2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Balancing the Power Grid with Cheap Assets

Sean Meyn*, Fan Lu, and Joel Mathias

***Abstract*— We have all heard that there is growing need to secure resources to obtain supply-demand balance in a power grid facing increasing volatility from renewable sources of energy. There are mandates for utility scale battery systems in regions all over the world, and there is a growing science of "demand dispatch" to obtain *virtual* energy storage from flexible electric loads such as water heaters, air conditioning, and pumps for irrigation.**

The question addressed in this tutorial is how to manage a large number of assets for balancing the grid. The focus is on variants of the economic dispatch problem, which may be regarded as the "feed-forward" component in an overall control architecture.

1) The resource allocation problem is identical to a finite horizon optimal control problem with degenerate cost—so called "cheap control". This implies a form of state space collapse, whose form is identified: the marginal cost for each load class evolves in a two-dimensional subspace, spanned by a scalar co-state process and its derivative.

2) The implication to distributed control is remarkable. Once the co-state process is synthesized, this common signal may be broadcast to each asset for optimal control. However, the optimal solution is extremely fragile, in a sense made clear through results from numerical studies.

3) Several remedies are proposed to address fragility. One is described through "robust training" in a particular Q-learning architecture (one approach to reinforcement learning). In numerical studies it is found that specialized training leads to more robust control solutions.

I. Controlling the Power Grid

This tutorial concerns control of a balancing area, with consideration of three classes of agents: 1. consumers of electricity, both residential and commercial, 2. generators of various types, 3. a balancing authority (BA) that has some authority over all agents, whose mandate is to ensure reliability within its territory. In some regions of the US such as Florida, a utility company may serve roles 2 and 3. A resource aggregator may serve to facilitate interaction between the BA and consumers.

In the past the BA has been concerned largely with reliability and cost. Our goal, and increasingly the goal of policy makers, is to address the need for balancing services in a power grid with large amounts of energy from the wind and sun, ultimately resulting in lower emissions.

*S. Meyn and F. Lu are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA (e-mail: fan.lu@ufl.edu; meyn@ece.ufl.edu).

J. Mathias is with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, 85281, USA (e-mail: Joel.Mathias@asu.edu).

SM acknowledges support from ARO award W911NF2010055, National Science Foundation awards 2122313 and EPCN 1935389, and from an Inria International Chair, Paris, France. Portions of the results are based on joint research with Karan Kalsi, Robert Moye and Joseph Warrington.

An example of a BA is CAISO (California Independent System Operator). Fig. 1 provides an illustration of the challenges they face. The demand at 2pm was just under 20 GW, while the net-demand (demand minus generation from renewables; also called net-load) was less than 5 GW. This good news is offset by the tremendous ramps in net-demand observed between 5pm and 8pm, coinciding with the setting sun. Generators must be ready in advance of this surge, and ramp up their production to ensure supply meets demand at all times.

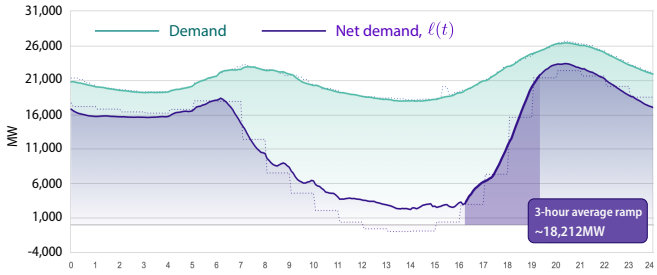

Fig. 1: Demand and net-demand at CAISO on Monday, May 8, 2023, as shown on their website the following day. The net-demand shows some volatility on time scales of minutes. The massive afternoon ramp is the largest challenge to CAISO on this day.

It is commonly proposed that net-demand can be flattened with a big enough battery. On this day, the required size for entirely flat net-demand (in particular, compensating for the 18GW ramp) would be approximately 10 GW in terms of power, and 40 GWh in terms of energy. The largest battery in the world today is about 0.35 GW / 1.5 GWh, taking up 100 acres of public land[1].

