Kullback-Leibler-Quadratic Optimal Control in a Stochastic Environment

Neil Cammardella, Ana Bušić, Sean Meyn

Kullback-Leibler-Quadratic Optimal Control in a Stochastic Environment

Neil Cammardella[1], Ana Bušić[2], and Sean Meyn[1]

***Abstract*— This paper presents advances in Kullback-Leibler-Quadratic (KLQ) optimal control: a stochastic control framework for Markovian models. The motivation is distributed control of large networked systems. The objective function is composed of a control cost in the form of Kullback-Leibler divergence plus a quadratic cost on the sequence of marginal distributions. With this choice of objective function, the optimal probability distribution of a population of agents over a finite time horizon is shown to be an exponential tilting of the nominal probability distribution. The same is true for the controlled transition matrices that induce the optimal probability distribution.**

However, one limitation of the previous work is that randomness can only be introduced via the control policy; all uncontrolled processes must be modeled as deterministic to render them immutable under an exponential tilting. In this work, only the controlled dynamics are subject to tilting, allowing for more general probabilistic models.

Numerical experiments are conducted in the context of power networks. The distributed control techniques described in this paper can transform a large collection of flexible loads into a 'virtual battery' capable of delivering the same grid services as traditional batteries. Additionally, quality of service to the load owner is guaranteed, privacy is preserved, and computation and communication requirements are reduced, relative to alternative centralized control techniques.

I. Introduction

The setting of this paper is optimal control of Markov Decision Processes (MDPs). The state space S and input space U are assumed to be finite. A finite time horizon is considered, indexed by $\{k : 1 \le k \le K\}$. The controlled transition matrix T_k defines the statistics of the state process $\boldsymbol{S}$ with input process $\boldsymbol{U}$: $T_k(x, s') =$

$$\mathsf{P}\{S_{k+1} = s' \mid S_i,\ U_i,\quad 0 \le i \le k;\ S_k = s,\ U_k = u\}$$

The policies $\{\phi_k\}$ are assumed to be Markovian: $\phi_k(u \mid s) =$

$$\mathsf{P}\{U_k = u \mid S_i\,, U_i\,, 0 \le i < k;\ S_k = s\} \tag{1}$$

As in [1], [2], the Kullback-Leibler-Quadratic (KLQ) optimization criterion is based on convex functions of the marginal probability mass functions (pmfs) of the joint state-input process $X_k = (S_k, U_k)$:

$$\nu_k(x) = \mathsf{P}\{S_k = s, U_k = u\}\,, \quad x = (s, u) \in \mathsf{X} \tag{2}$$

Funding from the National Science Foundation under award EPCN 1935389 and French National Research Agency grant ANR-16-CE05-0008 is gratefully acknowledged.

[1]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611. SM's research also supported through an INRIA International Chair.

[2]Inria and DI ENS, École Normale Supérieure, CNRS, PSL Research University, Paris, France

where $\mathsf{X} = \mathsf{S} \times \mathsf{U}$ is the state-input space.

A sequence $\{\mathcal{C}_k\}$ of cost functions is given, and the control objective is to obtain the solution to the optimization problem

$$J^*(\nu_0) = \min_{\boldsymbol{U}} \sum_{k=1}^{K} \mathcal{C}_k(\nu_k) \tag{3}$$

where the minimum is over all randomized policies. Two classes of constraints are imposed. First, the initial pmf ν_0 for X_0 is fixed. Second are the dynamics, which can be expressed as a sequence of linear constraints on the marginals:

$$\sum_{u'} \nu_k(s', u') = \sum_{s,u} \nu_{k-1}(s, u) T_k(x, s')\,, \quad s' \in \mathsf{S} \tag{4}$$

One standard application is variance penalized MDPs. Our motivation is applications to mean field control as an approach to distributed control

A. Distributed control

The general optimization problem (3) falls outside of textbook stochastic control problems. It is inspired by mean-field game theory [3], [4], [5], [6], [7], [8] (see [9], [10] for recent surveys), and motivated in particular by applications to distributed control [11].

The control objective emerges from the approximation of a particular distributed control problem: a central authority wishes to shape the aggregate behavior of $\mathcal{N} \gg 1$ homogeneous agents, each modeled by the transition kernel T_k, with state-input denoted $\{X_k^i = (S_k^i, U_k^i) : 1 \le i \le \mathcal{N}\}$. The sequence of empirical distributions is denoted

$$\nu_k^{\mathcal{N}}(x) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \mathbb{I}\{S_k^i = s\,,\ U_k^i = u\}, \quad x = (s, u) \in \mathsf{X}$$

The optimization criterion of interest is (3), but with ν_k replaced by $\nu_k^{\mathcal{N}}$. A mean-field control approximation is justified by applying the law of large numbers: fix a sequence of randomized policies $\{\phi_k : 0 \le k \le K\}$, and consider $\mathcal{N}$ as a variable. The empirical distribution $\{\nu_k^{\mathcal{N}}\}$ converges as $\mathcal{N} \to \infty$ for each k, and the limit satisfies the linear constraints (4).

Section IV focuses on a homogeneous population of residential refrigerators for the creation of "virtual energy storage" for power grid applications. The goal is to shape the power usage of the population of loads. Let $\mathcal{Y}\colon \mathsf{X} \to \mathbb{R}_+$ denote power consumption as a function of state, so that the average power consumption of the population tracks the reference signal r, with acceptable error. The temperature of

the ith load at time k is denoted Θ_k^i, and the power mode (0 or 1) is denoted M_k^i. A typical linear model is given by

$$\Theta_{k+1}^i = \Theta_k^i + \alpha[\Theta^a - \Theta_k^i] - \beta M_k^i \tag{5}$$

where $\alpha > 0$, $\beta > 0$ and $\Theta^a \in \mathbb{R}$ denotes ambient temperature. This is a (deterministic) MDP model with state-input given by $X_k^i = (\Theta_k^i, M_k^i)$. In this work, our control design allows for the inclusion of disturbances in the model (5).

B. Literature review

Our primary motivation is application to distributed control of power systems, specifically *Demand Dispatch*. The term was introduced in the conceptual article [12] to describe the possibility of distributed intelligence in electric loads, designed so that the population would help provide supply-demand balance in the power grid. Contributing to this science has been a focus of the authors for the past decade [13], [14], [11], and many others (see [11] for a recent bibliography).

