Optimal Sensor Querying: General Markovian and LQG Models With Controlled Observations

Wei Wu, *Member, IEEE*, and Ari Arapostathis, *Fellow, IEEE*

***Abstract*—This paper is motivated by networked control systems deployed in a large-scale sensor network where data collection from all sensors is prohibitive. We model it as a class of discrete-time stochastic control systems for which the observations available to the controller are not fixed, but there are a number of options to choose from, and each choice has a cost associated with it. The observation costs are added to the running cost of the optimization criterion and the resulting optimal control problem is investigated. Since only part of the observations are available at each time step, the controller has to balance the system performance with the penalty of the requested information (query). We first formulate the problem for a general partially observed Markov decision process model and then specialize to the stochastic linear quadratic Gaussian problem. We focus primarily on the ergodic control problem and analyze this in detail.**

***Index Terms*—Dynamic programming, Kalman filter, linear quadratic Gaussian (LQG) control, networked control systems (NCS), partially observable Markov decision processes (POMDP).**

I. Introduction

MUCH attention has been recently paid to networked control systems (NCS), in which the sensors, the controllers and the actuators are located in a distributed manner and are interconnected by communication channels. In such systems, the information collected by sensors and the decisions made by controllers are not instantly available to the controllers and actuators, respectively. Rather they are transmitted through communication channels, which might suffer delay and/or transmission errors, and as such this transmission carries a cost. Understanding the interaction between the control system and the communication system becomes more and more important and plays a key role on the overall performance of NCS.

Stability is a basic requirement for a control system, and for NCS a key issue is how much information does a feedback controller need in order to stabilize the system. Questions of this kind have motivated much of the study of NCS: stability under communication constraints of linear control systems is studied by Wong and Brockett [1], [2], Tatikonda and Mitter [3], [4], Elia and Mitter [5], Nair and Evans [6], Liberzon [7] and many others; stability of nonlinear control systems is further studied in [8] and [9].

Broadly speaking, the amount of information the controller receives, affects the performance of estimation and control. However, information is not free. On the one hand, it consumes resources such as bandwidth, and power (i.e., in the case of a wireless channel), while on the other, by generating more traffic in the network it induces delays. If one incorporates in a standard optimal control problem an additional running penalty, associated with receiving the observations at the controller, then a tradeoff would result that balances the cost of observation and the performance of the control. In this paper, we consider a simple network scenario: a network of sensors, provides observations on the system state that are sent to the controller through a communication channel. The controller has the option of requesting different amounts of information from the sensors (i.e., more detailed or coarser observations), and can do so at each discrete time step. Based on the information received, an estimate of the state is computed and a control action is decided upon. However, what is different here is that there is a running cost, associated with the information requested, which is added to the running cost of the original control criterion. As a result the observation space is not static, rather changes dynamically as the controller issues different queries to the sensors.

Early work on the control of the observation process can be traced back to the seminal paper of Meier *et al.* [10], in which a separation principle between the optimal plant control and optimal measurement control was proved for finite-horizon linear quadratic Gaussian (LQG) control. Later on, work has focused on the optimal measurement control, or the so-called sensor scheduling problem [11]–[19], in which there are a number of sensors with different levels of precision and operation costs and the controller can access only one sensor at a time to receive the observation. The objective is to minimize a weighted average of the estimation error and observation cost. In [18], the sensor scheduling problem is addressed for continuous-time linear systems, while in [19], the dynamics correspond to a hidden Markov chain. Recently in [20], Gupta *et al.* propose computationally tractable algorithms to solve the stochastic sensor scheduling problem for the finite-horizon LQG problem.

In this work, we use a general model in the context of partially observed Markov decision processes (POMDP) with controlled observations and then specialize to the analogous infinite-horizon LQG problem. The feature that distinguishes

Manuscript received July 17, 2006; revised June 8, 2007. Published August 29, 2008 (projected). This paper was presented in part at the American Control Conference, Portland, OR, June 2005. This work was supported in part by the Office of Naval Research through the Electric Ship Research and Development Consortium, by the National Science Foundation under Grants ECS-0218207, ECS-0225448, and ECS-0424169, and the Hemphill-Gilmore Student Endowed Fellowship through the University of Texas at Austin. Recommended by Associate Editor I.-J. Wang.

The authors are with the Wireless Networking and Communications Group, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: wwu@ece.utexas.edu; ari@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2008.925817

our work from others' is that we study the optimal control problem over the infinite horizon—both the discounted (DC) and the long-term average (AC) control problems, for finite-state Markov chains and LQG control with controlled observations. We prove the existence of stationary optimal policies for MDPs with controlled observations and hierarchically structured observation spaces under a mild assumption. Then we consider the LQG control problem and prove that a partial separation principle of estimation and control holds over the infinite horizon: the optimal control can be decoupled into two subproblems, an optimal control problem with full observations and an optimal query/estimation problem requiring the knowledge of the controller gain. The estimation problem reduces to a Kalman filter, with the gain computed by a discrete algebraic Riccati equation (DARE). On the other hand, the optimal query is characterized by a dynamic programming equation.

The main contributions in this paper are highlighted as follows:

- existence of a stationary optimal policy for the ergodic control of finite-state POMDPs with controlled observations exhibiting hierarchical structure;
- the separation principle of estimation and control for the infinite-horizon LQG control problems with controlled observations;
- the characterization of optimal control for LQG control with controlled observations, including existence of a solution to the Hamilton-Jacobi-Bellman (HJB) equation for the ergodic control problem.

The rest of this paper is organized as follows. In Section II, we describe a model of POMDPs with controlled observations, and formulate an optimal control problem which includes a running penalty for the observation. Section III is devoted to the LQG problem. We present some examples in Section IV, and conclude the paper in Section V.

II. POMDPs With Controlled Observations

A. System Model and Problem Formulation

We consider the control of a dynamical system, which is governed by a Markov chain $(\boldsymbol{X},\boldsymbol{U},P,\mu)$, where $\boldsymbol{X}$ is the state space (assumed to be a Borel space), μ is the initial distribution of the state variable X_0 and $\boldsymbol{U}$ is the set of actions, which is assumed to be a compact metric space. We use capital letters to denote random processes and variables and lower case letters to denote the elements of the space they live in. We denote by $\mathcal{P}(\boldsymbol{X})$ the set of probability measures on $\boldsymbol{X}$. The dynamics of the process are governed by a transition kernel P on $\boldsymbol{X}$ given $\boldsymbol{X}\times\boldsymbol{U}$, which may be interpreted as

$$P_u(A\mid x)=\mathrm{Prob}(X_{t+1}\in A\mid X_t=x,\ U_t=u)$$

for $t=0,1,\ldots$, with A an element of the set of the Borel σ-field of $\boldsymbol{X}$, the latter denoted by $\mathfrak{B}(\boldsymbol{X})$.

The model includes ℓ distinct observation processes, but only one of these can be accessed at a time. Consider for example, a network of sensors providing observations for the control of a dynamical system. Suppose that there are ℓ levels of sensor information, and at each time t, Y_t^i represents the set of data provided at the i-th level, which lives in a space $\boldsymbol{Y}^{(i)}$. In as much as the complete set of data is a partial measurement of the state X_t of the system, we are provided with stochastic kernels $\mathcal{K}^i$ on $\mathcal{P}(\boldsymbol{Y}^{(i)})$ given $\boldsymbol{X}$, which may be interpreted as the conditional distribution of Y_t^i given X_t, i.e.,

$$\mathcal{K}^i(y\mid x)=\mathrm{Prob}(Y_t^i=y\mid X_t=x)\,.$$

The mechanism of sensor querying is facilitated by the query variable Q_t which chooses the subset of sensors to be queried at time t, i.e., takes values in $\boldsymbol{Q}=\{1,\ldots,\ell\}$. The evolution of the system is as follows: at each time t an action and query $(U_t,Q_t)=(u,q)\in\boldsymbol{U}\times\boldsymbol{Q}$ are chosen and the system moves to the next state X_{t+1} according to the probability transition function P_u, and the data set $Y_{t+1}^q\in\boldsymbol{Y}^{(q)}$, corresponding to the queried sensors, is obtained.

One special case of this model is when the levels of sensor information constitute a hierarchy, i.e., the data set becomes richer as we move up in the levels, meaning that the σ-fields are ordered by the inclusion $\sigma(Y_t^1)\subset\cdots\subset\sigma(Y_t^\ell)$. Another scenario, in the sensor scheduling problem, involves ℓ independent sensors with observations Y_t^i, and at each time t, only one can be accessed (e.g., due to interference).

Following the standard POMDP model formulation (e.g., [21]), we define $\boldsymbol{Y}\triangleq\bigcup_{q\in\boldsymbol{Q}}\boldsymbol{Y}^{(q)}$, and the history spaces $\{\boldsymbol{H}_t\}$ by $\boldsymbol{H}_0\triangleq\mathcal{P}(\boldsymbol{X})$ and

$$\boldsymbol{H}_{t+1}\triangleq\boldsymbol{H}_t\times\boldsymbol{U}\times\boldsymbol{Q}\times\boldsymbol{Y}\,,\quad t=0,1,\ldots.$$

Markov controls and stationary controls are defined in the standard manner. We let $\mathcal{V}$ denote all admissible controls, and $\mathcal{V}_M$, $\mathcal{V}_S$ all the Markov, stationary (Markov) controls respectively. Under a Markov control v the probability measure $\mathbb{P}_\mu^v$ renders (X_t,Y_t) a Markov process.

Following the theory of partially observed stochastic systems, we obtain an equivalent completely observed model through the introduction of the conditional distribution Ψ_t of the state given the observations [21]–[23]. The process Ψ_t lives in $\boldsymbol{\Psi}\triangleq\mathcal{P}(\boldsymbol{X})$. An important difference from the otherwise routine construction is that the observation process does not live in a fixed space but varies dynamically based on the query process. The query variable selects the observation space, and the nonlinear Bayesian filter that updates the state estimate is chosen accordingly. Let

$$\begin{aligned}\tilde{P}(dx,dy\mid\psi,u,q)&\triangleq\int_{x'\in\boldsymbol{X}}\mathcal{K}^q(dy\mid x)P_u(dx\mid x')\psi(dx')\\ V(dy,\psi,u,q)&\triangleq\int_{x\in\boldsymbol{X}}\tilde{P}(dx,dy\mid\psi,u,q)\,.\end{aligned}$$

Decomposing the measure $\tilde{P}$ as

$$\tilde{P}(dx,dy\mid\psi,u,q)=\mathcal{T}(\psi,y,u,q)(dx)\,V(dy,\psi,u,q)$$

we obtain the filtering equation

$$\psi_{t+1}=\mathcal{T}(\psi_t,y_{t+1},u_t,q_t)(dx)\,. \tag{1}$$

The model includes a running penalty $r:\boldsymbol{X}\times\boldsymbol{U}\to\mathbb{R}$, which is assumed to be continuous and non-negative, as well as a penalty function $c:\boldsymbol{Q}\to\mathbb{R}$ that represents the cost of information. Let $g = r + c$. We are interested primarily in the long-term average, or ergodic criterion. In other words, we seek to minimize, over all admissible policies $v \in \mathcal{V}$

$$J^v(x) \triangleq \limsup_{N\to\infty} \frac{1}{N} \mathbb{E}_x^v \left[\sum_{t=0}^{N-1} g(X_t, U_t, Q_t) \right]. \quad (2)$$

When the optimal value of $J^v(x)$ is independent of the initial condition x, then (2) is referred to as the ergodic criterion. We also consider the β-discounted criterion

$$J_\beta^v(x) \triangleq \mathbb{E}_x^v \left[\sum_{t=0}^{N-1} \beta^t g(X_t, U_t, Q_t) \right], \quad \beta \in (0,1). \quad (3)$$

We define

$$J^* \triangleq \inf_{v\in\mathcal{V}} J^v, \qquad J_\beta^*(x) \triangleq \inf_{v\in\mathcal{V}} J_\beta^v(x).$$

If we let $\tilde{g}(\psi, u, q) = \int g(x, u, q)\psi(dx)$, the control criteria in (2)–(3) can be expressed in the equivalent CO model.