The reader is likely aware that flexible loads may be aggregated to provide services similar to a large battery system, without impacting consumer quality of service. High quality service requires a sensible control strategy: The term *demand dispatch* was introduced in [3] to convey control of flexible loads for the creation of *virtual energy storage* (VES). The priority-stack approach of [10] is ideal for fully centralized control, in which the aggregator can observe in real time the state of each load under its control. A distributed control solution for building HVAC systems is introduced in [9]. Approaches for on-off loads are typically based on a mean field model. Examples include feedback linearization [21], [25], requiring estimates of the histogram of states; load-level control techniques to enable control of the aggregate through scalar broadcast from the BA or aggregator (see [6],

[1]https://www.blm.gov/press-release/blm-announces-completion-crimson-energy-storage-project

[16], [17] for history), and feedforward control techniques [7], [2], [4].

A large collection of flexible loads in a given class (e.g., one million residential water heaters) will be called a distributed energy resource (DER). The same term will be used for a battery system that is engaged with the BA.

Each approach to demand dispatch is designed to provide grid balancing and regulation service while also imposing constraints on consumer-side quality of service (QoS). Consequently, while any DER is expensive to acquire in terms of installation cost, the operating costs are essentially zero. This has implications to control, and this *zero marginal cost* aspect of grid services also has significant implications to economics. In particular, it is demonstrated in [1], [12] that the deferrable nature of the loads considered here and in prior research implies that aggregate power consumption does not vary smoothly with small changes in price.

Contributions and organization. The remainder of this tutorial is organized in two sections, followed by the conclusions.

Section II contains a full description of the resource allocation problem considered. It is a formulated as a finite horizon optimal control problem, for which the cost is degenerate—so called "cheap control". This implies a form of state space collapse, whose form is identified: the marginal cost for each load class evolves in a two-dimensional subspace, spanned by a scalar co-state process and its derivative.

The implication to distributed control is remarkable: the scalar co-state signal may be broadcast to each asset to achieve the optimal control solution.

However, the optimal solution is extremely fragile, in a sense made clear through results from numerical studies. Several remedies are proposed to address fragility in Section III. One is described through "robust training" in a particular Q-learning architecture (one approach to reinforcement learning). In numerical studies it is found that specialized training leads to more robust control solutions.

II. Blessings from Mismatched Needs

Let's consider the needs of each agent: 1. residential consumers want hot water, a cold refrigerator, and a comfortable home (the needs of commercial consumers vary widely); 2. each generation company wants to maximize profits; and as stated previously 3. the BA engages with all the agents to ensure a reliable grid.

The mismatch between residential consumers and the BA is massive. For example, if every standard water heater (not tankless) is turned off for one hour, a consumer will likely suffer no loss of QoS, because temperature changes very slowly without usage. One person taking a 30 minute shower may notice a change in temperature. On the other hand, if all of these loads are shut off simultaneously, this represents a massive shock to the grid in terms of a downward ramp in demand; perhaps one GW in the case of several million water heaters.

With thoughtful design, this mismatch is a great blessing. Power consumption can be varied wildly without imposing cost to the consumer. This is why virtual energy storage is a resource of enormous untapped potential.

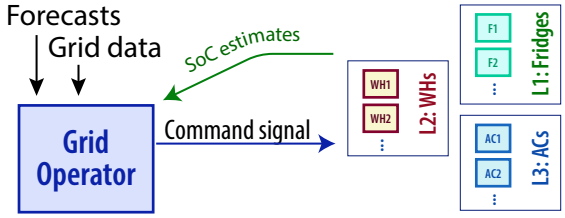

Fig. 2: Distributed control of distributed energy resources.

A. Control model

The control problem centers on the interaction between the three entities. The BA is regarded as the leading agent that optimizes generation and aggregate demand based on forecasts of nominal net-demand.

Fig. 2 shows an idealized control architecture, ignoring complications from engagements with resource aggregators. The BA must ensure that QoS constraints are satisfied for each load in its territory, and ensure that it will provide desired grid services, such as addressing the ramp shown in Fig. 1. In addition, generation is scheduled so as to minimize cost (including the costs associated with ramping).

Each of $M \geq 2$ DERs is modeled as a scalar linear state space model:

$$\tfrac{d}{dt} x_t^i = -\alpha_i x_t^i - z_t^i, \qquad 1 \leq i \leq M\,, \tag{1}$$

in which x_t^i is the *state of charge* (SoC) of the ith DER, and $-z_t^i$ is power deviation at time t. The coefficient $\alpha_i \geq 0$ models leakage for a battery system, and has a similar interpretation for a TCL (thermostatically controlled loads, such as a water heater). Justification of this model may be found in [11] for TCLs, and a similar construction was proposed as an approximation for pool cleaning (as well as irrigation) in [23].