The goal in much of this prior work is to modify the behavior of loads so that their aggregate power consumption tracks the reference signal r that is broadcast by a *Balancing authority* (BA), based on distributed control, with local randomized decision rules. Randomized control techniques for Demand Dispatch have been proposed in [15], [16], [17], [18] based on entirely different control architectures.

The following control strategy is common to all of the approaches described in [14], [11]. It is assumed that a family of transition matrices $\{P_\zeta : \zeta \in \mathbb{R}\}$ is available at each load. A sequence $\{\zeta_0, \zeta_1, \dots\}$ is broadcast from the BA, based on measurements of the grid, and at time k an individual load transitions according to this law:

$$\mathsf{P}\{X_{k+1} = x' \mid X_k = x,\ \zeta_k = \zeta\} = P_\zeta(x, x')$$

The paper [19] re-interprets the control solution of [20] as a technique to create the family $\{P_\zeta\}$ through the solution to the nonlinear program:

$$P_\zeta := \max_{\pi, P} \left\{\zeta \langle \pi, \mathcal{Y} \rangle - \mathcal{K}(P \| P^0)\right\}, \qquad \zeta \in \mathbb{R}, \tag{6}$$

where $\mathcal{K}$ denotes the infinite-horizon relative entropy rate for two Markov chains:

$$\mathcal{K}(P\|P^0) := \sum_{x,x'} \pi(x) P(x,x') \log\Big(\frac{P(x,x')}{P^0(x,x')}\Big) \tag{7}$$

in which π is the invariant pmf for P. The maximum in (6) is over all (π, P) subject to the invariance constraint $\pi P = \pi$ [19], [21], [11].

The optimal control formulation here is also based on (7), with similar motivation, but the control approaches are entirely different. Further discussion is postponed to Section II-C.

Organization. The remainder of this paper is organized as follows: Section II contains a description of our control objective and a characterization of our main result. Computation of the optimal policy $\{\phi_k^*\}$ is described in Section III. Results from numerical experiments are collected together in Section IV. Conclusions and directions for future research are contained in Section V. The appendix contains abbreviated proofs of some of the results of this paper. Complete proofs can be found in [22].

II. Kullback-Leibler-Quadratic control

A. Optimal control formulation

KLQ is designed to balance two objectives:

(i) $\nu_k \sim \nu_k^0$, where $\{\nu_k^0\}$ models nominal behavior.

(ii) $\langle \nu_k, \mathcal{Y} \rangle \approx r_k$, where $\{r_k\}$ is a reference signal, and $\mathcal{Y} \colon \mathsf{X} \to \mathbb{R}$.

A pmf p^0 on X^{K+1} defines the nominal model:

$$p^0(\vec{x}) = \nu_0^0(x_0) P_0^0(x_0, x_1) P_1^0(x_1, x_2) \cdots \tag{8}$$

where $\vec{x}$ denotes the elements of X^{K+1}, $\{P_k^0\}$ are the nominal Markov transition matrices with transition probabilities $P_k^0(x,x') := \mathsf{P}\{X_{k+1} = x' \mid X_k = x\}$, and the nominal marginal pmfs are:

$$\nu_k^0(x_k) = \sum_{x_i : i \neq k} p^0(\vec{x}), \qquad \vec{x} \in \mathsf{X}^{K+1}$$

The nominal transition probabilities are represented as a product of two conditional pmfs:

$$P_k^0(x,x') = T_k(x,s') \phi_{k+1}^0(u' \mid s'), \qquad x, x' \in \mathsf{X} \tag{9}$$

where $\{\phi_k^0\}$ denotes the nominal randomized policy. Also, any randomized policy $\{\phi_k\}$ produces Markov transition matrices $\{P_k\}$ with transition probabilities

$$P_k(x,x') = T_k(x,s') \phi_{k+1}(u' \mid s'), \quad x = (s,u) \in \mathsf{X} \tag{10}$$

The marginal pmfs evolve according to linear dynamics, which are shown to be equivalent to (4):

$$\nu_k = \nu_{k-1} P_{k-1}, \qquad 1 \le k \le K \tag{11}$$

where the kth marginal ν_k is interpreted as a d-dimensional row vector.

The two control objectives motivate the cost function considered in this paper:

$$\mathcal{C}_k(\nu) = \mathcal{D}(\nu, \nu_k^0) + \frac{\kappa}{2}\big[\langle \nu, \mathcal{Y} \rangle - r_k\big]^2$$

in which $\kappa > 0$ is a penalty parameter, and $\mathcal{D}$ penalizes deviation from nominal behavior. The finite-horizon optimal control problem is thus

$$J^*(\nu_0^0) = \min \sum_{k=1}^K \left[\mathcal{D}(\nu_k, \nu_k^0) + \frac{\kappa}{2}\big[\langle \nu_k, \mathcal{Y} \rangle - r_k\big]^2\right] \tag{12}$$

where the initial pmf ν_0^0 is given.

The relative entropy rate (7) will be adopted as the cost of deviation. Under our assumptions, this reduces to

$$\mathcal{D}(\nu_k, \nu_k^0) := \sum_{s,u} \nu_k(s,u) \log\Big(\frac{\phi_k(u \mid s)}{\phi_k^0(u \mid s)}\Big) \tag{13}$$

The terminology is justified through the following steps. First, we have seen that any randomized policy gives rise to a pmf p that is Markovian:

$$p(\vec{x}) = \nu_0^0(x_0)P_0(x_0,x_1)P_1(x_1,x_2)\cdots$$

where P_k is defined in (10), and the initialization ν_0^0 is specified. The *relative entropy* is the mean log-likelihood:

$$D(p\|p^0) = \sum L(\vec{x})\, p(\vec{x}) \tag{14}$$

where $L = \log(p/p^0)$ is an extended-real-valued function on X^{K+1}. The expression for P_k in (10) and the analogous formula for P_k^0 using ϕ_{k+1}^0 gives

$$L(\vec{x}) = \log\Big(\frac{p(\vec{x})}{p^0(\vec{x})}\Big) = \sum_{k=1}^K \log\Big(\frac{\phi_k(u_k \mid s_k)}{\phi_k^0(u_k \mid s_k)}\Big) \tag{15}$$

Consequently, $D(p\|p^0) = \sum_{k=1}^K \mathcal{D}(\nu_k, \nu_k^0)$.