Stationary optimal policies for the β-discounted cost objective can be characterized via the HJB equation, shown at the bottom of the page, where J_β is the optimal value function. For the long-term average or ergodic objective, the HJB equation takes the form of (4), shown at the bottom of the page. In (4), J^* is the optimal average cost, and h is called the bias function.

$$J_\beta(\psi) = \min_{(u,q)\in\boldsymbol{U}\times\boldsymbol{Q}} \Big\{ \tilde{g}(\psi, u, q) + \beta \int_{\boldsymbol{Y}^{(q)}} V(dy, \psi, u, q) J_\beta\big(\mathcal{T}(y, \psi, u, q)\big) \Big\}$$

$$J^* + h(\psi) = \min_{(u,q)\in\boldsymbol{U}\times\boldsymbol{Q}} \Big\{ \tilde{g}(\psi, u, q) + \int_{\boldsymbol{Y}^{(q)}} V(dy, \psi, u, q) h\big(\mathcal{T}(y, \psi, u, q)\big) \Big\} \quad (4)$$

B. POMDPs With Hierarchical Observations

In this section, we focus on models with the hierarchical structure $\sigma(Y_t^1) \subset \cdots \subset \sigma(Y_t^\ell)$. A simple example of such a structure is a temperature monitoring system in a large building that consists of a two-level hierarchy: a cluster of sensors at each room (fine observation space), and a cluster of sensors at each floor of the building (coarse observation space).

We consider a POMDP model with finite state space $\boldsymbol{X} = \{1, \ldots, n\}$ and observation spaces $\boldsymbol{Y}^{(q)} = \{1, \ldots, l_q\}$, $q \in \boldsymbol{Q}$. The action space $\boldsymbol{U}$ is assumed to be a compact metric space. The dynamics of the process are governed by a transition kernel on $\boldsymbol{X} \times \boldsymbol{Y}$

$$Q_{ij}^q(y, u) = \mathbb{P}(X_{t+1} = j, Y_{t+1}^q = y \mid X_t = i, U_t = u).$$

For fixed q, y and u, Q^q can be viewed as an $n \times n$ substochastic matrix and is assumed continuous with respect to u. Representing ψ as a row vector of dimension n, (1) takes the form

$$V(\psi, y, u, q) = \psi Q^q(y, u)\mathbf{1}$$
$$\mathcal{T}(\psi, y, u, q) = \begin{cases} \frac{\psi Q^q(y,u)}{V(\psi,y,u,q)}, & \text{if } V(\psi, y, u, q) > 0 \\ \bar{\psi}, & \text{otherwise} \end{cases}$$

where $\bar{\psi}$ can be chosen arbitrarily.

Under the hierarchical structure assumed, the observation space $\boldsymbol{Y}^{(q+1)}$, admits a partition $\{S_y^q,\ y \in \boldsymbol{Y}^{(q)}\}$, satisfying the property

$$Q^q(y, u) = \sum_{y' \in S_y^q} Q^{q+1}(y', u), \quad \forall (y, u) \in \boldsymbol{Y}^{(q)} \times \boldsymbol{U}. \quad (5)$$

Note that (5) implies that for any $y \in \boldsymbol{Y}^{(q)}$, $\mathcal{T}(\psi, y, u, q)$ can be expressed as a convex combination of

$$\{\mathcal{T}(\psi, y', u, q+1),\ y' \in S_y^q\}.$$

C. Existence of Stationary Optimal Policies for the Average Cost

In this section we employ results from [24] to show existence of a solution to (4) for a POMDP with hierarchical observations, by imposing a condition on the (finest) observation space $\boldsymbol{Y}^{(\ell)}$.

We adopt the following notation. For $k \in \mathbb{N}$, let

$$q^k \triangleq (q_0, \ldots, q_{k-1}) \in \boldsymbol{Q}^k, \qquad u^k \triangleq (u_0, \ldots, u_{k-1}) \in \boldsymbol{U}^k$$
$$y^k \triangleq (y_1, \ldots, y_k), \qquad \boldsymbol{Y}^{(q^k)} \triangleq \boldsymbol{Y}^{(q_0)} \times \cdots \times \boldsymbol{Y}^{(q_{k-1})}.$$

Then, provided q^k and y^k are compatible, i.e., $y^k \in \boldsymbol{Y}^{(q^k)}$, we define

$$Q(y^k, u^k, q^k) \triangleq Q^{q_{k-1}}(y_k, u_{k-1}) \cdots Q^{q_1}(y_2, u_1) Q^{q_0}(y_1, u_0).$$

Also, denote by $\{\ell\}^k$ the string $\{\ell, \ldots, \ell\}$ of length k.

Assumption 2.1: There exists $k_0 \in \mathbb{N}$ such that, for each $i \in \boldsymbol{X}$

$$\max_{1\le k\le k_0} \min_{u^k\in\boldsymbol{U}^k} \Big\{ \max_{y^k\in\boldsymbol{Y}^k} \min_{j\in\boldsymbol{X}} Q_{ij}(y^k, u^k, \{\ell\}^k) \Big\} > 0. \quad (6)$$

Remark 2.1: Perhaps a more transparent way of stating Assumption 2.1 is that for each $i \in \boldsymbol{X}$, and for each sequence

$u^{k_0} = (u_0, \dots, u_{k_0-1})$, there exists some $k \le k_0$ and a sequence y^k, such that $Q_{ij}(y^k, u^k, \{\ell\}^k) > 0$, for all $j \in \boldsymbol{X}$.

Remark 2.2: According to the results in [24], under Assumption 2.1, there exists a solution to (4), provided the observation space is restricted to $\boldsymbol{Y}^{(\ell)}$.

We have the following existence theorem.

Theorem 2.1: Let Assumption 2.1 hold. Then, there exists a solution (J^*, h) to (4), with $J^* \in \mathbb{R}$ and $h : \boldsymbol{\Psi} \to \mathbb{R}$, a concave function. Moreover, the minimizer in (4) defines a stationary optimal policy relative to the ergodic criterion, and J^* is the optimal cost.

Proof: By (5), for each $q \in \boldsymbol{Q}$, there exists a partition $\{\tilde{S}^q_y\,,\ y \in \boldsymbol{Y}^{(q)}\}$ of $\boldsymbol{Y}^{(\ell)}$, satisfying

$$Q^q(y, u) = \sum_{y' \in \tilde{S}^q_y} Q^\ell(y', u)\,, \quad \forall (y, u) \in \boldsymbol{Y}^{(q)} \times \boldsymbol{U}\,.$$

Then for any $q^k \in \boldsymbol{Q}^k$, $y^k \in \boldsymbol{Y}^{(q^k)}$, and $u^k \in \boldsymbol{U}^k$, we have

$$\begin{aligned} Q(y^k, u^k, q^k) &= \prod_{t=0}^{k-1} Q^{q_t}(y_{t+1}, u_t) \\ &= \prod_{t=0}^{k-1} \sum_{y' \in \tilde{S}^{q_t}_{y_{t+1}}} Q^\ell(y', u_t) \\ &= \sum_{\hat{y}^k \in \tilde{S}^{q_0}_{y_1} \times \cdots \times \tilde{S}^{q_{k-1}}_{y_k}} Q(\hat{y}^k, u^k, \{\ell\}^k)\,. \end{aligned} \tag{7}$$

Assuming (6) holds, fix $i \in \boldsymbol{X}$, $u^{k_0} = \{u_0, \dots, u_{k_0-1}\}$ and $q^{k_0} = \{q_0, \dots, q_{k_0-1}\}$, and let $\bar{y}^k$, $k \le k_0$ be such that

$$Q_{ij}(\bar{y}^k, u^k, \{\ell\}^k) > 0\,, \quad \forall j \in \boldsymbol{X}\,. \tag{8}$$

Since $\{\tilde{S}^q_y\,,\ y \in \boldsymbol{Y}^{(q)}\}$ is a partition of $\boldsymbol{Y}^{(\ell)}$, we can choose $y_t \in \boldsymbol{Y}^{(q_t)}$ such that $\bar{y}_t \in \tilde{S}^{q_t}_{y_t}$, for all $t = 0, \dots, k$. By (7)–(8)

$$Q_{ij}(y^k, u^k, q^k) \ge Q_{ij}(\bar{y}^k, u^k, \{\ell\}^k) > 0\,, \quad \forall j \in \boldsymbol{X}\,.$$

Therefore, Assumption 4 in [24] is satisfied, which yields the result. ■

III. Linear Systems With Controlled Observations

In this section, we consider a stochastic linear system in discrete-time, with quadratic running penalty. In Section III-A we introduce the LQG control model, and in Section III-C we study stability issues. The dynamic programming equation is further simplified and decoupled into two separate problems: (a) optimal estimation problem and (b) control, the latter being a standard LQG optimal control problem.

A. Linear Quadratic Gaussian Control: The Model

Consider a linear system governed by

$$X_{t+1} = AX_t + BU_t + DW_t\,, \quad t = 0, 1, \dots \tag{9}$$

where $X_t \in \mathbb{R}^{N_x}$ is the system state, $U_t \in \mathbb{R}^{N_u}$ is the control, and the noise process $\{W_t\}$ is i.i.d., and normally distributed. We assume that X_0 is Gaussian with mean $\bar{x}_0$ and covariance matrix Σ_0, and denote this by $X_0 \sim \mathcal{N}(\bar{x}_0, \Sigma_0)$. We also assume that X_0 and $\{W_t\,,\ t \ge 0\}$ are independent. The observation process is being observed by

$$Y_t = C_{Q_{t-1}} X_t + FW_t\,, \quad t \ge 1 \tag{10}$$

with $Y_t \in \mathbb{R}^{N_y}$, and $\det(FF^\mathsf{T}) \ne 0$. Moreover, we assume the system noise and observation noise are independent, i.e., $DF^\mathsf{T} = 0$. This independence assumption results in a simplification of the algebra; otherwise, it is not essential.

The running cost r is quadratic in the state and control, and takes the form

$$r(x, u) = x^\mathsf{T} R x + u^\mathsf{T} S u$$

where R and S belong to $\mathcal{M}^+$, the set of symmetric, positive definite matrices.