The state space model for control will be the $2M$-dimensional process $x_t^a = \{x_t^{a\,i} := (x_t^i; z_t^i) : 1 \leq i \leq M\}$, with input $u_t = \frac{d}{dt} z_t^i$. We find that this state augmentation is a convenient way to incorporate the cost of ramping generation in the optimization problems surveyed here.

With forecast net-demand at time t denoted ℓ_t, the realized load is $\ell_t - z_t^\sigma$, in which the superscript denotes summation, $z_t^\sigma = \sum_i z_t^i$; it is interpreted as the *virtual discharge* rate of the aggregate when positive. With g denoting total generation from traditional sources (e.g., fossil-based, nuclear, and hydro power plants), balancing supply and demand requires $g_t = \ell_t - z_t^\sigma$ for all t.

The economic dispatch problem is posed in continuous time as the finite horizon optimal control problem, with time-horizon $[0, \mathcal{T}]$: with $(x_0, z_0) = x^a = (x, u) \in \mathbb{R}^{2M}$ given,

$J^\star(x^a)$ is the value of

$$\begin{aligned}
\underset{g,\gamma,x}{\text{minimize}} \quad & \int_0^\mathcal{T} \big[c_{\mathrm{g}}(g_t) + c_{\mathrm{d}}(\gamma_t) + c_{\mathrm{x}}(x_t)\big]\, dt && \text{(2a)}\\
\text{subject to} \quad & \ell_t = g_t + z_t^\sigma \quad , && \text{(2b)}\\
& \tfrac{d}{dt} g_t = \gamma_t \quad , && \text{(2c)}\\
& \tfrac{d}{dt} x_t^i = -\alpha_i x_t^i - z_t^i \quad , && \text{(2d)}\\
& \tfrac{d}{dt} z_t^i = u_t^i, \quad i \in \{1,\dots,M\}\,, && \text{(2e)}
\end{aligned}$$

in which (2b) is the supply-demand constraint, and (2c) is a notational convention. The cost functions are defined as follows: c_{g} is generation cost, c_{d} is the cost of ramping generation, and c_{x} is a penalty or barrier function designed to impose QoS bounds. We take

$$c_{\mathrm{x}}(x) = \sum_{i=1}^M c_i(x^i)\,, \quad x \in \mathbb{R}^M\,, \tag{3}$$

in which each c_i is strongly convex. Given desired capacity bounds $|x_t^i| \le C_i$ for each i, a typical choice for barrier function is $c_i(x^i) = -\delta \log(1 - [x_i/C_i]^2)$ with $\delta > 0$ a small scalar.

The zero marginal cost assumption for balancing services implies this optimization problem falls in the category of "cheap optimal control" [8], [24]. This is seen through elimination of variables g and γ via the equality constraints (2b) and (2e) to obtain the following alternative expression for the cost at time t (the integrand in (2a)):

$$c(x_t^a, u_t, t) := c_{\mathrm{g}}(\ell_t - z_t^\sigma) + c_{\mathrm{d}}(\tfrac{d}{dt}\ell_t - u_t^\sigma) + c_{\mathrm{x}}(x_t)$$

The cost is a function of u_t^σ, and hence cannot be coercive in the M-dimensional input u_t.

B. State space collapse

Subject to smoothness assumptions on the cost functions and the nominal net-demand ℓ, the following conclusions are obtained in [19], [20]:

1 $J^\star$ is convex in x^a and finite-valued. Moreover, there is a function $K^\star\colon \mathbb{R}^2 \to \mathbb{R}$ such that $J^\star(x,z) = K^\star(x^\sigma, z^\sigma)$ for each $x, z \in \mathbb{R}^M$.