The optimal control problem (12), subject to the constraint (11), can be expressed

$$J^*(\nu_0^0) := \min_{\nu,\gamma} \sum_{k=1}^K \mathcal{D}(\nu_k,\nu_k^0) + \frac{\kappa}{2}\sum_{k=1}^K \gamma_k^2 \tag{16a}$$

$$\text{s.t. } \gamma_k + r_k - \langle\, \nu_k, \mathcal{Y}\,\rangle = 0 \tag{16b}$$

$$\sum_{u'} \nu_k(s',u') - \sum_{s,u} \nu_{k-1}(s,u)T_k(x,s') = 0 \tag{16c}$$

Prop. 2.1 asserts that the objective function is convex. It is also evident that the constraints (16b) and (16c) are linear in ν_k; hence, the optimization problem (16) is convex.

Proposition 2.1: The optimization problem (16) is jointly convex in $\{\nu_k, \gamma_k : 1 \le k \le K\}$. Furthermore, the constraint (16c) is equivalent to (11). □

An abbreviated proof can be found in Appendix A.

B. Dead-beat control

In Section 5 of the book chapter [23] a similar optimal control formulation is proposed:

$$\begin{aligned} &\min_p D(p\|p^0) \\ &\text{subject to } \mathsf{E}_p\big[\mathcal{Y}(X_k)\big] = r_k\,, \quad 1 \le k \le K \end{aligned} \tag{17}$$

Relative entropy is a useful measure of cost of deviation from nominal behavior because the optimizer has a simple form: a "tilting" (or "twisting") of the nominal model [20], [14]. This is motivation for the use of relative entropy in this prior work, which leads to the solution to (17):

$$p^*(\vec{x}) = p^0(\vec{x})\exp\Big(\sum_{k=1}^K \beta_k \mathcal{Y}(x_k) - \Lambda(\beta)\Big)$$

in which $\beta \in \mathbb{R}^K$ are Lagrange multipliers corresponding to the average power constraints, and $\Lambda(\beta)$ a normalizing constant.

The optimization criterion (17) is a form of dead-beat control, which might cause concern: is the optimization problem feasible? are there stability issues, as is well known for dead-beat control of linear systems?

For the KLQ formulation described in this paper, the tracking constraint in (17) is replaced by a quadratic loss function. As $\kappa \to \infty$ we recover the solution to the dead-beat control problem (however, our final result is different than the solution in [23] wherein ν_0 is not constrained).

The convex program formulation (16) has many advantages. First, (16) is always feasible, while feasibility of (17) requires conditions on p^0 and r. Theorem 3.1 requires no assumptions on the model or reference signal. Second is the value of flexibility in choice of κ, so that we can *learn* what is an "expensive" reference signal. It is anticipated that the penalty parameter κ can be used to make tradeoffs between tracking performance and robustness to modeling error: robustness and sensitivity analysis will be a topic of future research.

C. KLQ and IPD

The finite-horizon version of (6) is also considered in [14], [21], similar to the KLQ formulation:

$$p^\zeta := \arg\max_p \Big\{\zeta \mathsf{E}_p\Big[\sum_{k=1}^K \mathcal{Y}(x_k)\Big] - D(p\|p^0)\Big\}\,. \tag{18}$$

This and (6) are versions of the *Individual Perspective Design* (IPD) [21].

The IPD design (18) has the following alternative interpretation. For a scalar $r_0 \in \mathbb{R}$, consider the constrained optimization problem

$$\begin{aligned} &\max_p \{-D(p\|p^0)\} \\ &\text{subject to } \mathsf{E}_p\Big[\sum_{k=1}^K \mathcal{Y}(x_k)\Big] = Kr_0\,, \quad 1 \le k \le K \end{aligned} \tag{19}$$

The dual function $\varphi^*\colon \mathbb{R} \to \mathbb{R}$ is defined by

$$\varphi^*(\lambda) = \max_p \Big\{\lambda \mathsf{E}_p\Big[\sum_{k=1}^K \mathcal{Y}(x_k)\Big] - D(p\|p^0)\Big\} - \lambda K r_0$$

where $\lambda \in \mathbb{R}$ is a Lagrange multiplier. It is evident that the optimizer $p^{*\lambda}$ is an IPD solution for each λ. Consequently, for each ζ, the IPD solution (18) also solves (19) for some scalar $r_0(\zeta)$.

Note however that the IPD approach is designed to construct a family of transition matrices $\{P_\zeta\}$. This is then used in a pure *feedback* control design in which $\{\zeta_k\}$ is broadcast to the loads from the BA. The present work is focused on feed-forward control.

III. Duality

Structure for the solution of (16) will be obtained by consideration of a dual, in which $\lambda \in \mathbb{R}^K$ and $g \in \mathbb{R}^{K\times|\mathsf{S}|}$ denote the vectors of Lagrange multipliers for the first and

second set of constraints, respectively. The Lagrangian is thus

$$\begin{aligned}\mathcal{L}(\nu,\gamma,\lambda,g) = &\sum_{k=1}^{K}\mathcal{D}(\nu_k,\nu_k^0) + \frac{\kappa}{2}\sum_{k=1}^{K}\gamma_k^2 \\ &+\sum_{k=1}^{K}\lambda_k \mathrm{b}_k + \sum_{k=1}^{K}\sum_{s'} g_k(s')\mathrm{c}_k\end{aligned} \tag{20}$$

where b_k and c_k refer to the left-hand-side of equations (16b) and (16c), respectively. The dual function is the minimum:

$$\varphi^*(\lambda,g) := \min_{\nu,\gamma}\mathcal{L}(\nu,\gamma,\lambda,g)$$

and the *dual* of the optimization problem (16) is defined as the maximum of the dual function φ^* over λ and g. We will see that there is no duality gap, so that for a quadruple $(\nu^*,\gamma^*,\lambda^*,g^*)$,

$$J^*(\nu_0^0) = \mathcal{L}(\nu^*,\gamma^*,\lambda^*,g^*) = \varphi^*(\lambda^*,g^*)\,.$$

In the following subsections we obtain a representation of the dual function that is suitable for optimization, and in doing so we obtain a representation for the optimal policy. Properties of the dual function are contained in Theorem 3.1 and Prop. 3.2 that follow. The statement of these results requires additional notation: define a function $\mathcal{T}_k^\lambda\colon \mathbb{R}^{|S|}\to\mathbb{R}^{|S|}$, for $f\colon S\to\mathbb{R}$, $\lambda\in\mathbb{R}^K$, and $s\in\mathbb{R}^{|S|}$ via $\mathcal{T}_k^\lambda(f;s) =$

$$\log\Big(\sum_u \phi_k^0(u\mid s)\exp\big(\sum_{s'}T_k(x,s')f(s') + \lambda_k\mathcal{Y}(s,u)\big)\Big)$$