B. Optimal Control Over a Finite Horizon

For an initial condition X_0 and an admissible policy $v = \{(U_t, Q_t),\ t \ge 0\}$, let $\mathbb{P}^v$ denote the unique probability measure on the path space of the process, and $\mathbb{E}^v$ the corresponding expectation operator. Whenever needed we indicate the dependence on X_0 explicitly (or more precisely the dependence on the law of X_0), by using the notation $\mathbb{P}^v_{X_0}$ and $\mathbb{E}^v_{X_0}$. The optimal control problem over a finite horizon N, amounts to minimizing over all admissible controls v the functional

$$J^v_N \triangleq \mathbb{E}^v \left[\sum_{t=0}^{N-1} \big(c(Q_t) + r(X_t, U_t) \big) + X_N^\mathsf{T} \Pi_N X_N \right] \tag{11}$$

where $\Pi_N \in \mathcal{M}^+_0$, the set of symmetric, positive semi-definite matrices in $\mathbb{R}^{N_x \times N_x}$. In (11), J^v_N is of course a function of the law of X_0, and hence can be parameterized as $J^v_N = J^v_N(\bar{x}_0, \Sigma_0)$. The solution of the problem over a finite horizon is well known [10]. Nevertheless, we summarize the key results in the theorem below. A proof is included because some of the derivations are needed later.

Theorem 3.1: Consider the control system in (9)–(10), under the assumptions stated in Section III-A, and let $\bar{x}_0$ and Σ_0, be the mean and covariance matrix of X_0, respectively. Let

$$J^*_N(\bar{x}_0, \Sigma_0) \triangleq \inf_{v \in \mathcal{V}} J^v_N(\bar{x}_0, \Sigma_0)\,.$$

Let $v^* = \{(U^*_t, Q^*_t)\,,\ t = 0, \dots, N-1\}$, where $\{U^*_t\}$ is defined by

$$U^*_t = -K_t \hat{X}_t \tag{12}$$

with

$$K_t = (B^\mathsf{T}\Pi_{t+1}B + S)^{-1} B^\mathsf{T} \Pi_{t+1} A \tag{13a}$$

$$\begin{aligned} \Pi_t = {} & R + A^\mathsf{T}\Pi_{t+1}A \\ & - A^\mathsf{T}\Pi_{t+1}B(B^\mathsf{T}\Pi_{t+1}B + S)^{-1}B^\mathsf{T}\Pi_{t+1}A \end{aligned} \tag{13b}$$

and Q^*_t is a selector of the minimizer in the dynamic programming equation

$$f_t(\hat{\Pi}) = \min_q \left\{ c(q) + \operatorname{tr}(\tilde{\Pi}_t \hat{\Pi}) + f_{t+1}(\mathcal{T}_q(\hat{\Pi})) \right\} \tag{14}$$

$t=0,\ldots,N-1$, with $f_N \triangleq 0$, the Riccati map $\mathcal{T}_q$ as defined in (17), and

$$\tilde{\Pi}_t \triangleq R - \Pi_t + A^{\mathsf{T}}\Pi_{t+1}A, \quad t=0,\ldots,N-1.$$

Then v^* is optimal with respect to the cost functional J_N^v and

$$J_N^*(\bar{x}_0,\Sigma_0) = \tilde{J}_N^{U^*}(\bar{x}_0,\Sigma_0) + f_0(\Sigma_0)$$

where

$$\tilde{J}_N^{U^*}(\bar{x}_0,\Sigma_0) \triangleq \bar{x}_0^{\mathsf{T}}\Pi_0\bar{x}_0 + \mathrm{tr}(\Pi_0\Sigma_0) + \sum_{k=1}^{N}\mathrm{tr}(\Pi_k DD^{\mathsf{T}}).$$

Proof: Let $Y^t=\{Y_1,\ldots,Y_t\}$, and $\mathcal{Y}^t=\sigma(Y^t)$. Invoking the results of the general POMDP model in Section II, we can obtain an equivalent completely observed model using the conditional distribution of X_t given $\mathcal{Y}^t$ as the new state. It is well known that with respect to $\mathbb{P}^v$ the conditional distribution of X_t given $\mathcal{Y}^t$ is Gaussian [25]. Let $\hat{X}_t=\mathbb{E}^v[X_t \mid \mathcal{Y}^t]$. Since there is no observation Y_0 in our model, we set $\mathcal{Y}^0$ as the trivial σ-field. Hence, $\hat{X}_0=\mathbb{E}[X_0]=\bar{x}_0$. Then, a standard derivation, yields

$$\hat{X}_{t+1} = A\hat{X}_t + BU_t + \hat{K}_{Q_t}(\hat{\Pi}_t)\big(Y_{t+1} - C_{Q_t}(A\hat{X}_t+BU_t)\big) \quad (15)$$

where

$$\hat{K}_q(\hat{\Pi}) \triangleq \Xi(\hat{\Pi})C_q^{\mathsf{T}}\big(C_q\Xi(\hat{\Pi})C_q^{\mathsf{T}} + FF^{\mathsf{T}}\big)^{-1} \quad (16a)$$

$$\hat{\Pi}_{t+1} = \Xi(\hat{\Pi}_t) - \hat{K}_{Q_t}(\hat{\Pi}_t)C_{Q_t}\Xi(\hat{\Pi}_t) \quad (16b)$$

$$\Xi(\hat{\Pi}) \triangleq DD^{\mathsf{T}} + A\hat{\Pi}A^{\mathsf{T}}. \quad (16c)$$

In (16), $\hat{\Pi}_t$ is the conditional covariance of $X_t-\hat{X}_t$ under $\mathbb{P}^v$ given $\mathcal{Y}^t$, and $\hat{\Pi}_0=\Sigma_0$. By (16b), the conditional error covariance matrix $\hat{\Pi}_t$ satisfies $\hat{\Pi}_{t+1}=\mathcal{T}_{Q_t}(\hat{\Pi}_t)$, where

$$\mathcal{T}_q(\hat{\Pi}) \triangleq \Xi(\hat{\Pi}) - \hat{K}_q(\hat{\Pi})C_q\Xi(\hat{\Pi}). \quad (17)$$

If an admissible sequence $\{Q_t,\ t\ge 0\}$ is specified, then standard LQG theory shows that the policy $\{U_t^*,\ t=0,\ldots,N-1\}$, given by (12)–(13), is optimal relative to the functional J_N^v. In other words, if we denote $J_N^v = J_N^{U,Q}$ with $Q=\{Q_t,\ t\ge 0\}$ fixed, then $U^*=\{U_t^*,\ t\ge 0\}$ defined in (12) satisfies

$$J_N^{U^*,Q} = \inf_{\tilde{U}} J_N^{\tilde{U},Q}$$

where the infimum is over all admissible policies $\tilde{U}$.

Combining the feedback policy in (12) with (15), we obtain

$$\hat{X}_{t+1} = (A-BK_t)\hat{X}_t + \hat{K}_{Q_t}(\hat{\Pi}_t)C_{Q_t}A(X_t-\hat{X}_t) + \hat{K}_{Q_t}(\hat{\Pi}_t)(C_{Q_t}DW_t + FW_{t+1}). \quad (18)$$

A straightforward computation using (13)–(18), yields

$$\begin{aligned}\mathbb{E}^{U^*,Q}[X_t^{\mathsf{T}}\Pi_tX_t] &= \mathbb{E}^{U^*,Q}[\hat{X}_t^{\mathsf{T}}\Pi_t\hat{X}_t] + \mathbb{E}^{U^*,Q}[(X_t-\hat{X}_t)^{\mathsf{T}}\Pi_t(X_t-\hat{X}_t)]\\ &= \mathbb{E}^{U^*,Q}[\hat{X}_{t-1}^{\mathsf{T}}(\Pi_{t-1}-R - K_{t-1}^{\mathsf{T}}SK_{t-1})\hat{X}_{t-1}]\\ &\quad + \mathbb{E}^{U^*,Q}[\mathrm{tr}(\Pi_t\hat{\Pi}_t) + \mathrm{tr}(\Pi_t\hat{\tilde{\Pi}}_t)] \end{aligned} \quad (19)$$

where

$$\hat{\tilde{\Pi}}_{t+1} \triangleq \Xi(\hat{\Pi}_t)C_{Q_t}^{\mathsf{T}}(C_{Q_t}\Xi(\hat{\Pi}_t)C_{Q_t}^{\mathsf{T}} + FF^{\mathsf{T}})^{-1}C_{Q_t}\Xi(\hat{\Pi}_t).$$

Similarly, for $t=0,\ldots,N-1$, we have

$$\mathbb{E}^{U^*,Q}[r(X_t,U_t^*)] = \mathbb{E}^{U^*,Q}\big[\hat{X}_t^{\mathsf{T}}(R+K_t^{\mathsf{T}}SK_t)\hat{X}_t\big] + \mathbb{E}^{U^*,Q}\big[\mathrm{tr}(R\hat{\Pi}_t)\big]. \quad (20)$$

Thus, by (19)–(20), for $t=0,\ldots,N-1$

$$\begin{aligned}&\mathbb{E}^{U^*,Q}[r(X_t,U_t^*) + X_{t+1}^{\mathsf{T}}\Pi_{t+1}X_{t+1}]\\ &\quad = \mathbb{E}^{U^*,Q}[\hat{X}_t^{\mathsf{T}}\Pi_t\hat{X}_t + \mathrm{tr}(\Pi_{t+1}\hat{\Pi}_{t+1}) + \mathrm{tr}(\Pi_{t+1}\hat{\tilde{\Pi}}_{t+1})]\\ &\quad + \mathbb{E}^{U^*,Q}[\mathrm{tr}(R\hat{\Pi}_t)]. \end{aligned} \quad (21)$$

Since

$$\mathrm{tr}(R\hat{\Pi}_t) = \mathrm{tr}(\tilde{\Pi}_t\hat{\Pi}_t) - \mathrm{tr}(A^{\mathsf{T}}\Pi_{t+1}A\hat{\Pi}_t) + \mathrm{tr}(\Pi_t\hat{\Pi}_t),$$
$$\text{and } \mathrm{tr}(\Pi_t\hat{\Pi}_t) + \mathrm{tr}(\Pi_t\hat{\tilde{\Pi}}_t) = \mathrm{tr}(\Pi_tDD^{\mathsf{T}}) + \mathrm{tr}(\Pi_tA\hat{\Pi}_{t-1}A^{\mathsf{T}})$$

we have

$$\mathrm{tr}(R\hat{\Pi}_{t-1}) + \mathrm{tr}(\Pi_t\hat{\Pi}_t) + \mathrm{tr}(\Pi_t\hat{\tilde{\Pi}}_t) = tr(\tilde{\Pi}_{t-1}\hat{\Pi}_{t-1}) + tr(\Pi_{t-1}\hat{\Pi}_{t-1}) + tr(\Pi_tDD^{\mathsf{T}}). \quad (22)$$

Define

$$J_{t,N}^v \triangleq \mathbb{E}^v\left[\sum_{k=t}^{N-1}\big(c(Q_k) + r(X_k,U_k)\big) + X_N^{\mathsf{T}}\Pi_NX_N\right].$$

Simple induction using (21) and (22) yields, for $t=0,\ldots,N$

$$J_{t,N}^{U^*,Q} = \mathbb{E}^{U^*,Q}[\hat{X}_t^{\mathsf{T}}\Pi_t\hat{X}_t + tr(\Pi_t\hat{\Pi}_t)] + \mathbb{E}^{U^*,Q}\left[\sum_{k=t}^{N-1}\big(c(Q_k) + tr(\tilde{\Pi}_k\hat{\Pi}_k) + tr(\Pi_{k+1}DD^{\mathsf{T}})\big)\right].$$

Therefore, $J_N^{U^*,Q} = \tilde{J}_N^{U^*} + \hat{J}_N^{U^*,Q}$, where $\tilde{J}_N^{U^*}$ does not depend on Q, and

$$\hat{J}_N^{U^*,Q} \triangleq \mathbb{E}^{U^*,Q}\left[\sum_{k=0}^{N-1}\big(c(Q_k) + tr(\tilde{\Pi}_k\hat{\Pi}_k)\big)\right]. \quad (23)$$

If we define $f_t(\hat{\Pi})$ as the cost-to-go function for (23), then the optimal policy Q^* can be determined by (14) via the dynamic programming principle. ■

As in the standard theory of LQG control with partial observations, the optimal control of (9)–(10) is a *certainty equivalence* control, namely, the optimization problem can be separated into two stages: first, the optimal control U_t^* is the linear feedback control in (12) whose gain does not depend on the choice of the query policy $\{Q_t\}$; second, the conditional distribution of the system state is obtained recursively via the filtering (15) which is coupled with the dynamic programming (14) to determine the optimal query policy. The difference from the standard LQG problem is that the dynamic programming equation depends on the controller gain which evolves according to (13). Thus, (14) can be viewed as the solution of an optimal estimation problem, in which the cost function is the sum of the cost of the query and a weighted estimation error (23).