2 The optimal SoC evolves on a two-dimensional manifold and can be computed based on a scalar dual variable $\lambda^\star$ and its derivative:

$$c_i'\big(x_t^{i\star}\big) = \alpha_i \lambda_t^\star - \tfrac{d}{dt}\lambda_t^\star\,. \tag{4}$$

with c_i' the derivative of c_i, i.e. *marginal cost*. Under strong convexity the inverse $F_i = (c_i')^{-1}$ exists, giving

$$x_t^{i\star} = F_i\big(\alpha_i \lambda_t^\star - \tfrac{d}{dt}\lambda_t^\star\big) \tag{5}$$

A similar conclusion is reached in [5], but in an entirely different context.

The scalar signal $\lambda^\star$ represents the Lagrange multiplier for the supply-demand constraint (2b), and satisfies $\lambda_\mathcal{T}^\star = 0$. Consequently, if $x^{i\star}$ is known for just one i, then the Lagrange multiplier can be obtained by solving a first order differential equation, and from this $x_j^\star$ can be recovered for each j.

Fig. 3 is based on numerical results from [19]. In this experiment, $\lambda_t^\star$ and $\frac{d}{dt}\lambda_t^\star$ are treated as independent variables. Given two optimal SoC trajectories these values can be computed, and then any other optimal SoC trajectories are obtained. In this case, the SoC trajectory of pool pumps is recovered using the trajectories of ACs and residential water heaters via (5) combined with

$$\begin{bmatrix} \lambda_t^\star \\ \frac{d}{dt}\lambda_t^\star \end{bmatrix} = \begin{bmatrix} \alpha_{ac} & -1 \\ \alpha_{rwh} & -1 \end{bmatrix}^{-1} \begin{bmatrix} c_i'(x_{ac}^\star(t)) \\ c_i'(x_{rwh}^\star(t)) \end{bmatrix}$$

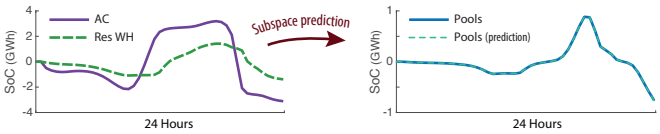

Fig. 3: SoC for pool pumps recovered using those of ACs and WHs.

Price discovery. One might assume that $\lambda^\star$ represents a *price*, since prices are typically interpreted as a Lagrange multiplier for a supply-demand constraint. It is argued in [19], [20] that this interpretation cannot be useful in practice. For one, there is no possibility of price discovery in this real-time setting. And remember, this is a finite horizon optimal control problem performed by the BA based on *forecast* net-demand. This information is not available to all of the other participants, and as forecasts change the BA may decide unilaterally to re-solve the optimization problem as part of an MPC (model predictive control) control architecture.

C. Solutions from reinforcement learning

In the recent work [14], [15] (see also the dissertation of Fan Lu [13]) the solution to the BA's optimization is approximated using techniques from reinforcement learning (RL).

An explanation of this approach requires a few more words on optimal control. First, recall that the *cost-to-go* is defined for each $T_0 \in [0, \mathcal{T})$ by

$$J^\star(x^a, T_0) := \min_{u_{T_0}^\mathcal{T}} \int_{T_0}^\mathcal{T} c(x_t^a, u_t, t)\, dt\,, \quad x_{T_0}^a = x^a$$

This is used in Bellman's *principle of optimality*, expressed as the family of fixed point equations: for $\tau \in [0, \mathcal{T})$ and with initial condition $x_0^a = x^a$,

$$J^\star(x^a) = \min_{u_0^\tau} \Big\{ \int_0^\tau c(x_t^a, u_t, t)\, dt + J^\star(x_\tau^a, \tau) \Big\} \tag{6}$$

This is one step in a proof of the Hamilton-Jacobi-Bellman (HJB) equation.

An important implication is described here—a cousin to the Minimum Principle. Consider any solution $\{(x_t^a, u_t) : t \ge 0\}$ to the state equations, and denote

$$Q^\star(x_t^a, u_t, t) = c(x_t^a, u_t, t) + \tfrac{d}{dt} J^\star(x_t^a, t)\,, \quad t \ge 0 \tag{7}$$