The maximum of the dual function over g is denoted

$$\varphi^*(\lambda) := \max_g \phi^*(\lambda,g) = \varphi^*(\lambda,g^\lambda)$$

where g^λ is a maximizer:

$$g^\lambda \in \arg\max_g \phi^*(\lambda,g)$$

We will show that the sequence of functions g^λ is given by the recursion

$$g_k^\lambda = \mathcal{T}_k^\lambda(g_{k+1}^\lambda)\,,\quad 1\le k\le K\,,\quad \text{where } g_{K+1}^\lambda\equiv 0 \tag{21}$$

and denote:

$$G_k^\lambda(x_{k-1}) = \sum_s T_{u_{k-1}}(s_{k-1},s)g_k^\lambda(s) \tag{22}$$

Theorem 3.1: There exists a maximizer $\{\lambda_k^*, g_k^* : 1\le k\le K\}$ for φ^*, and there is no duality gap:

$$\varphi^*(\lambda^*,g^*) = J^*(\nu_0^0)$$

The optimal policy is obtained from $\{g_k^*\}$ via:

$$\begin{aligned}\phi_k^*(u\mid s) =& \phi_k^0(u\mid s)\exp\big(\sum_{s'}T_k(x,s')g_{k+1}^*(s') \\ &+\lambda_k^*\mathcal{Y}(s,u) - g_k^*(s)\big)\\ \textit{where}\ \ g_k^*(s) =& \mathcal{T}_k^\lambda(g_{k+1}^*;s)\ \ \textit{and}\ \ g_{K+1}^*\equiv 0\end{aligned} \tag{23}$$

□

Proposition 3.2: The following hold for the dual of (16): for each $\lambda\in\mathbb{R}^K$,

(i) A maximizer g^λ is given by (21)

(ii) The maximum of the dual function over g is the concave function

$$\varphi^*(\lambda) = \lambda^T r - \frac{1}{2\kappa}\|\lambda\|^2 - \langle\,\nu_0, G_1^\lambda\,\rangle \tag{24}$$

(iii) The function (24) is continuously differentiable, and

$$\frac{\partial}{\partial\lambda_k}\varphi^*(\lambda) = r_k - \frac{1}{\kappa}\lambda_k - \langle\,\nu_k^\lambda,\mathcal{Y}\,\rangle \tag{25}$$

where $\{\nu_k^\lambda\}$ is the sequence of marginals obtained from the randomized policy defined in (23), substituting $\{g_k^*\}$ by $\{g_k^\lambda\}$ defined in (i). □

To conclude this section, we provide representations of the log-likelihood ratio $L(\vec{x})$, relative entropy $D(p^\lambda\|p^0)$, and primal objective function,

$$J(p^\lambda,\nu_0^0) := D(p^\lambda\|p^0) + \frac{\kappa}{2}\sum_{k=1}^{K}\big(\langle\,\nu_k^\lambda,\mathcal{Y}\,\rangle - r_k\big)^2 \tag{26}$$

where p^λ is the pmf obtained from the randomized policy defined in (23), substituting $\{g_k^*\}$ by $\{g_k^\lambda\}$ defined in Prop. 3.2, part (i).

Corollary 3.3: The following hold for all $\{\lambda_k, g_k^\lambda : 1\le k\le K\}$:

(i) The log-likelihood ratio can be expressed:

$$L(\vec{x}) = \sum_{k=1}^{K}\{\Delta_k(x_{k-1},s_k) + \lambda_k\mathcal{Y}(x_k)\} - G_1^\lambda(x_0) \tag{27}$$

where for each k *(recalling* $x_k=(s_k,u_k)$*)*,

$$\Delta_k(x_{k-1},s_k) = G_k^\lambda(x_{k-1}) - g_k^\lambda(s_k) \tag{28}$$

(ii) The relative entropy is given by

$$D(p^\lambda\|p^0) = \sum_{k=1}^{K}\lambda_k\langle\,\nu_k^\lambda,\mathcal{Y}\,\rangle - \langle\,\nu_0, G_1^\lambda\,\rangle \tag{29}$$

(iii) The value of the primal is $J(p^\lambda,\nu_0^0) =$

$$-\langle\,\nu_0,G_1^\lambda\,\rangle + \sum_{k=1}^{K}\Big(\lambda_k\langle\,\nu_k^\lambda,\mathcal{Y}\,\rangle + \frac{\kappa}{2}\big(\langle\,\nu_k^\lambda,\mathcal{Y}\,\rangle - r_k\big)^2\Big)$$

□

The stochastic process $\{\Delta_k(X_{k-1},S_k)\}$ is a martingale difference sequence; it vanishes when nature is deterministic, reducing to the solution obtained in [1], [2].

IV. Numerical Experiments

Numerical experiments are conducted in the context of Demand Dispatch. The goal here is to modify the behavior of flexible loads such that their aggregate power consumption tracks a reference signal $\{r_k\}$ that is broadcast by a BA. Previous work has demonstrated the potential for pool pumps [14], HVAC systems [13], water heaters [24], and refrigerators [25] to provide grid services. The following numerical experiments demonstrate distributed control of a collection of homogeneous residential refrigerators.

A. Algorithms

We have found in examples that using gradient ascent on the dual function curve may be slow to converge, likely due to a large "overshoot" when applying standard first-order methods [1]. In the numerical results that follow we opt for proximal gradient methods [26]. Monte Carlo methods have also been used to estimate λ^*. This is motivated by the representation of the gradient in terms of the first-order statistics of the random variable $\{\mathcal{Y}(\mathsf{X})\}$ when $\mathsf{X} \sim \nu_k^\lambda$:

$$\mathsf{E}_k[\mathcal{Y}(\mathsf{X})] = \sum_{x_k \in \mathsf{X}} \nu_k^\lambda(x_k)\mathcal{Y}(x_k) = \langle \nu_k^\lambda, \mathcal{Y} \rangle \tag{30}$$

Lemma 4.1 follows from (25) combined with (30):

Lemma 4.1: For any $\lambda \in \mathbb{R}^K$ and $1 \le k \le K$,

$$\frac{\partial}{\partial \lambda_k} \varphi^*(\lambda) = r_k - \frac{1}{\kappa}\lambda_k - \mathsf{E}_k[\mathcal{Y}(\mathsf{X})]\,, \quad \mathsf{X} \sim \nu_k^\lambda. \tag{31}$$