C. Stabilization

Stability considerations are important in the analysis of optimal control over the infinite horizon. The study of reachability and stabilization of switched linear systems has attracted considerable interest recently [26]–[31]. Necessary and sufficient conditions for stabilizability for the continuous-time counterpart of (24) are obtained in [27], [32]. Switched discrete-time linear systems are studied in [28], [31], [33] under different scenarios. We start with the following definition.

Definition 3.1: The stochastic system (9)–(10) is *stabilizable in the mean with bounded second moment (SMBSM)*, if there exist an admissible policy $v \in \mathcal{V}$, such that

$$\mathbb{E}^v_{X_0}[X_t] \xrightarrow[t\to\infty]{} 0\,, \quad \text{and} \quad \sup_{t\ge 0}\ \mathbb{E}^v_{X_0}\|X_t\|^2 < \infty$$

for any initial condition $X_0 \sim \mathcal{N}(\bar{x}_0, \Sigma_0)$. A policy v having this property is called *stable*.

We begin by discussing the deterministic system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t)\\ y(t+1) &= C_{q(t)}x(t+1) \end{aligned} \tag{24}$$

whose state, observation and controls live in the same Euclidean spaces as (9)–(10), and the pair $(u(t), q(t))$ is chosen as a function of $\{y(1), \dots, y(t)\}$. Let

$$\bar{C} \triangleq \left[C_1^{\mathsf{T}} \mid C_2^{\mathsf{T}} \mid \cdots \mid C_\ell^{\mathsf{T}}\right]^{\mathsf{T}}.$$

Then, a necessary condition for the existence of a control $\{u(t), q(t)\}$ such that the closed loop system is asymptotically stable to the origin is that the pair (A, B) be stabilizable and the pair $(\bar{C}, A)$ be detectable. This condition is also sufficient, as shown in the following theorem, whose proof is contained in Appendix A.

Theorem 3.2: Suppose (A, B) is stabilizable and $(\bar{C}, A)$ is detectable and $K \in \mathbb{R}^{N_u \times N_x}$ is such that the matrix $A - BK$ is stable, i.e., has its eigenvalues in the open unit disc of the complex plane. Then, there exist a collection of matrices $\{L_q\,,\ q \in \boldsymbol{Q}\}$, and a sequence $\{q(0), q(1), \dots\}$ such that the controlled system (24), under the dynamic feedback control $u(t) = -K\hat{x}(t)$, with

$$\begin{aligned} \hat{x}(t+1) = (A - BK)\hat{x}(t) + L_{q(t)}\big[&y(t+1)\\ &-C_{q(t)}(A - BK)\hat{x}(t)\big] \end{aligned}$$

is uniformly geometrically stable to the origin.

Theorem 3.2 can be applied to characterize the stabilizability of (9)–(10).

Corollary 3.3: The stochastic linear system (9)–(10) is SMBSM if and only if (A, B) is stabilizable and $(\bar{C}, A)$ is detectable.

Proof: Consider dynamic output feedback of the form:

$$\begin{aligned} Z_{t+1} &= (A - L_{Q_{t-1}}C_{Q_{t-1}})Z_t + L_{Q_{t-1}}Y_t + BU_t\\ U_t &= -KZ_t \end{aligned} \tag{25}$$

with $Z_0 = 0$, and let $\tilde{Z} \triangleq X_t - Z_t$. Then, by (9)–(10) and (25), we obtain

$$\begin{aligned} X_{t+1} &= (A - BK)X_t + BK\tilde{Z}_t + DW_t\\ \tilde{Z}_{t+1} &= (A - L_{Q_{t-1}}C_{Q_{t-1}})\tilde{Z}_t + (D - L_{Q_{t-1}}F)W_t\,. \end{aligned}$$

By Theorem 3.2, provided (A, B) is observable and $(\bar{C}, A)$ is stabilizable, there exist gain matrices K, and $\{L_q\,,\ q \in \boldsymbol{Q}\}$, and a periodic sequence $\{Q_t\}$, such that under this policy (i.e., with $U_t = -KZ_t$), which is denoted by $v_s \in \mathcal{V}$, we have $\mathbb{E}^{v_s}[X_t] \xrightarrow[t\to\infty]{} 0$, while $\mathbb{E}^{v_s}\|X_t\|^2$ is bounded. Furthermore, since by the proof of Theorem 3.2, the product $\prod_{t=0}^n (A - L_{Q_t}C_{Q_t})$ decays geometrically in norm, there exist constants $\gamma_s \in (0, 1)$ and $M_s > 0$, such that

$$\begin{aligned} \|\mathbb{E}^{v_s}X_t\|^2 &\le M_s\gamma_s^t\,\|\bar{x}_0\|^2\,, &&\forall t \ge 0\\ \mathbb{E}^{v_s}\|X_t\|^2 &\le M_s\left[\gamma_s^t\,\mathrm{tr}(\Sigma_0) + \frac{1}{2}\right], &&\forall t \ge 0\,. \end{aligned} \tag{26}$$

This completes the proof. ∎

Remark 3.1: Note that under the policy v_s in the proof of Corollary 3.3, $\mathbb{E}^{v_s}\|U_t\|^2$ remains bounded, and redefining M_s as the largest of the two bounds, in addition to (26), we can assert that

$$\mathbb{E}^{v_s}\|U_t\|^2 \le M_s\,, \qquad \forall t \ge 0\,. \tag{27}$$

Remark 3.2: The control U_t used in (25) is $\mathcal{Y}^{t-1}$-adapted, whereas the admissible controls for (9)–(10) were defined as $\mathcal{Y}^t$-adapted. However, there is no discrepancy: on the one hand, sufficiency is not affected, while on the other (A, B) observable and $(\bar{C}, A)$ stabilizable is necessary for (9)–(10) to be SMBMS.

D. Optimal Control Over the Infinite Horizon

In this section we study the optimal control problem over the infinite horizon. We are particularly interested in the ergodic control problem, and we approach this via the β-discounted one.

Let $\beta \in (0, 1)$ be the discount factor. For a policy $v \in \mathcal{V}$, define

$$J^v_\beta(\bar{x}_0, \Sigma_0) \triangleq \mathbb{E}^v_{X_0}\left[\sum_{t=0}^\infty \beta^t\big(c(Q_t) + r(X_t, U_t)\big)\right]$$

and let $J^*_\beta \triangleq \inf_{v\in\mathcal{V}}\ J^v_\beta$.

Provided (A, B) is stabilizable, and $(\bar{C}, A)$ is detectable, J^*_β is finite. Indeed, since $J^*_\beta \le J^{v_s}_\beta$, with $v_s \in \mathcal{V}$ the policy in (26), an easy calculation shows that there exists a constant $\tilde{M}$ such that

$$\begin{aligned} J^*_\beta(\bar{x}_0, \Sigma_0) &\le \tilde{M}\left(\frac{\|\bar{x}_0\|^2 + \mathrm{tr}(\Sigma_0)}{1 - \beta\gamma_s} + (1 - \beta)^{-1}\right)\\ &\le \tilde{M}\left(\frac{1}{1 - \beta} + \frac{\|\bar{x}_0\|^2 + \mathrm{tr}(\Sigma_0)}{1 - \gamma_s}\right). \end{aligned} \tag{28}$$

The existence and characterization of stationary optimal policies for the β-discounted control problem is the topic of the following theorem.

Theorem 3.4: For the control system (9)–(10), assume that (A,B) is stabilizable, and $(\bar{C},A)$ is detectable. Then there exists a unique positive definite solution Π^*_β to the algebraic Riccati equation

$$\Pi^*_\beta = R + \beta A^\mathsf{T}\Pi^*_\beta A - \beta^2 A^\mathsf{T}\Pi^*_\beta B(S+\beta B^\mathsf{T}\Pi^*_\beta B)^{-1}B^\mathsf{T}\Pi^*_\beta A. \tag{29}$$

Define the functional map S_β, $\beta\in(0,1]$, by

$$S_\beta(f)(\hat{\Pi}) \triangleq \min_q \left\{c(q)+\mathrm{tr}(\tilde{\Pi}^*_\beta\hat{\Pi})+\beta f(\mathcal{T}_q(\hat{\Pi}))\right\} \tag{30}$$

where $\tilde{\Pi}^*_\beta \triangleq R-\Pi^*_\beta+\beta A^\mathsf{T}\Pi^*_\beta A$. Let

$$U^*_t = -(S+\beta B^\mathsf{T}\Pi^*_\beta B)^{-1}\beta B^\mathsf{T}\Pi^*_\beta A\hat{X}_t\,,$$
$$\text{where}\quad \hat{X}_{t+1} = A\hat{X}_t + BU_t + \hat{K}_{q^*(\hat{\Pi}_t)}(\hat{\Pi}_t)\big(Y_{t+1} - C_{q^*(\hat{\Pi}_t)}(A\hat{X}_t+BU_t)\big)$$

with

$$\hat{\Pi}_{t+1} = \mathcal{T}_{q^*(\hat{\Pi}_t)}(\hat{\Pi}_t)\,,\qquad \hat{\Pi}_0=\Sigma_0\,.$$

Let

$$\tilde{J}^{U^*}_\beta(\bar{x}_0,\Sigma_0) \triangleq \bar{x}_0^\mathsf{T}\Pi^*_\beta\bar{x}_0 + \mathrm{tr}(\Pi^*_\beta\Sigma_0) + \beta(1-\beta)^{-1}\mathrm{tr}(\Pi^*_\beta DD^\mathsf{T})\,. \tag{31}$$

There exists a lower semicontinuous $f^*_\beta:\mathcal{M}^+_0\to\mathbb{R}_+$ satisfying

$$f^*_\beta(\hat{\Pi}) = S_\beta(f^*_\beta)(\hat{\Pi}) \tag{32}$$

such that if $q^*_\beta:\mathcal{M}^+_0\to\boldsymbol{Q}$ is a selector of the minimizer in (30), with $f=f^*_\beta$, then $v^*_\beta=(\{U^*_t\},q^*_\beta)$ is optimal for the discounted control problem, and for each $\beta\in(0,1)$, the optimal discounted cost is given by

$$J^*_\beta(\bar{x}_0,\Sigma_0) = \tilde{J}^{U^*}_\beta(\bar{x}_0,\Sigma_0) + f^*_\beta(\Sigma_0)\,. \tag{33}$$