We have $Q^\star(x_t^a, u_t, t) \ge 0$ for all $0 \le t \le \mathcal{T}$, and this lower bound is achieved along an optimal solution. Following the

treatment of the infinite horizon problem considered in [22], take $\sigma > 0$ and consider

$$H^\star(x^a, u, t) := -\sigma J^\star(x^a, t) + Q^\star(x^a, u, t) \tag{8}$$

If we manage to compute $Q^\star$ or $H^\star$, then the optimal input is obtained via state feedback $u_t = \phi^\star(x^a_t, t)$ with

$$\begin{aligned}\phi^\star(x^a, t) &= \arg\min_u H^\star(x^a, u, t) \\ &= \arg\min_u Q^\star(x^a, u, t)\end{aligned} \tag{9}$$

Writing $\underline{H}(x^a, t) = \min_u H(x^a, u, t)$ for any function of (x^a, u, t), we have $\underline{H}^\star(x^a, t) = -\sigma J^\star(x^a, t)$. Further manipulations in [14], [15] lead to the ODE,

$$\begin{aligned}\tfrac{d}{dt}\underline{H}^\star(x^a_t, t) &= \sigma \underline{H}^\star(x^a_t, t) \\ &\quad + \sigma\big[c(x^a_t, u_t, t) - H^\star(x^a_t, u_t, t)\big]\end{aligned} \tag{10}$$

The boundary condition $\underline{H}^\star(x^a_\tau, \tau) = 0$ then gives the filtered Bellman equation: for any input-state trajectory and $t \in [0, \tau]$,

$$\underline{H}^\star(x^a_t, t) = -\int_t^\tau \sigma e^{\sigma(r-t)}[c_r - H^\star_r]\, dr \tag{11}$$

with $c_r = c(x_r, u_r, r)$ and $H^\star_r = H^\star(x_r, u_r, r)$. It is worth emphasizing that (11) holds for *any* input-state pair $\{x^a_r, u_r : 0 \le r \le \tau\}$.

Sample path representations of a Bellman equation are a starting point of many approaches to RL, such as Watkins' Q-learning algorithm.

Q learning: Given a family of approximations $\{H^\theta : \theta \in \mathbb{R}^d\}$, obtain θ^* so that (11) is approximately solved using training data $\{(x^k_t, z^k_t, u^k_t) : 0 \le t \le \tau, 1 \le k \le N\}$ where $N \ge 1$ denotes the number of independent runs. For any parameter θ, the policy ϕ^θ is defined in analogy with (9):

$$\phi^\theta(x^a, t) = \arg\min_u H^\theta(x^a, u, t) \tag{12}$$

This is of course a vague definition of an algorithm. A version of convex Q-learning is considered in [14], [15], in which the approximation is defined via a convex program.

Challenges with the optimal control solution are discussed in the next section, along with a remedy based on RL.

III. Fragility of Cheap Control

Fig. 3 may raise concern about the optimal control solution. The goal in the experiments leading to this figure was to flatten net-demand, with ℓ not nearly as volatile as the net-demand shown in Fig. 1. We see in Fig. 3 that the ACs and residential water heaters are roughly aligned in power deviation, but the downward ramp in power consumption from the ACs in the afternoon occurs at nearly the same time as an upward ramp in power consumption by the pools.

Fig. 4 shows results from other experiments surveyed in [18], [16]. The sum $z^{\sigma\star}$ of the power deviations is smooth, and serves to flatten nominal net-demand. The massive volatility of individual DERs might present problems, say, at the distribution level. Moreover, these experiments suggest very high sensitivity to model parameters. If the values of $\{\alpha_i\}$ in (1) are off by a few percent, we can expect massive changes in the optimizer $z^\star$.

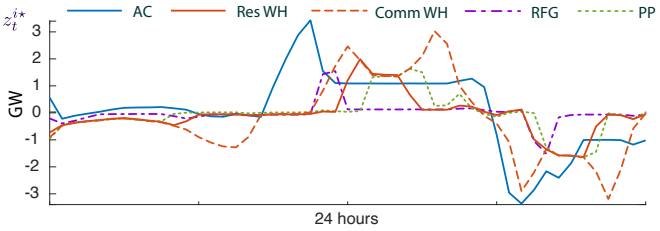

Fig. 4: Optimal trajectories $\{z^{i\star}_t\}$ based on five classes of loads to create five DERs: residential air-conditioning (AC), residential water heaters, commercial water heaters, residential refrigeration, and pool pumps for cleaning.