□

B. Designing the nominal model

The nominal model is carefully designed to ensure the uncontrolled dynamics remain unchanged. The state-input space is the cartesian product of two state-input spaces: $\mathsf{X} = \mathsf{X}_\mathrm{n} \times \mathsf{X}_\mathrm{u}$, where X_n contains the uncontrollable components and X_u contains the the controllable components. The nominal state transition matrices (9) that define the nominal model (8) are products of T_k, which represents the uncontrollable components, and ϕ^0, which represents the nominal control policy. In order to satisfy this decomposition, the state at time k for a refrigerator is defined to be

$$S_k = (\theta_k, \theta_{k-1}, U_{k-1}) \tag{32}$$

where $\theta_k \in \mathbb{R}$ denotes the temperature inside the refrigerator and $U_k \in \{0, 1\}$ is the power mode. The uncontrollable component T_k can not be modified and is derived from the linear model (5), with the addition of disturbances to represent randomness from nature. The nominal policy ϕ_k^0 is designed to approximate the deterministic control that is standard for a refrigerator [11].

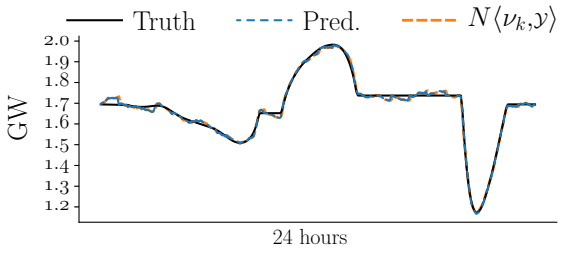

Fig. 1. 20 million refrigerators tracking a feasible signal

C. Designing the Experiment

The output process as a function of the state is defined to be the power consumption of a single refrigerator

$$\mathcal{Y}(x) = u\varrho\,, \qquad x = (s, u) \in \mathsf{X}\,.$$

where ϱ is its rated power consumption. In the plots shown below, the expected power consumption $\langle \nu_k, \mathcal{Y} \rangle$ and the reference signal r_k are multiplied by the number of refrigerators N to show the aggregate response. The reference signal is obtained from the solution to a separate optimization problem [27] which was designed to calculate optimal power trajectories for distinct classes of flexible loads. These power trajectories are optimal with respect to minimizing peaks and ramps in net load while satisfying quality of service constraints such as maintaining a given temperature range inside a refrigerator.

To simulate prediction error, the reference signal is created by adding random noise to the optimal power trajectory. Data from the California Independent System Operator (CAISO) show that hour-ahead predictions are fairly accurate, which inspires our use of model predictive control (MPC): fix two time periods T_0 and T, where $T_0 \ll T$; we choose $T_0 = 1$ hour and $T = 12$ hours. At the initial time t_0, the marginal pmf ν_0 is estimated, the reference signal is predicted over the time window $[t_0, t_0 + T]$, where its value at time t_0 is observed and has zero prediction error, and a solution is computed over the time window $[t_0, t_0 + T]$. Then, the solution is implemented, but restricted to the smaller time window $[t_0, t_0 + T_0]$. At time $t_0 + T_0$, the marginal pmf is estimated, the reference signal is predicted over the time window $[t_0 + T_0, t_0 + T_0 + T]$, where its value at time $t_0 + T_0$ is observed and has zero prediction error, a solution is computed over $[t_0 + T_0, t_0 + T_0 + T]$, and the process continues.

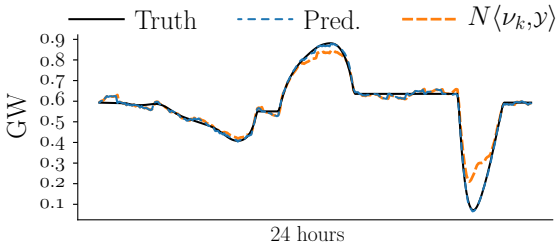

Fig. 2. 7 million refrigerators tracking an infeasible signal

D. Results

The first experiment, shown in Fig. 1, depicts an ideal case, where the reference signal is feasible, with respect to satisfying strict temperature bounds for each refrigerator. The variance of the prediction error is chosen to match the variance of the prediction error in hour-ahead forecasts observed by CAISO on February 25, 2021 [28]. The reference signal is an optimal power trajectory for 20 million refrigerators on a typical sunny day in California. Notice how the refrigerators are asked to consume more during the day when solar output is high and to ramp down during the evening peak. Tracking is nearly perfect in this ideal case.

The second experiment, shown in Fig. 2, depicts a case where the reference signal is made infeasible by reducing the number of participating loads to 7 million. Satisfying the energy requirements of this signal would require refrigerators

to violate their temperature bounds. This is not possible since quality of service is guaranteed with our distributed control control architecture, where switching decisions are still made at the load level [1]. Notice how the collective power consumption is gracefully truncated at the peak and valley of the reference signal while tracking remains nearly perfect at all other times.

V. Conclusions

A Lagrangian decomposition separates the finite-horizon optimal control problem into K separate convex programs, one for each time step. By applying a few well-known concepts from information theory, the optimal policy at each time step is found to be an exponential tilting of the nominal policy. The main novelty is to allow for more general Markovian models. Numerical experiments demonstrate the usefulness of this distributed control technique in a power system setting: a collection of flexible loads can be controlled such that their aggregate power consumption tracks a reference signal.

Plans for future research include:

(i) Evaluate robustness and sensitivity

(ii) Investigate techniques to reduce computation

(iii) Extend this stochastic KLQ formulation into a continuous-time setting and search for more effective relaxation techniques; initial work may be found in [29].

(iv) Consider other cost functions, e.g., the Wasserstein distance.

(v) Investigate the relationship between optimality and coupling of the pmfs, and the implications to control design.

(vi) Careful design of a terminal cost function may result in better performance for smaller time horizons [30].

Appendix

This appendix contains abbreviated proofs of some of the results of this paper. Complete proofs of every result can be found in [22].

A. Proof of Prop. 2.1

Proof: Clearly, (16) is convex in γ_k, a quadratic function of ν_k. Convexity in ν_k is established by constructing a sub-gradient for $\mathcal{D}(\nu_k, \nu_k^0)$: the function

$$g_{\mu,k}(s,u) = \log\Big(\frac{\mu(s,u)}{\nu_k^0(s,u)}\Big) - \log\Big(\frac{\hat{\mu}(s)}{\hat{\nu}_k^0(s)}\Big), \tag{33}$$

satisfies the sub-gradient property

$$\mathcal{D}(\nu_k, \nu_k^0) \geq \mathcal{D}(\mu, \nu_k^0) + \langle\, \nu_k - \mu, g_{\mu,k} \,\rangle \tag{34}$$

where $\hat{\mu}(s) = \sum_{u'} \mu(s,u')$ and $\hat{\nu}_k(s) = \sum_{u'} \nu_k(s,u')$. Hence, (16) is jointly convex in ν_k and γ_k since it is a sum of convex functions.