Proof: It is well known that, provided (A,B) is stabilizable, the matrix recursive iteration (13) for Π_t converges to a positive definite matrix Π^*_β satisfying (29). Moreover,(29) has a unique solution in $\mathcal{M}^+$. Consider the finite-horizon problem with initial condition $X_0\sim\mathcal{N}(\bar{x}_0,\Sigma_0)$

$$J^v_k(\bar{x}_0,\Sigma_0) \triangleq \mathbb{E}^v_{X_0}\left[\sum_{t=0}^{k-1}\beta^t\big(c(Q_t)+r(X_t,U_t)\big)+\beta^k X_k^\mathsf{T}\Pi^*_\beta X_k\right],\ k\in\mathbb{N}. \tag{34}$$

It follows by Section III-B that the optimal cost J^*_k is given by

$$J^*_k(\bar{x}_0,\Sigma_0) = \bar{x}_0^\mathsf{T}\Pi^*_\beta\bar{x}_0+\mathrm{tr}(\Pi^*_\beta\Sigma_0) + \sum_{t=1}^{k}\beta^t\mathrm{tr}(\Pi^*_\beta DD^\mathsf{T}) + f^{(k)}_0(\Sigma_0)\,,$$

where $f^{(k)}_0:\mathcal{M}^+_0\to\mathbb{R}$ satisfies

$$f^{(k+1)}_0(\hat{\Pi}) = \min_q\left\{c(q)+tr(\tilde{\Pi}^*_\beta\hat{\Pi})+\beta f^{(k)}_0(\mathcal{T}_q(\hat{\Pi}))\right\} \tag{35}$$

with $f^{(0)}_0=0$. Since $J^*_k\le J^{v_s}_k$, where v_s is the policy in Theorem 3.2, it follows that $\{f^{(k)}_0\}$ is bounded pointwise in $\mathcal{M}^+_0$. Since, in addition, $f^{(k)}_0\uparrow$, it converges to a lower semicontinuous function f^*_β, and taking monotone limits, (35) yields(32). Since f^*_β is locally bounded, it follows by (20) and (34) that $\beta^t\mathbb{E}^{q^*_\beta}_{X_0}[\mathrm{tr}(\hat{\Pi}_t)]\to 0$ as $t\to\infty$. Thus, the estimate in (28) yields $\beta^t\mathbb{E}^{q^*_\beta}_{X_0}[f^*_\beta(\hat{\Pi}_t)]\to 0$, as $t\to\infty$. Using the dynamic programming (32), we have

$$f^*_\beta(\Sigma_0) = \mathbb{E}^{q^*_\beta}_{X_0}\left[\sum_{k=0}^{t-1}\beta^k\big(c(Q_k)+\mathrm{tr}(\tilde{\Pi}^*_\beta\hat{\Pi}_k)\big)\right] + \beta^t\mathbb{E}^{q^*_\beta}_{X_0}[f^*_\beta(\hat{\Pi}_t)]$$

and taking limits as $t\to\infty$, we obtain

$$f^*_\beta(\Sigma_0) = \mathbb{E}^{q^*_\beta}_{X_0}\left[\sum_{k=0}^{\infty}\beta^k\big(c(Q_k)+\mathrm{tr}(\tilde{\Pi}^*_\beta\hat{\Pi}_k)\big)\right].$$

One more application of (32) shows that for all $v\in\mathcal{V}$, such that $J^v_\beta(\bar{x}_0,\Sigma_0)<\infty$

$$f^*_\beta(\Sigma_0) \le \mathbb{E}^{v}_{X_0}\left[\sum_{k=0}^{\infty}\beta^k\big(c(Q_k)+\mathrm{tr}(\tilde{\Pi}^*_\beta\hat{\Pi}_k)\big)\right]$$

and thus, q^*_β is optimal. The proof is complete. ■

Before proceeding to the analysis of the ergodic control problem, we establish some useful properties of the Riccati map $\mathcal{T}_q$ defined in (17). For $\varepsilon>0$, let

$$\mathcal{M}^+_\varepsilon \triangleq \{\hat{\Pi}\in\mathcal{M}^+ : \min\{\lambda\mid\lambda\in\sigma(\hat{\Pi})\}>\varepsilon\}$$

and "$\circ$" denote composition of functions. To prove the existence of a stationary optimal policy for the ergodic control problem, we employ Lemmas 3.5–3.6 below, whose proofs are in Appendices B and C.

Lemma 3.5: There exists $\varepsilon>0$ and $\kappa\in\mathbb{N}$, such that

$$\mathcal{T}_{q_{\kappa-1}}\circ\mathcal{T}_{q_{\kappa-2}}\circ\cdots\circ\mathcal{T}_{q_0}(0)\in\mathcal{M}^+_\varepsilon$$

for every sequence $\{q_0,q_1,\ldots,q_{\kappa-1}\}$, if and only if the pair (A,D) is controllable, in which case $\kappa\le N_x$.

Lemma 3.6: The functions $\mathcal{T}_q:\mathcal{M}^+_0\to\mathcal{M}^+_0$ and $f^*_\beta:\mathcal{M}^+_0\to\mathbb{R}_+$ are concave.

To characterize the ergodic control problem, we adopt the vanishing discount method, i.e., an asymptotic analysis as the discount factor $\beta\to 1$. By (31)–(33), for any $\bar{x}_1,\bar{x}_2$ in $\mathbb{R}^{N_x}$ and Σ_1, Σ_2 in $\mathcal{M}^+_0$

$$f^*_\beta(\Sigma_1)-f^*_\beta(\Sigma_2) = J^*_\beta(\bar{x}_1,\Sigma_1)-J^*_\beta(\bar{x}_2,\Sigma_2) - (\bar{x}_1^\mathsf{T}\Pi^*_\beta\bar{x}_1-\bar{x}_2^\mathsf{T}\Pi^*_\beta\bar{x}_2) - \mathrm{tr}\big(\Pi^*_\beta(\Sigma_1-\Sigma_2)\big)\,. \tag{36}$$

Also using

$$J^*_\beta(\bar{x}_0,\Sigma_0) \le \mathbb{E}^{v_s}_{X_0}\left[\sum_{t=0}^{k-1}\beta^t\big(c(Q_t)+r(X_t,U_t)\big)\right] + \beta^k\mathbb{E}^{v_s}_{X_0}[J^*_\beta(\hat{X}_k,\hat{\Pi}_k)]$$

we obtain, that for some constant $\tilde{M}' > 0$, and M_s the constant in (26)–(27)

$$J^*_\beta(\bar{x}_0, \Sigma_0) \le \tilde{M}'\big(1 + \|x_0\|^2 + \operatorname{tr}(\Sigma_0)\big) + \sup_{\substack{\|x\|^2 \le M_s \\ \operatorname{tr}(\Sigma) \le M_s}} J^*_\beta(\bar{x}, \Sigma). \quad (37)$$

Let $\mathcal{B}_s \subset \mathbb{R}^{N_x} \times \mathcal{M}_0^+$ be a bounded ball in $\mathcal{M}_0^+$ containing the set $\{\Sigma : \operatorname{tr}(\Sigma) \le M_s\}$, and such that

$$\mathcal{T}_{q_\kappa} \circ \mathcal{T}_{q_{\kappa-1}} \circ \cdots \circ \mathcal{T}_{q_0}(0) \in \mathcal{B}_s, \quad \forall \{q_0, q_1, \dots, q_\kappa\} \subset \boldsymbol{Q}^{\kappa+1}.$$

Since f^*_β depends only on Σ, and since Π^*_β converges to a limit in $\mathcal{M}^+$, as $\beta \to 1$, it follows from (36) and (37) that there exists a continuous function $G_s : \mathbb{R}_+ \to \mathbb{R}_+$, having affine growth, such that

$$f^*_\beta(\Sigma) - \sup_{\Sigma' \in \mathcal{B}_s} \{f^*_\beta(\Sigma')\} \le G_s(\operatorname{tr}(\Sigma)) \quad (38)$$

for all $\Sigma \in \mathcal{M}_0^+$. Define

$$\bar{f}_\beta \triangleq f^*_\beta(\Sigma) - \inf_{\Sigma' \in \mathcal{M}_0^+} f^*_\beta(\Sigma')$$
$$\operatorname{span}_{\mathcal{B}_s}(f^*_\beta) \triangleq \sup_{\Sigma \in \mathcal{B}_s} f^*_\beta(\Sigma) - \inf_{\Sigma \in \mathcal{B}_s} f^*_\beta(\Sigma).$$

Equicontinuity of the differential discounted value function $\bar{f}_\beta$, is established in the following lemma, whose proof is contained in Appendix D.

Lemma 3.7: Under the assumptions of Theorem 3.4

i) $\inf_{\Sigma \in \mathcal{M}_0^+} f^*_\beta(\Sigma) = f^*_\beta(0)$, for any $\beta \le 1$.
ii) Suppose (A, D) is a controllable pair. Then, $\bar{f}_\beta$ is locally bounded, uniformly in $\beta \in (0, 1)$.
iii) Provided (A, D) is a controllable pair, $\{\bar{f}_\beta,\ 0 < \beta < 1\}$ is equicontinuous on compact subsets of $\mathcal{M}_0^+$.

We now turn to the ergodic control problem. For a policy $v \in \mathcal{V}$, define

$$J^v \triangleq \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}^v_{X_0} \sum_{t=0}^{N-1} \big(c(Q_t) + X_t^{\mathsf{T}} R X_t + U_t^{\mathsf{T}} S U_t\big)$$

and let $J^* = \inf_{v \in \mathcal{V}} J^v$. The main result of this section is the following.

Theorem 3.8 (Ergodic Control): Assume that (A, B) is stabilizable, $(\bar{C}, A)$ is detectable, and (A, D) is controllable. Define the functional map S, by

$$S(h)(\hat{\Pi}) \triangleq \min_q \left\{c(q) + \operatorname{tr}(\tilde{\Pi}^* \hat{\Pi}) + h(\mathcal{T}_q(\hat{\Pi}))\right\} \quad (39)$$

where $\tilde{\Pi}^* \triangleq R - \Pi^* + A^{\mathsf{T}}\Pi^* A$, and $\Pi^* \in \mathcal{M}^+$ solves

$$\Pi^* = A^{\mathsf{T}}\Pi^* A + R - A^{\mathsf{T}}\Pi^* B(S + B^{\mathsf{T}}\Pi^* B)^{-1} B^{\mathsf{T}}\Pi^* A. \quad (40)$$

There exists a nonnegative constant ϱ^* and a continuous $h : \mathcal{M}_0^+ \to \mathbb{R}_+$ satisfying

$$h(\hat{\Pi}) + \varrho^* = S(h)(\hat{\Pi}). \quad (41)$$

Let $q^* : \mathcal{M}_0^+ \to \boldsymbol{Q}$ be a selector of the minimizer in (39). Set

$$U^*_t = -(S + B^{\mathsf{T}}\Pi^* B)^{-1} B^{\mathsf{T}}\Pi^* A \hat{X}_t \quad (42)$$

where

$$\hat{X}_{t+1} = A\hat{X}_t + BU_t + \hat{K}_{q^*(\hat{\Pi}_t)}(\hat{\Pi}_t)\big(Y_{t+1} - C_{q^*(\hat{\Pi}_t)}(A\hat{X}_t + BU_t)\big)$$

with $\hat{K}_q$ as in (16a), and

$$\hat{\Pi}_{t+1} = \mathcal{T}_{q^*(\hat{\Pi}_t)}(\hat{\Pi}_t), \qquad \hat{\Pi}_0 = \Sigma_0. \quad (43)$$

Then, $v^* = (\{U^*_t\}, q^*)$ is optimal for the ergodic control problem, and

$$J^* = \operatorname{tr}(\Pi^* D D^{\mathsf{T}}) + \varrho^*.$$

Furthermore, v^* is stable.

