_t - e^{-\gamma\Delta t}\bar{y}_{t+\Delta t}^T\bar{P}\bar{y}_{t+\Delta t}}{\Delta t}\\ &= x(t)^T(A^TP + PA - \gamma P + C^TQC)x(t) + \hat{x}PBR^{-1}B^TP\hat{x}\end{aligned}$$

Now, let G_2 be a filter with the same dimension of G, then $\hat{x}\to x$ as $G_2\to G$. □

Although the main advantage of this idea is to work in a PI setting, it has the drawback of requiring the convergence $G_2\to G$. This can be in principle guaranteed by the probing/exploration noise. Nevertheless an open problem is how to avoid such an extra filter.

REFERENCES

[1] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389–1400, 2009.

[2] S. Koch, M. Zima, and G. Andersson, "Potentials and applications of coordinated groups of thermal household appliances for power system control purposes," in *2009 IEEE PES/IAS Conf. on Sustainable Alternative Energy (SAE)*, pp. 1–8, IEEE, 2009.

[3] S. Baldi, A. Karagevrekis, I. T. Michailidis, and E. B. Kosmatopoulos, "Joint energy demand and thermal comfort optimization in photovoltaic-equipped interconnected microgrids," *Energy Conversion and Management*, vol. 101, pp. 352 – 363, 2015.

[4] C. D. Korkas, S. Baldi, and E. B. Kosmatopoulos, "9 - grid-connected microgrids: Demand management via distributed control and human-in-the-loop optimization," in *Advances in Renewable Energies and Power Technologies* (I. Yahyaoui, ed.), pp. 315 – 344, Elsevier, 2018.

[5] T. Ericson, "Direct load control of residential water heaters," *Energy Policy*, vol. 37, no. 9, pp. 3502–3512, 2009.

[6] J. Taneja, D. Culler, and P. Dutta, "Towards cooperative grids: Sensor/actuator networks for renewables integration," in *2010 First IEEE Intl. Conf. on Smart Grid Communications*, pp. 531–536, IEEE, 2010.

[7] S. Kundu, N. Sinitsyn, S. Backhaus, and I. Hiskens, "Modeling and control of thermostatically controlled loads," *arXiv preprint arXiv:1101.2157*, 2011.

[8] S. Bashash and H. K. Fathy, "Modeling and control insights into demand-side energy management through setpoint control of thermostatic loads," in *American Control Conf. (ACC), 2011*, pp. 4546–4553, IEEE, 2011.

[9] S. Koch, J. L. Mathieu, and D. S. Callaway, "Modeling and control of aggregated heterogeneous thermostatically controlled loads for ancillary services," in *Proc. PSCC*, pp. 1–7, Citeseer, 2011.

[10] M. Liu and Y. Shi, "Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services," *IEEE trans. on power systems*, vol. 31, no. 3, pp. 1963–1971, 2016.

[11] A. Ghaffari, S. Moura, and M. Krstić, "Modeling, control, and stability analysis of heterogeneous thermostatically controlled load populations using partial differential equations," *Journal of Dynamic Systems, Measurement, and Control*, vol. 137, no. 10, p. 101009, 2015.

[12] M. Chertkov and V. Chernyak, "Ensemble of thermostatically controlled loads: statistical physics approach," *Scientific Reports*, vol. 7, no. 1, p. 8673, 2017.

[13] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE circuits and systems magazine*, vol. 9, no. 3, 2009.

[14] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, 2009.

[15] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Trans. on Autom. Control*, vol. 59, no. 11, pp. 3051–3056, 2014.

[16] B. Luo and H.-N. Wu, "Online adaptive optimal control for bilinear systems," in *2012 American Control Conf. (ACC)*, pp. 5507–5512, IEEE, 2012.

[17] Y. Jiang and Z. P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[18] T. Bian and Z. P. Jiang, "Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design," *Automatica*, vol. 71, pp. 348–360, 2016.

[19] W. Gao and Z.-P. Jiang, "Adaptive dynamic programming and adaptive optimal output regulation of linear systems," *IEEE Trans. on Automatic Control*, vol. 61, no. 12, pp. 4164–4169, 2016.

[20] H. Modares, F. L. Lewis, and Z. P. Jiang, "Optimal Output-Feedback Control of Unknown Continuous-Time Linear Systems Using Off-policy Reinforcement Learning," *IEEE Trans. on Cybernetics*, vol. 46, no. 11, pp. 2401–2410, 2016.

[21] I. Michailidis, S. Baldi, E. B. Kosmatopoulos, and P. A. Ioannou, "Adaptive optimal control for large-scale nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5567–5577, 2017.

[22] S. Coimbatore Anand, *Optimal tracking strategies for uncertain ensembles of thermostatically controlled loads*. MSc thesis, TU Delft, 2019.

[23] L. Chang, X. Wang, and M. Mao, "Forecast of schedulable capacity for thermostatically controlled loads with big data analysis," in *Power Electronics for Distributed Generation Systems (PEDG), 2017 IEEE 8th Intl. Symposium on*, pp. 1–6, IEEE, 2017.

[24] Y. Jiang and Z.-P. Jiang, *Robust Adaptive Dynamic Programming*. John Wiley & Sons, 2017.

[25] S. K. Jha, S. B. Roy, and S. Bhasin, "Direct adaptive optimal control for uncertain continuous-time lti systems without persistence of excitation," *IEEE Trans. on Circuits and Systems II: Express Briefs*, vol. 65, no. 12, pp. 1993–1997, 2018.

[26] N. Cho, H. Shin, Y. Kim, and A. Tsourdos, "Composite model reference adaptive control with parameter convergence under finite excitation," *IEEE Trans. on Automatic Control*, vol. 63, no. 3, pp. 811–818, 2018.

[27] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 32, no. 2, pp. 140–153, 2002.