The equivalence claim is proven by multiplying both sides of (16c) by $\phi_k(u' \mid s')$, yielding

$$\mathsf{P}\{\mathsf{X}_k = x'\} = \sum_x \mathsf{P}\{\mathsf{X}_{k-1} = x\}\mathsf{P}\{\mathsf{X}_k = x' \mid \mathsf{X}_{k-1} = x\}$$

which is identical to (11). The proof of the implication (11) $\Longrightarrow$ (16c) is similar. A complete proof can be found in [22]. ■

B. Relative entropy and duality

The proofs of Theorem 3.1 and Prop. 3.2 make use of the following four lemmas. The first is based on a well known result regarding relative entropy. For any function $h\colon \mathsf{X}^{K+1} \to \mathbb{R}$ denote

$$Ł^0(h) := \sup_p \big\{ \langle p, h\rangle - D(p\|p^0) \big\} \tag{35}$$

Lemma 1.1 (Convex dual of relative entropy): For each p^0 and function $h\colon \mathsf{X}^{K+1} \to \mathbb{R}$, the (possibly infinite) value of (35) coincides with the log moment generating function:

$$Ł^0(h) = \log\langle\, p^0, e^h \,\rangle$$

Moreover, provided $Ł^0(h) < \infty$, the supremum in (35) is uniquely attained with $p^* = p^0 \exp(h - Ł^0(h))$. That is, the log-likelihood $L^* = \log(dp^*/dp^0)$ is given by

$$L^*(\vec{x}) = h(\vec{x}) - \Lambda_0(h)$$

□

Lemma 1.2: The dual function can be expressed

$$\begin{aligned}\varphi^*(\lambda, g) =& \lambda^T r - \frac{1}{2\kappa}\|\lambda\|^2 - \langle\, \nu_0, G_1^\lambda \,\rangle \\ &+ \sum_{k=1}^K \min_s \Big[g_k(s) - \mathcal{T}_k^\lambda(g_{k+1}; s) \Big]\end{aligned} \tag{36}$$

where $g_{K+1} \equiv 0$. □

Proof: First, substitute $\nu_k(s,u) = \hat{\nu}_k(s)\phi_k(u \mid s)$ in the Lagrangian (20) so that its minimization requires obtaining each of the K minimizers $\{\nu_k^{\lambda,g} : \nu_k^{\lambda,g}(s,u) = \hat{\nu}_k^{\lambda,g}(s)\phi_k^{\lambda,g}(u \mid s)\}$. It follows from Lemma 1.1 that the minimizers are given by

$$\begin{aligned}\phi_k^{\lambda,g}(u \mid s) =& \phi_k^0(u \mid s) \exp\Big(\sum_{s'} T_k(x,s') g_{k+1}(s') \\ &+ \lambda_k \mathcal{Y}(s,u) - \Lambda_k(s)\Big)\end{aligned} \tag{37}$$

with $\Lambda_k(s) = \mathcal{T}_k^\lambda(g_{k+1}; s)$.

Lemma 1.1 also gives the value:

$$\arg\min_{\phi_k} \mathcal{L} = -\mathcal{T}_k^\lambda(g_{k+1}; s) \tag{38}$$

resulting in $\min_\nu \mathcal{L}(\nu, \gamma, \lambda, g) =$

$$\begin{aligned}&\sum_{k=1}^K \Big(\frac{\kappa}{2}\gamma_k^2 + \lambda_k\gamma_k + \lambda_k r_k \Big) - \sum_{s,u} \nu_0^0(s,u) G_1^\lambda(s,u) \\ &+ \sum_{k=1}^K \min_{\hat{\nu}_k} \langle\, \hat{\nu}_k, g_k - \mathcal{T}_k^\lambda(g_{k+1}) \,\rangle\end{aligned} \tag{39}$$

Next, observe that the minimizer $\hat{\nu}_k^{\lambda,g}$ is obtained when the support of each $\hat{\nu}_k$ satisfies

$$\operatorname{supp}\big(\hat{\nu}_k(s)\big) \subseteq \arg\min_s \Big[g_k(s) - \mathcal{T}_k^\lambda(g_{k+1}; s) \Big]$$

so that

$$\min_s \Big[g_k(s) - \mathcal{T}_k^\lambda(g_{k+1}; s) \Big] = \langle\, \hat{\nu}_k^{\lambda,g}, g_k - \mathcal{T}_k^\lambda(g_{k+1}) \,\rangle$$

Also, the minimizer γ_k^λ is

$$\gamma_k^\lambda = -\frac{1}{\kappa}\lambda_k \tag{40}$$

Substituting the minimizers $\{\nu_k^{\lambda,g}, \gamma_k^\lambda\}$ into (39) results in (36). ■

C. Duality

Lemma 1.3: The maximum of the dual function over g is

$$\varphi^*(\lambda) := \max_g \varphi^*(\lambda, g) = \lambda^T \hat{r} - \frac{1}{2\kappa}\|\lambda\|^2 - \langle\, \nu_0, G_1^\lambda \,\rangle \tag{41}$$

with $G_1^\lambda(s,u) = \sum_{s'} T_k(x,s') g_1^\lambda(s')$. A maximizer g^λ is given by the recursive formula:

$$g_k^\lambda = \mathcal{T}_k^\lambda(g_{k+1}^\lambda)\,, \quad 1 \le k \le K\,, \quad \textit{where } g_{K+1}^\lambda \equiv 0\,, \tag{42}$$