Proof: It is well known that, provided (A, B) is stabilizable, Π^*_β converges as $\beta \to 1$ to $\Pi^* \in \mathcal{M}^+$, which is the unique positive definite solution to the algebraic Riccati (40). Thus it suffices to turn our attention to the query policy. By Lemma 3.7, $\{\bar{f}_\beta\}$ is locally equicontinuous and bounded, and thus along some sequence $\beta_k \to 0$, $\bar{f}_{\beta_k}$ converges to some continuous function h, while at the same time $(1 - \beta_k) f_{\beta_k}(0)$ converges to some constant ϱ^*. Taking limits in (32), we obtain (41).

By (41), there exists $M_0 > 0$ such that $\operatorname{tr}(\Sigma) > M_0$ implies

$$h(\mathcal{T}_{q^*(\Sigma)}(\Sigma)) - h(\Sigma) < -1, \qquad \forall q \in \boldsymbol{Q}.$$

This shows that $\sup_{t \ge 0} \mathbb{E}^{q^*}_{X_0}[\hat{\Pi}_t] < \infty$, for all X_0. Let

$$K^* \triangleq (S + B^{\mathsf{T}}\Pi^* B)^{-1} B^{\mathsf{T}}\Pi^* A.$$

Since

$$(A - BK^*)^{\mathsf{T}}\Pi^*(A - BK^*) - \Pi^* = -R - (K^*)^{\mathsf{T}} S K^*$$

it follows that $(A - BK^*)$ is a stable matrix. Thus, from

$$X_{t+1} = (A - BK^*)X_t + BK^*(X_t - \hat{X}_t) + DW_t$$

we obtain

$$\mathbb{E}^{q^*}_{X_0}[X_t] \xrightarrow[t \to \infty]{} 0, \quad \text{and} \quad \sup_{t \ge 0} \mathbb{E}^{q^*}_{X_0} \|X_t\|^2 < \infty.$$

By (41), for any admissible $\{Q_t\}$

$$\varrho^* + \frac{h(\Sigma_0) - h(\hat{\Pi}_N)}{N} \le \frac{1}{N} \sum_{t=0}^{N-1} \big(c(Q_t) + \operatorname{tr}(\tilde{\Pi}^* \hat{\Pi}_t)\big) \quad (44)$$

with equality when $Q_t = q^*$. Since the function G_s in (38) has affine growth, it follows that $h(\Sigma) \le G_s^*(\mathrm{tr}(\Sigma))$, for some affine function G_s^*. Therefore, since q^* is stable, $h(\hat{\Pi}_N)/N \to 0$, as $N \to \infty$, which in turn implies by (44) that

$$\varrho^* = \limsup_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} \left(c(Q_t) + \mathrm{tr}(\tilde{\Pi}^* \hat{\Pi}_t)\right), \quad \mathbb{P}_{X_0}^{q^*} - \text{a.s.}.$$

Also for any policy $v \in \mathcal{V}$ such that the limit supremum of the expectation of the right hand side of (44) is finite, we have $\mathbb{E}_{X_0}^v[h(\hat{\Pi}_{N_k})]/N_k \to 0$, along some subsequence $N_k \to \infty$. Thus

$$\liminf_{n\to\infty} \frac{h(\hat{\Pi}_n)}{n} = 0, \quad \mathbb{P}_{X_0}^v - \text{a.s.} \tag{45}$$

By (44)–(45),

$$\varrho^* \le \limsup_{k\to\infty} \frac{1}{N} \mathbb{E}_{X_0}^v \left[\sum_{t=0}^{N-1} \left(c(Q_t) + \mathrm{tr}(\tilde{\Pi}^* \hat{\Pi}_t)\right) \right].$$

Hence q^* is optimal. This completes the proof. ∎

Remark 3.3: The assumption (A, D) controllable cannot be relaxed in general. Lack of this assumption may result in the long-run optimal cost to depend on the initial condition Σ_0. A possible weakening is to require that (A, D) is stabilizable, but we do not pursue this analysis here.

Remark 3.4: In summary, the steps to compute the optimal controller are as follows: First we solve the Riccati (40) for $\Pi^* \in \mathcal{M}^+$. The optimal control is the linear feedback controller in (42) with a constant gain. Next, we solve the HJB (41) to obtain a stationary optimal policy q^* for the query. The optimal query is a function of $\hat{\Pi}_t$, and the state estimates are updated according to (43).

IV. Example: Optimal Switching Estimation

Since the switching control for the observation is the key feature of the problem, the examples presented in this section concentrate on the optimal estimation problem. In other words, the objective is to estimate the system state X_t while minimizing the infinite-horizon criteria with respect to the running cost $\tilde{g}(\hat{\Pi}, q) = c(q) + \mathrm{tr}(\hat{\Pi})$. We present examples of 1–D and 2–D systems with a binary query variable, i.e., $\boldsymbol{Q} = \{1, 2\}$.

A. 1-D Case

Consider a 1-D system as in (9)–(10), with $C_q \ne 0$, $q \in \{1, 2\}$. If we let $V_q \triangleq F/C_q$, i.e., the normalized noise, $\mathcal{T}_q$ takes the form

$$\mathcal{T}_q(\hat{\Pi}) = A^2\hat{\Pi} + D^2 - \frac{(A^2\hat{\Pi} + D^2)^2}{A^2\hat{\Pi} + D^2 + V_q^2}$$

and the HJB equations for the discounted and ergodic criteria take the form

$$f_\beta^*(\hat{\Pi}) = \min_q \left\{ c(q) + \hat{\Pi} + \beta f_\beta^*(\mathcal{T}_q(\hat{\Pi})) \right\} \tag{46a}$$

$$\varrho + f^*(\hat{\Pi}) = \min_q \left\{ c(q) + \hat{\Pi} + f^*(\mathcal{T}_q(\hat{\Pi})) \right\}. \tag{46b}$$

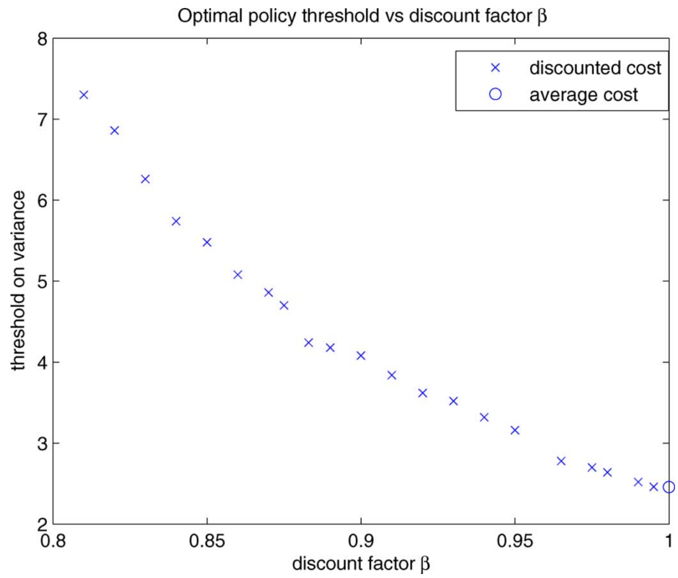

Fig. 1. Optimal policy threshold versus the discount factor β.

Suppose $V_1 > V_2$ and that the cost of observation satisfies $c(1) < c(2)$. In other words, Sensor 1 has a lower sensing capability and lower cost, while Sensor 2 has a higher sensing capability and cost.

Let $\hat{\Pi}_1^*$, $\hat{\Pi}_2^*$ denote the unique fixed points of $\mathcal{T}_1$, $\mathcal{T}_2$, respectively. Since $V_1 > V_2$, we have $\hat{\Pi}_1^* > \hat{\Pi}_2^*$. The iterates of the map $\mathcal{T}_q$, converge to $\hat{\Pi}_q^*$, hence we restrict our attention to the set of initial conditions $[\hat{\Pi}_2^*, \hat{\Pi}_1^*]$, which is invariant under $\mathcal{T}_q$, $q \in \{1, 2\}$. For $\hat{\Pi} \in (\hat{\Pi}_2^*, \hat{\Pi}_1^*)$, $T_2(\hat{\Pi}) < \hat{\Pi} < T_1(\hat{\Pi})$.

Using the method of successive iterates of the dynamic programming operator, we can derive sharp conditions for the optimal query policy to be switching between the two sensors, and not to be a constant. This is summarized in the following proposition, whose proof is omitted.

Proposition 4.1: Let q_β^* denote the minimizer in (46a), for $\beta \in (0, 1)$, and the minimizer in (46b), when $\beta = 1$. Then, given $\beta \in (0, 1]$, there exists $\delta > 0$ such that

i) $q_\beta^*(\hat{\Pi}) = 1$, for $\hat{\Pi} \in [\hat{\Pi}_2^*, \hat{\Pi}_2^* + \delta]$, if and only if

$$c(2) - c(1) > \sum_{k=0}^{\infty} \beta^k \left(\mathcal{T}_2^k \circ \mathcal{T}_1(\hat{\Pi}_2^*) - \hat{\Pi}_2^*\right).$$

ii) $q_\beta^*(\hat{\Pi}) = 2$, for $\hat{\Pi} \in [\hat{\Pi}_1^* - \delta, \hat{\Pi}_1^*]$, if and only if

$$c(2) - c(1) < \sum_{k=0}^{\infty} \beta^k \left(\hat{\Pi}_1^* - \mathcal{T}_1^k \circ \mathcal{T}_2(\hat{\Pi}_1^*)\right).$$

The optimal query policy for the 1-D example can be easily obtained numerically by standard algorithms, like value iteration or policy iteration. After running numerous simulations, it appears that the optimal query policy for both the discounted and average costs is a threshold policy, namely, the optimal q_β^* takes the form

$$q_\beta^* = \begin{cases} 1, & \hat{\Pi} < \hat{\Pi}_\beta^* \\ 2, & \hat{\Pi} \ge \hat{\Pi}_\beta^*. \end{cases}$$

The threshold point $\hat{\Pi}_\beta^*$, as a function of the discount factor β, is displayed in Fig. 1. As β approaches 1, the optimal threshold

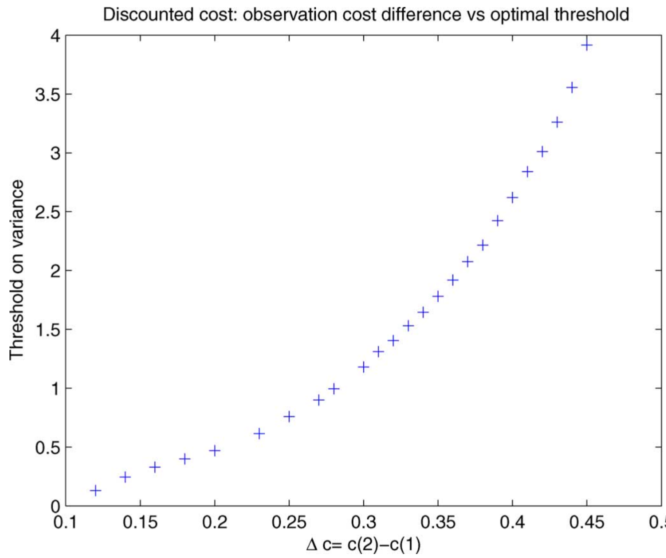

Fig. 2. Cost difference versus the optimal threshold.

for the discounted cost converges to that of the average cost. Furthermore, the optimal threshold is a decreasing function of β. This agrees with Proposition 4.1, and also agrees with intuition that as the future is weighted more in the criterion, the frequency with which the optimal policy chooses the more accurate and costly observation increases.