A. Addressing volatility

Two potential solutions might be considered in future research, each based on a modification of the objective function:

Cost on DER ramping. If the objective function is modified to include a penalty, such as $\sum_i \big(\frac{d}{dt} z^{i\star}_t\big)^2$ then we can expect a smoother optimal solution.

Cost to encourage consensus. This might involve adding to the objective the term

$$\sum_{i=1}^M \int_0^\tau \big[z^{i\star}_t - M^{-1} z^{\sigma\star}_t\big]^2\, dt$$

The first approach addresses volatility directly. However, we lose state space collapse since the control cost is no longer "cheap": recall that $\frac{d}{dt} z^i = u^i$.

State space collapse is preserved in the second approach, which deserves further study.

These approaches do not directly address model uncertainty, a topic considered only in very recent research based on techniques from reinforcement learning.

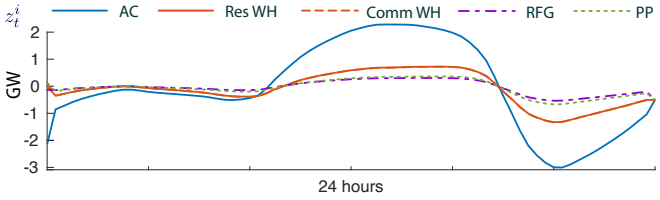

Fig. 5: Deviations $\{z^i_t\}$ from a convex Q-learning approximation.

B. Robust training for convex Q-learning

Without space for details on the Q-learning algorithm considered, we describe here an approach to marry MPC and Q-learning for policy synthesis. Suppose that H^{θ^*} is an approximation of $H^\star$, and let $J^{\theta^*} = -\underline{H}^{\theta^*}/\sigma$ be the corresponding approximation for the family of cost to go functions. For a given time-horizon $\tau \in (0, \tau]$, Bellman's principle (6) suggests the following feedback policy:

MPC-Q The policy is obtained through the following steps. For $t \le \tau - \tau$, obtain

$$u^{\theta^*}_{[t,t+\tau]} = \arg\min_{u^{t+\tau}_t} \Big\{ \int_t^{t+\tau} c(x^a_r, u_r, r)\, dr + J^{\theta^*}(x^a_{t+\tau}, t+\tau) \Big\} \tag{13}$$

and then set $\phi^{\text{MPC-Q}}(x_t^a, t) := u_t^{\theta^*}$.

That is, the input at time t for MPC-Q is defined by $u_t = \phi^{\text{MPC-Q}}(x_t^a, t)$. Of course, in any practical implementation the integral in (13) is replaced by a sum in an Euler approximation.

Note that MPC is not usually proposed as state feedback, but it does allow this interpretation.

In the experiments that follow the horizon τ is chosen far smaller than the total horizon $\mathcal{T}$. In addition the objective (2a) was modified to include a terminal cost,

$$\int_0^{\mathcal{T}} \left[c_{\text{g}}(g_t) + c_{\text{d}}(\gamma_t) + c_{\text{x}}(x_t)\right] dt + J_0(x_{\mathcal{T}}^a)$$

The terminal cost J_0 was set to zero in the foregoing to simplify exposition. The significant change in Section II-C is that $\underline{H}^\star(x_\tau^a, \mathcal{T}) = J_0(x_\tau^a)$ is the boundary condition for the ODE (10).

Theory-informed function class. It is reasonable to take into consideration both the model and theory in the construction of a function class.

The construction of [15] begins with an affine function class for the value function:

$$J^\theta(x^a, t) = J_0(x^a) + \theta^\intercal \psi(x^{\sigma,a}, t)\,, \quad \theta \in \mathbb{R}^d\,, \tag{14}$$

with $\psi\colon \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}^d$. A typical basis function was taken of the form $\psi_i(x^\sigma, z^\sigma, t) = q_i(x^\sigma, z^\sigma)p_i(t)$ in which q_i is quadratic in its two-dimensional argument, and p_i a mixture of Fourier basis elements and polynomials. To match the boundary condition $J^\theta(x^a, \mathcal{T}) = J_0(x^a)$, these functions were constructed so that $p_i(\mathcal{T}) = 0$ for each i.