□

Proof: Adding a constant to any of the $(g_1, g_2, \ldots, g_K)$ does not change the value of $\mathcal{L}$ or φ^* (this follows from (21)), so without loss of generality we assume, for each k,

$$\min_s \Big[g_k(s) - \mathcal{T}_k^\lambda(g_{k+1}; s) \Big] = 0 \tag{43}$$

and consequently

$$g_k \ge \mathcal{T}_k^\lambda(g_{k+1}) \;\; \textit{for each } k\,. \tag{44}$$

Thus, in view of (36), $\varphi^*(\lambda) =$

$$\lambda^T r - \frac{1}{2\kappa}\|\lambda\|^2 - \min_{g_1} \sum_{s,u} \nu_0^0(s,u) \sum_{s'} T_k(x,s') g_1(s')\,, \tag{45}$$

where the minimum is subject to the constraint (44). Next, observe that $\mathcal{T}_k^\lambda$ is a monotone operator, so that for each $k \le K$,

$$g_k \ge \mathcal{T}_k^\lambda \circ \mathcal{T}_{k+1}^\lambda \circ \cdots \circ \mathcal{T}_K^\lambda(g_{K+1}) \doteq g_k^\lambda\,, \;\; \textit{where } g_{K+1} \equiv 0$$

Based on the expression (45), we now show that the maximum $\arg\max_g \phi^*(\lambda, g)$ is obtained by choosing each g_k to reach this lower bound, giving (42). Indeed, g_1^λ achieves the minimum in (45), since $g_1^\lambda \le g_1$ for any g_1 for which (44) holds. This result along with (43) yields (41). ■

Lemma 1.4: The maximizers $\{g_k^\lambda\}$ have at most linear growth in $\|\lambda\|$:

$$|g_k^\lambda(s)| \le \|\mathcal{Y}\|_\infty \sum_{i=k}^K |\lambda_i| \le \sqrt{K}\|\mathcal{Y}\|_\infty \|\lambda\| \tag{46}$$

□

Proof: The proof is by induction, starting with the base case $k = K$. Then, we assume the hypothesis is true for $k \le K$ and show it holds for $k - 1$. ■

Next, we present a proof of Theorem 3.1.

Proof: We prove the existence of a maximizer λ^* by showing that $\phi^*(\lambda)$ is an anti-coercive function, i.e., $\phi^*(\lambda) \to -\infty$ as $\|\lambda\| \to \infty$. By Lemma 1.4,

$$\begin{aligned}
\varphi^*(\lambda) &= \lambda^T r - \frac{1}{2\kappa}\|\lambda\|^2 - \sum_{s,u} \nu_0^0(s,u) \sum_{s'} T_k(x,s') g_1^\lambda(s') \\
&\le \|\lambda\|\|r\| - \frac{1}{2\kappa}\|\lambda\|^2 + \max_{s'} |g_1^\lambda(s')| \\
&\le \|\lambda\|\|r\| - \frac{1}{2\kappa}\|\lambda\|^2 + \sqrt{K}\|\mathcal{Y}\|_\infty \|\lambda\|
\end{aligned}$$

Since $\phi^*(\lambda)$ is upper-bounded by an anti-coercive function, $\phi^*(\lambda)$ itself is an anti-coercive function. Thus a maximizer λ^* exists, and $(\lambda^*, g^*) = (\lambda^*, g^{\lambda^*})$ by (42).

The primal is a convex program, as established in Prop. 2.1. To show that there is no duality gap it is sufficient that Slater's condition holds [31, Section 5.3.2]. This condition holds: the relative interior of the constraint-set for the primal is non-empty since it contains $\{\nu_k^0\}$. Optimality of (23) is established by substituting g_{k+1}^* into (37) and by making the substitution $g_k^* = \mathcal{T}_k^\lambda(g_{k+1}^*)$ implied by (42). ■

Next, we present a proof of Prop. 3.2.

Proof: This proof has three parts:

(i) (21) is proven by Lemma 1.3.

(ii) (24) is proven by Lemma 1.3.

(iii) The representation of the derivative in part (iii) is standard (e.g., Section 5.6 of [31]), but we provide the proof for completeness. The representation (20) implies that φ^* is concave in (λ, g), since it is the infimum of linear functions. This representation also gives a formula for a derivative:

$$\frac{\partial}{\partial \lambda_k}\varphi^*(\lambda, g) = r_k - \frac{1}{\kappa}\lambda_k - \langle\, \nu_k^{\lambda,g}, \mathcal{Y} \,\rangle$$

where $\nu_k^{\lambda,g}$ is any optimizer in (39). Using $\varphi^*(\lambda) = \varphi^*(\lambda, g^\lambda)$ then gives

$$\frac{\partial}{\partial \lambda_k}\varphi^*(\lambda) = r_k - \frac{1}{\kappa}\lambda_k - \langle\, \nu_k^\lambda, \mathcal{Y} \,\rangle + \frac{\partial}{\partial g}\varphi^*\left(\lambda, g^\lambda\right) \cdot \frac{\partial}{\partial \lambda_k} g^\lambda$$

The first order condition for optimality gives $\frac{\partial}{\partial g}\varphi^*\left(\lambda, g^\lambda\right) = 0$, which completes the proof of the representation. It is evident that φ^* is continuously differentiable since ν_k^λ is continuously differentiable for each k by construction.

■

Next, we present a proof of Corollary 3.3.

Proof: This proof has three parts:

(i) Application of (15) and (23) results in the log-likelihood ratio:

$$\begin{aligned}
L(\vec{x}) &= \sum_{k=1}^K \Big(\sum_s T_k(x_k, s) g_{k+1}^\lambda(s) + \lambda_k \mathcal{Y}(x_k) - g_k^\lambda(s_k) \Big) \\
&= \sum_{k=1}^K \Big(G_{k+1}^\lambda(x_k) + \lambda_k \mathcal{Y}(x_k) - g_k^\lambda(s_k) \Big)
\end{aligned}$$

where the second identity follows from the definition (22). We have from the definitions, $G^\lambda_{K+1} \equiv 0$, which results in

$$L(\vec{x}) = -G^\lambda_1(x_0) + \sum_{k=1}^{K} \Big(G^\lambda_k(x_{k-1}) + \lambda_k \mathcal{Y}(x_k) - g^\lambda_k(s_k) \Big)$$

This combined with (28) yields (27).

(ii) Applying the definition of relative entropy as the mean log-likelihood, and noticing that $\mathsf{E}_{p^\lambda}\big[\Delta_k(X_{k-1}, S_k)\big] = 0$ for $1 \le k \le K$, results in $\sum_{\vec{x}} p^\lambda(\vec{x}) L(\vec{x}) =$

$$\sum_{\vec{x}} p^\lambda(\vec{x}) \left(\sum_{k=1}^{K} \lambda_k \mathcal{Y}(x_k) - G^\lambda_1(x_0) \right)$$

$$= \sum_{k=1}^{K} \sum_{x_k} \sum_{x_i, i \neq k} p^\lambda(\vec{x}) \lambda_k \mathcal{Y}(x_k) - \sum_{x_0} \sum_{x_i, i \neq 0} p^\lambda(\vec{x}) G^\lambda_1(x_0)$$

$$= \sum_{k=1}^{K} \lambda_k \langle\, \nu^\lambda_k, \mathcal{Y} \,\rangle - \langle\, \nu_0, G^\lambda_1 \,\rangle$$

(iii) Substitution of (29) into (26) results in (3).