Fig. 2, shows the variation of the optimal threshold as a function of the cost differential. The threshold point is an increasing function of the cost differential and once the latter increases in value beyond 0.45 the optimal policy is a constant, and the controller chooses to use the least costly observation all the time.

B. 2-D Case

We present an example of a 2-D system with system state $X = [X^1, X^2]^\mathsf{T}$, a scalar observation, and the following parameters:

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \qquad DD^\mathsf{T} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}$$
$$F^2 = 0.2 \qquad C_1 = [1 \quad 0], \qquad C_2 = [0 \quad 1].$$

The running cost is $c(1) = c(2) = 0$. Since the pairs (A, C_q) are not detectable, this example can be viewed as a problem of optimal switching estimation.

Fig. 3 shows the optimal switching curve that minimizes the trace of estimation error variance, and can be interpreted as follows: when $\hat{\Pi}^1$, the estimation variance of X^1 is larger than the estimation variance $\hat{\Pi}^2$ of X^2, we query Sensor 1, and vice versa. The switching curve is a straight line due to symmetry.

Next suppose that Sensor 2 has lower observation noise and higher price, i.e.,

$$F_1^2 = 0.1\,, \quad F_2^2 = 0.2\,, \quad c(1) = 0.05\,, \quad c(2) = 0$$

while the rest of the parameters are kept the same as before. This has the following impact on the optimal switching curve, as shown in Fig. 4: Near the origin, where the penalty on the estimation errors is small, Sensor 2 is used, due to its lower operation cost; far away from the origin, where the estimation error dominates the cost of querying, the symmetry of Fig. 3 is broken, and Sensor 1 is favored.

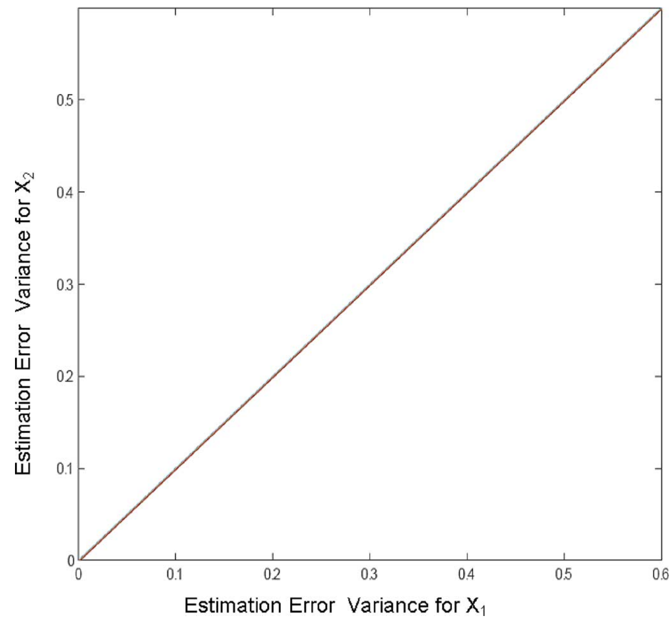

Fig. 3. Optimal switching curve for the first 2-D example.

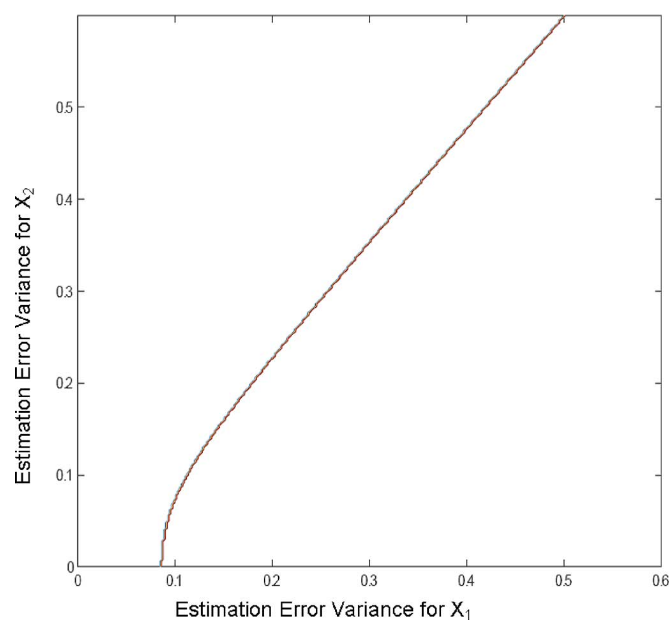

Fig. 4. Optimal switching curve for the second 2-D example.

In the third 2-D example, both sensors can fully detect the unstable eigenmode of the system state, i.e.

$$C_1 = [1.0 \;\; 1.0], \qquad C_2 = [1.1 \;\; 1.1]\,,$$
$$\text{and } F_1^2 = 0.2\,, \quad F_2^2 = 0.1\,, \qquad c(1) = 0\,, \quad c(2) = 0.05\,.$$

Fig. 5 portrays the optimal switching curve for this example. When the estimation error lies in the interior of the switching curve, Sensor 1 is queried due to its low cost. Outside the switching curve, the estimation error is large enough to necessitate querying Sensor 2, which has higher precision.

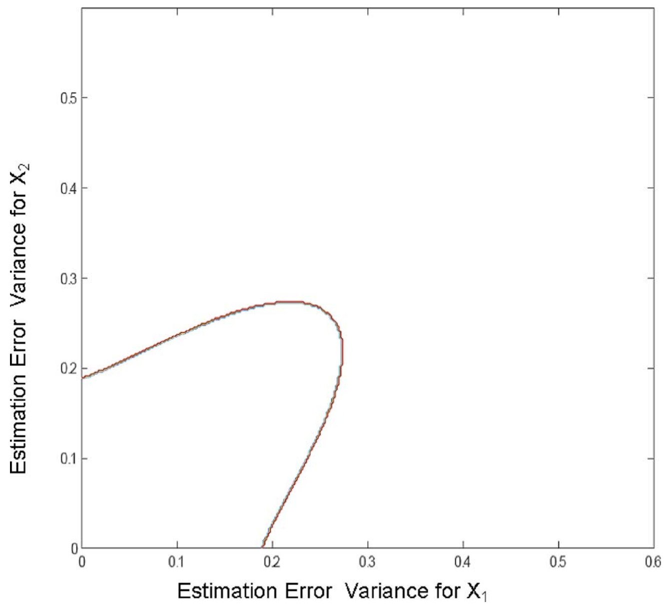

Fig. 5. Optimal switching curve for the third 2-D example.

V. Conclusion

In this paper, we described a general model for partially observable Markov processes with controlled observations and infinite-horizon cost. The existence of ergodic control in the special case of switched hierarchical observations was proven under an assumption on the finest observation space. Specializing to the LQG problem with controlled observations, we obtained existence results for the discounted and ergodic optimal control problem, under sharp conditions. The structure of optimal sensor query for 1-D LQG problem was investigated both analytically and numerically. Some simple numerical examples were presented for 2-D linear systems.

Appendix A
Proof of Theorem 3.2

It is enough to show that the system

$$z(t+1) = (A + L_{q(t)}C_{q(t)})z(t)$$

is uniformly geometrically stable to the origin. Consider first the case $\ell = 2$, that lends itself to simpler notation. Without loss of generality assume $(\bar{C}, A)$ is observable. Then, there exist row vectors k_q of dimension N_y such that with $c_q = k_q C_q$, we have

$$\begin{aligned}\mathbb{R}^{N_x} =& \mathcal{R}\big[c_1^\mathsf{T} \mid A^\mathsf{T} c_1^\mathsf{T} \mid \cdots \mid (A^{n_1-1})^\mathsf{T} c_1^\mathsf{T}\big] \\ &\oplus \mathcal{R}\big[c_2^\mathsf{T} \mid A^\mathsf{T} c_2^\mathsf{T} \mid \cdots \mid (A^{n_2-1})^\mathsf{T} c_2^\mathsf{T}\big] \end{aligned} \tag{47}$$

and $n_1 + n_2 = N_x$. With respect to the ordered basis in (47), A and c_1, c_2 take the form

$$TA^\mathsf{T}T^{-1} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_{12} \\ 0 & \tilde{A}_2 \end{bmatrix}, \qquad \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} T^{-1} = \begin{bmatrix} \tilde{c}_1 & 0 \\ 0 & \tilde{c}_2 \end{bmatrix}$$

with $\tilde{A}_{12} \in \mathbb{R}^{n_1 \times n_2}$, $\tilde{A}_q \in \mathbb{R}^{n_q \times n_q}$, $\tilde{c}_q^\mathsf{T} \in \mathbb{R}^{n_q}$, and the pair $(\tilde{c}_q, \tilde{A}_q)$ is observable, for $q \in \{1, 2\}$.