The representation (8) then motivates

$$\begin{aligned} H^\theta(x^a, u, t) &:= -\sigma J^\theta(x^a, t) + Q^\theta(x^a, u, t) \\ Q^\theta(x^a, u, t) &:= c(x^a, u, t) \\ &\qquad + J_x^\theta(x^a, t)\cdot F(x^a, u, t) + J_t^\theta(x^a, t) \end{aligned} \tag{15}$$

with F the dynamics in the state space model, $\frac{d}{dt}x_t = F(x_t, u_t, t)$, defined by eqs. (2d) and (2e). The subscripts represent partial derivatives with respect to x and t.

Figs. 3 to 6 were generated under similar settings: five classes of loads, viz., air conditioning (ACs), residential water heating (res-WH), commercial water heating (comm-WH), refrigeration (RFG), and pool pumping (pp) are deployed in addition to traditional generation to balance the net load based on California's "duck curve" for a single day in March, 2020 (this data is obtained from CAISO). The parameters for the linear models are obtained from table I of [5]. The controlled vector field F appearing in (15) was based on the data from this table.

The data used for training the RL algorithm was obtained using perturbed dynamics:

Robust training. The approximation H^{θ^*} was obtained based on training data from a diverse collection of loads. In particular, during training the $\{\alpha_i\}$ vary widely, and disturbances were included in the load simulations. Details may be found in [15].

An example of numerical results. Fig. 5 shows results using MPC-Q (definition below (13)), with $\tau = 20$ mins. and $\mathcal{T} = 24$ hrs. The main conclusions:

1. A comparison of Figs. 4 and 5 shows that our primary goals has been achieved: the evolution of power deviations are far smoother and harmonious.

2. The policy obtained by setting $J^{\theta^*} \equiv 0$ results in the basic MPC algorithm. It is found that performance of this policy was not acceptable in this example using $\tau \le 40$ [15].

3. It was found that robust training typically improved performance *even on the nominal model*. Fig. 6 shows a comparison, taken from [14].

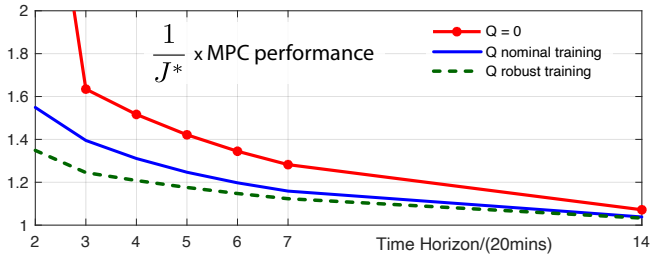

Fig. 6: Performance of MPC and MPC-Q on the nominal model.

IV. Conclusions

The optimal control problem (2) can be reduced to just two dimensions, regardless of the number of assets M. The solution can be represented as a distributed control architecture in which a *common* scalar command signal $\lambda^\star$ is broadcast to asset class (an aggregate of loads or a battery system).

We hope it is clear that the low marginal cost of balancing services from flexible loads and batteries comes with both blessings and potential curses. We have emphasized here the fragility of optimal control solutions, and surveyed potential approaches to mitigate this risk. The best way to avoid "spaghetti" outcomes, such as illustrated in Fig. 4, remains a topic for future research.

We are most excited about the potential for approaches based on reinforcement learning, building on what was briefly surveyed here. Theory is needed to better understand the benefits and risks when using robust training.

References

[1] H. Ballouz, J. Mathias, S. Meyn, R. Moye, and J. Warrington. Reliable power grid: Long overdue alternatives to surge pricing. *arXiv 2103.06355*, March 2021.

[2] E. Benenati, M. Colombino, and E. Dall'Anese. A tractable formulation for multi-period linearized optimal power flow in presence of thermostatically controlled loads. In *IEEE Conference on Decision and Control*, pages 4189–4194. IEEE, 2019.

[3] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Weihl. Demand dispatch. *IEEE Power and Energy Magazine*, 8(3):20–29, May 2010.

[4] N. Cammardella, A. Bušić, Y. Ji, and S. Meyn. Kullback-Leibler-Quadratic optimal control of flexible power demand. In *Proc. of the Conf. on Dec. and Control*, pages 4195–4201, Dec. 2019.

[5] N. Cammardella, J. Mathias, M. Kiener, A. Bušić, and S. Meyn. Balancing California's grid without batteries. In *Proc. of the Conf. on Dec. and Control*, pages 7314–7321, Dec 2018.