■

REFERENCES

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

[2] N. Cammardella, A. Bušić, and S. Meyn, "Simultaneous allocation and control of distributed energy resources via Kullback-Leibler-Quadratic optimal control," in *American Control Conference*, July 2020, pp. 514–520.

[3] J. M. Lasry and P. L. Lions, "Mean field games," *Japan. J. Math.*, vol. 2, pp. 229–260, 2007.

[4] M. Huang, P. E. Caines, and R. P. Malhame, "Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ε-Nash equilibria," *IEEE Trans. Automat. Control*, vol. 52, no. 9, pp. 1560–1571, 2007.

[5] M. Huang, R. P. Malhame, and P. E. Caines, "Large population stochastic dynamic games: closed-loop Mckean-Vlasov systems and the Nash certainty equivalence principle," *Communications in Information and Systems*, vol. 6, no. 3, pp. 221–251, 2006.

[6] P. E. Caines, "Mean field games," in *Encyclopedia of Systems and Control*, J. Baillieul and T. Samad, Eds. London: Springer London, 2015, pp. 706–712. [Online]. Available: https://doi.org/10.1007/978-1-4471-5058-9_30

[7] O. Guéant, J.-M. Lasry, and P.-L. Lions, *Mean Field Games and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 205–266. [Online]. Available: https://doi.org/10.1007/978-3-642-14660-2_3

[8] H. Yin, P. Mehta, S. Meyn, and U. Shanbhag, "Synchronization of coupled oscillators is a game," *Automatic Control, IEEE Transactions on*, vol. 57, no. 4, pp. 920–935, 2012.

[9] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I: Mean Field FBSDEs, Control, and Games*, ser. Probability Theory and Stochastic Modelling. Springer International Publishing, 2018.

[10] ——, *Probabilistic Theory of Mean Field Games with Applications II: Mean Field Games with Common Noise and Master Equations*, ser. Probability Theory and Stochastic Modelling. Springer International Publishing, 2018.

[11] Y. Chen, M. U. Hashmi, J. Mathias, A. Bušić, and S. Meyn, "Distributed control design for balancing the grid using flexible loads," in *Energy Markets and Responsive Grids: Modeling, Control, and Optimization*, S. Meyn, T. Samad, I. Hiskens, and J. Stoustrup, Eds. New York, NY: Springer, 2018, pp. 383–411. [Online]. Available: https://doi.org/10.1007/978-1-4939-7822-9_16

[12] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Weihl, "Demand dispatch," *IEEE Power and Energy Magazine*, vol. 8, no. 3, pp. 20–29, May 2010.

[13] H. Hao, T. Middelkoop, P. Barooah, and S. Meyn, "How demand response from commercial buildings will provide the regulation needs of the grid," in *50th Allerton Conference on Communication, Control, and Computing*, 2012, pp. 1908–1913.

[14] S. Meyn, P. Barooah, A. Bušić, and J. Ehren, "Ancillary service to the grid from deferrable loads: The case for intelligent pool pumps in Florida," in *Proc. of the IEEE Conf. on Dec. and Control*, Dec 2013, pp. 6946–6953.

[15] J. Mathieu, S. Koch, and D. Callaway, "State estimation and control of electric loads to manage real-time energy imbalance," *IEEE Trans. Power Systems*, vol. 28, no. 1, pp. 430–440, 2013.

[16] S. H. Tindemans, V. Trovato, and G. Strbac, "Decentralized control of thermostatic loads for flexible demand response," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 5, pp. 1685–1700, Sept 2015.

[17] M. Almassalkhi, J. Frolik, and P. Hines, "Packetized energy management: asynchronous and anonymous coordination of thermostatically controlled loads," in *Proc. of the American Control Conf.* IEEE, 2017, pp. 1431–1437.

[18] 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*. IEEE, 2019, pp. 4189–4194.

[19] S. Meyn, P. Barooah, A. Bušić, Y. Chen, and J. Ehren, "Ancillary service to the grid using intelligent deferrable loads," *IEEE Trans. Automat. Control*, vol. 60, no. 11, pp. 2847–2862, Nov 2015.

[20] E. Todorov, "Linearly-solvable Markov decision problems," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 1369–1376.

[21] A. Bušić and S. Meyn, "Distributed randomized control for demand dispatch," in *Proc. of the IEEE Conf. on Dec. and Control*, Dec 2016, pp. 6964–6971.

[22] N. Cammardella, A. Bušić, and S. Meyn, "Kullback-Leibler-quadratic optimal control," *arXiv e-prints*, p. arXiv:2004.01798, April 2020. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2020arXiv200401798C

[23] 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, 2017.

[24] J. Mathias, A. Bušić, and S. Meyn, "Demand dispatch with heterogeneous intelligent loads," in *Proc. 50th Annual Hawaii International Conference on System Sciences (HICSS), and arXiv 1610.00813*, 2017.

[25] J. Mathias, R. Kaddah, A. Bušić, and S. Meyn, "Smart fridge / dumb grid? Demand Dispatch for the power grid of 2020," in *Proc. 49th Annual Hawaii International Conference on System Sciences (HICSS)*, Jan 2016, pp. 2498–2507.

[26] N. Parikh and S. Boyd, *Proximal Algorithms*, ser. Foundations and Trends in Optimization. Now Publishers, 2013. [Online]. Available: https://books.google.com/books?id=DS04ngEACAAJ

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

[28] California ISO – Folsom, CA 95763-9014, "ISO Today," Online www.caiso.com/Pages/TodaysOutlook.aspx.

[29] A. Bušić and S. Meyn, "Distributed control of thermostatically controlled loads: Kullback-Leibler optimal control in continuous time," in *Proc. of the IEEE Conf. on Dec. and Control*, Dec 2019, pp. 7258–7265.

[30] R.-R. Chen and S. P. Meyn, "Value iteration and optimization of multiclass queueing networks," *Queueing Syst. Theory Appl.*, vol. 32, no. 1-3, pp. 65–97, 1999.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. New York: Cambridge University Press, 2004.