[6] Y. Chen, M. U. Hashmi, J. Mathias, A. Bušić, and S. Meyn. Distributed control design for balancing the grid using flexible loads. In S. Meyn, T. Samad, I. Hiskens, and J. Stoustrup, editors, *Energy Markets and Responsive Grids: Modeling, Control, and Optimization*, pages 383–411. Springer, New York, NY, 2018.

[7] M. Chertkov and V. Y. Chernyak. Ensemble control of cycling energy loads: Markov Decision Approach. In *IMA volume on the control of energy markets and grids*. Springer, 2018.

[8] B. Francis. The optimal linear-quadratic time-invariant regulator with cheap control. *IEEE Trans. Automat. Control*, 24(4):616–621, 1979.

[9] H. Hao, Y. Lin, A. Kowli, P. Barooah, and S. Meyn. Ancillary service to the grid through control of fans in commercial building HVAC systems. *IEEE Trans. on Smart Grid*, 5(4):2066–2074, July 2014.

[10] H. Hao, B. Sanandaji, K. Poolla, and T. Vincent. A generalized battery model of a collection of thermostatically controlled loads for providing ancillary service. In *51st Annual Allerton Conference on Communication, Control, and Computing*, pages 551–558, Oct 2013.

[11] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent. Aggregate flexibility of thermostatically controlled loads. *IEEE Trans. on Power Systems*, 30(1):189–198, Jan 2015.

[12] H. Lo, S. Blumsack, P. Hines, and S. Meyn. Electricity rates for the zero marginal cost grid. *The Electricity Journal*, 32(3):39 – 43, 2019.

[13] F. Lu. *Convex Q-learning: theory and applications*. PhD thesis, University of Florida, 2023.

[14] F. Lu, J. Mathias, S. Meyn, and K. Kalsi. Model-free characterizations of the Hamilton-Jacobi-Bellman equation and convex Q-learning in continuous time. *arXiv.2210.08131*, 2022.

[15] F. Lu, J. Mathias, S. Meyn, and K. Kalsi. Convex Q-learning in continuous time with application to dispatch of distributed energy resources. In *IEEE Conference on Decision and Control*, 2023.

[16] J. Mathias. *Balancing the power grid with distributed control of flexible loads*. PhD thesis, University of Florida, Gainesville, FL, USA, 2022.

[17] J. Mathias, A. Bušić, and S. Meyn. Load-level control design for demand dispatch with heterogeneous flexible loads. *IEEE Transactions on Control Systems Technology*, pages 1–14, 2023.

[18] J. Mathias, S. Meyn, H. Ballouz, and M. Ansari. A distributed control architecture for optimal allocation of grid-responsive load aggregations. In *Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2022.

[19] J. Mathias, R. Moye, S. Meyn, and J. Warrington. State space collapse in resource allocation for demand dispatch. In *Proc. of the Conf. on Dec. and Control*, pages 6181–6188 (and arXiv:1909.06869), Dec 2019.

[20] J. Mathias, R. Moye, S. Meyn, and J. Warrington. State space collapse in resource allocation for demand dispatch and its implications for distributed control design. *IEEE Trans. Automat. Control*, page pp, 2023.

[21] J. Mathieu and D. Callaway. State estimation and control of heterogeneous thermostatically controlled loads for load following. In *45th International Conference on System Sciences*, pages 2002–2011, Hawaii, 2012. IEEE.

[22] P. G. Mehta and S. P. Meyn. Q-learning and Pontryagin's minimum principle. In *Proc. of the Conf. on Dec. and Control*, pages 3598–3605, Dec. 2009.

[23] S. Meyn, P. Barooah, A. Bušić, Y. Chen, and J. Ehren. Ancillary service to the grid using intelligent deferrable loads. *IEEE Trans. Automat. Control*, 60(11):2847–2862, Nov 2015.

[24] V. Saksena, J. O'Reilly, and P. Kokotovic. Singular perturbations and time-scale methods in control theory: Survey 1976–1983. *Automatica*, 20(3):273 – 293, 1984.

[25] S. H. Tindemans, V. Trovato, and G. Strbac. Decentralized control of thermostatic loads for flexible demand response. *IEEE Transactions on Control Systems Technology*, 23(5):1685–1700, Sept 2015.