Let $\gamma > 0$ be such that

$$\max_{q \in \{1,2\}} \{|\lambda| : \lambda \in \sigma(\tilde{A}_q)\} < \gamma \tag{48}$$

where σ denotes the spectrum of the matrix, and let $\tilde{\gamma} > 0$ be defined by

$$\tilde{\gamma} \triangleq \frac{1}{2} \min\{\gamma, \gamma^{-1}\}. \tag{49}$$

Select gains $l_q \in \mathbb{R}^{n_q}$, such that

$$\max_{q \in \{1,2\}} \{|\lambda| : \lambda \in \sigma(\tilde{A}_q + l_q \tilde{c}_q)\} < \tilde{\gamma}. \tag{50}$$

Then, if we let

$$L_1 \triangleq T^{-1} \begin{bmatrix} l_1 \\ 0 \end{bmatrix} k_1, \qquad L_2 \triangleq T^{-1} \begin{bmatrix} 0 \\ l_2 \end{bmatrix} k_2,$$

and $\bar{A}_q \triangleq \tilde{A}_q + l_q \tilde{c}_q$, we obtain

$$\begin{aligned} \hat{A}_1 &\triangleq T(A + L_1 C_1)T^{-1} = \begin{bmatrix} \bar{A}_1 & \tilde{A}_{12} \\ 0 & \tilde{A}_2 \end{bmatrix}, \\ \hat{A}_2 &\triangleq T(A + L_2 C_2)T^{-1} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_{12} \\ 0 & \bar{A}_2 \end{bmatrix}. \end{aligned}$$

Expressing $x \in \mathbb{R}^n$ in block form $x = \binom{x_1}{x_2}$, with $x_q \in \mathbb{R}^{n_q}$, $q \in \{1, 2\}$, we define the block norm

$$\|x\|_* = \max_{q \in \{1,2\}} \{\|x_q\|\}.$$

By (48) and (50), there exists $M > 0$ such that for all $k \in \mathbb{N}$,

$$\left\|\tilde{A}_q^k\right\| \le M\gamma^k, \quad \left\|\bar{A}_q^k\right\| \le M\tilde{\gamma}^k, \quad q \in \{1, 2\} \tag{51}$$

and $\left\|\tilde{A}_{12}\right\| \le M$. For $k \in \mathbb{N}$ and $\bar{x} \in \mathbb{R}^n$, we have

$$\hat{A}_1^k \hat{A}_2^k \bar{x} = \begin{pmatrix} \bar{A}_1^k \tilde{A}_1^k \bar{x}_1 + \Gamma_k \bar{x}_2 \\ \tilde{A}_2^k \bar{A}_2^k \bar{x}_2 \end{pmatrix} \tag{52}$$

where

$$\Gamma_k \triangleq \sum_{i=0}^{k-1} \left(\bar{A}_1^k \tilde{A}_1^{k-1-i} \tilde{A}_{12} \bar{A}_2^i + \bar{A}_1^{k-1-i} \tilde{A}_{12} \tilde{A}_2^i \bar{A}_2^k \right).$$

Using (49) and (51), we calculate the following estimates

$$\begin{aligned} \left\|\bar{A}_1^k \tilde{A}_1^k \bar{x}_1\right\| &\le M^2 \tilde{\gamma}^k \gamma^k \|\bar{x}_1\| \le M^2 \left(\frac{1}{2}\right)^k \|\bar{x}_1\| \\ \left\|\tilde{A}_2^k \bar{A}_2^k \bar{x}_2\right\| &\le M^2 \tilde{\gamma}^k \gamma^k \|\bar{x}_2\| \le M^2 \left(\frac{1}{2}\right)^k \|\bar{x}_2\| \end{aligned} \tag{53}$$

and

$$\begin{aligned} \left\|\bar{A}_1^k \sum_{i=0}^{k-1} \tilde{A}_1^{k-1-i} \tilde{A}_{12} \bar{A}_2^i \bar{x}_2\right\| &\le M\tilde{\gamma}^k \sum_{i=0}^{k-1} M^3 \gamma^{k-1-i} \tilde{\gamma}^i \|\bar{x}_2\| \\ &\le 2M^4 \gamma^{-1} \tilde{\gamma}^k \gamma^k \|\bar{x}_2\| \\ &\le 2M^4 \gamma^{-1} \left(\frac{1}{2}\right)^k \|\bar{x}_2\|, \end{aligned} \tag{54}$$

$$\left\|\sum_{i=0}^{k-1}\bar{A}_1^{k-1-i}\tilde{A}_{12}\tilde{A}_2^i\bar{A}_2^k\bar{x}_2\right\| \leq \sum_{i=0}^{k-1}M^4\tilde{\gamma}^{k-1-i}\gamma^i\tilde{\gamma}^k\|\bar{x}_2\| \\ = \sum_{i=0}^{k-1}M^4\tilde{\gamma}^i\gamma^{k-1-i}\tilde{\gamma}^k\|\bar{x}_2\| \\ \leq 2M^4\gamma^{-1}\left(\frac{1}{2}\right)^k\|\bar{x}_2\|. \quad (55)$$

It follows by (52)–(55), that if we select

$$\hat{k} > \log_2\left(2M^2 + 8M^4\gamma^{-1}\right)$$

then

$$\left\|\hat{A}_1^{\hat{k}}\hat{A}_2^{\hat{k}}\bar{x}\right\|_* < \frac{1}{2}\|\bar{x}\|_*, \quad \forall \bar{x} \in \mathbb{R}^n.$$

Therefore, the periodic switching

$$q(t) = \begin{cases} 1 & t \in \{(2l+1)\hat{k}, \dots, (2l+2)\hat{k}-1\} \\ 2 & t \in \{2l\hat{k}, \dots, (2l+1)\hat{k}-1\} \end{cases}$$

for $l = 0, 1, \dots$, yields an asymptotically stable system. The general case $\ell \geq 2$, follows in exact analogy: one shows that the map $\hat{A}_1^{\hat{k}}\hat{A}_2^{\hat{k}}\cdots\hat{A}_\ell^{\hat{k}}$ is a contraction with respect to the block norm $\|\cdot\|_*$, for some $\hat{k} \in \mathbb{N}$. Thus, there exists a periodic switching sequence which is stabilizing.

Appendix B
Proof of Lemma 3.5

The Riccati map $\mathcal{T}_q$ defined in (17) satisfies the identity

$$\mathcal{T}_q(\hat{\Pi}) = (I - \hat{K}_q(\hat{\Pi})C_q)\Xi(\hat{\Pi})(I - \hat{K}_q(\hat{\Pi})C_q)^{\mathsf{T}} \\ + \hat{K}_q(\hat{\Pi})FF^{\mathsf{T}}\hat{K}_q^{\mathsf{T}}(\hat{\Pi}). \quad (56)$$

Since both terms on the right hand side of (56) are positive definite, and FF^{T} is nonsingular then $z \in \ker\big(\mathcal{T}_q(\hat{\Pi})\big)$ implies that $z \in \ker\big(\hat{K}_q^{\mathsf{T}}(\hat{\Pi})\big)$ and $z \in \ker\big(\Xi(\hat{\Pi})\big)$. Moreover, since $\ker\big(\Xi(\hat{\Pi})\big) \subset \ker\big(\hat{K}_q^{\mathsf{T}}(\hat{\Pi})\big)$, we deduce from (56) that

$$\ker\big(\mathcal{T}_q(\hat{\Pi})\big) = \ker\big(\Xi(\hat{\Pi})\big) \\ = \ker\big(\hat{\Pi}A^{\mathsf{T}}\big) \cap \ker\big(D^{\mathsf{T}}\big). \quad (57)$$

Therefore, if $\{q_0, q_1, \dots, q_{k-1}\}$ is an arbitrary sequence and $\hat{\Pi}_k = \mathcal{T}_{q_{k-1}} \circ \cdots \circ \mathcal{T}_{q_0}(0)$, we obtain by (57) that

$$\ker(\hat{\Pi}_k) = \ker\big(D^{\mathsf{T}}\big) \cap \ker\big(D^{\mathsf{T}}A^{\mathsf{T}}\big) \cap \\ \cdots \cap \ker\big(D^{\mathsf{T}}(A^{\mathsf{T}})^{k-1}\big). \quad (58)$$

When $k = N_x$, (58) implies that $\hat{\Pi}_{N_x}$ is nonsingular if and only if (A, D) is controllable. The existence of $\varepsilon > 0$ as asserted in the lemma stems from the fact that that collection of maps $\{\mathcal{T}_q\}$ is finite.

Appendix C
Proof of Lemma 3.6

Note that if the filtering at time t is based upon y^{t-1} instead of y^t, the corresponding Riccati map is different from $\mathcal{T}$ and its convexity has been shown in [34].

To prove the concavity of $\mathcal{T}_q$, we show that for any scalar θ and symmetric square matrix Z, $\partial^2\mathcal{T}_q(\hat{\Pi} + \theta Z)/\partial\theta^2 \leq 0$. To simplify the notation, we define

$$\hat{\Pi}' \triangleq \hat{\Pi} + \theta Z, \qquad \Lambda_q \triangleq (C_q\Xi(\hat{\Pi}')C_q^{\mathsf{T}} + FF^{\mathsf{T}})^{-1}.$$

After some algebra, we obtain

$$\frac{\partial^2\mathcal{T}_q(\hat{\Pi}')}{\partial\theta^2} = -2(C_q^{\mathsf{T}}\Lambda_qC_q\Xi - I)^{\mathsf{T}}AZ^{\mathsf{T}}A^{\mathsf{T}}C_q^{\mathsf{T}}\Lambda_qC_q \\ \cdot AZA^{\mathsf{T}}(C_q^{\mathsf{T}}\Lambda_qC_q\Xi - I).$$

Since $\Lambda_q > 0$ and Z is symmetric, we have $\partial^2\mathcal{T}_q(\hat{\Pi}')/\partial\theta^2 \leq 0$, which shows that $\mathcal{T}_q$ is concave. The concavity of f_β^* follows from the fact that the map S_β is concavity-conserving, namely, $S_\beta(f)$ is concave if f is concave.

Appendix D
Proof of Lemma 3.7

i) As mentioned in the proof of Lemma 3.5, $\Sigma' \geq \Sigma$ implies $\mathcal{T}_q(\Sigma') \geq \mathcal{T}_q(\Sigma)$. Hence, it follows by induction from (35) that if $\Sigma' \geq \Sigma$, then $f_\beta^*(\Sigma') \geq f_\beta^*(\Sigma)$. Thus $f_\beta^*(0) = \inf_{\Sigma \in \mathcal{M}_0^+} f_\beta^*(\Sigma)$.

ii) Let $\varepsilon > 0$ be the constant in Lemma 3.5. For a $\beta \in (0, 1)$, let $\Sigma_\beta^* \in \mathcal{B}_s$ be such that

$$f_\beta^*(\Sigma_\beta^*) \geq \sup_{\Sigma \in \mathcal{B}_s} f_\beta^*(\Sigma) - \varepsilon.$$

If v_β^* is an optimal β-discounted policy, then

$$f_\beta^*(0) = \mathbb{E}_0^{v_\beta^*}\left[\sum_{t=0}^{\kappa-1}\beta^t(c(Q_t) + \mathrm{tr}(\tilde{\Pi}_\beta^*\hat{\Pi}_t)) + \beta^\kappa f_\beta^*(\hat{\Pi}_\kappa)\right] \\ \geq \beta^\kappa\mathbb{E}_0^{v_\beta^*}\left[f_\beta^*(\hat{\Pi}_\kappa)\right].$$

Thus

$$\operatorname*{span}_{\mathcal{B}_s}(f_\beta^*) \leq f_\beta^*(\Sigma_\beta^*) - f_\beta^*(0) + \varepsilon \\ \leq f_\beta^*(\Sigma_\beta^*) - \beta^\kappa\mathbb{E}_0^{v_\beta^*}\left[f_\beta^*(\hat{\Pi}_\kappa)\right] + \varepsilon \\ = (1 - \beta^\kappa)f_\beta^*(\Sigma_\beta^*) \\ + \beta^\kappa\mathbb{E}_0^{v_\beta^*}\left[f_\beta^*(\Sigma_\beta^*) - f_\beta^*(\hat{\Pi}_\kappa)\right] + \varepsilon \\ \leq (1 - \beta^\kappa)f_\beta^*(\Sigma_\beta^*) + \beta^\kappa(1 - \varepsilon)\operatorname*{span}_{\mathcal{B}_s}(f_\beta^*) + \varepsilon$$

where the last inequality follows from the concavity of f^*_β, and the fact that $\hat{\Pi}_\kappa \in \mathcal{M}^+_\varepsilon$. Therefore

$$\begin{aligned}\operatorname*{span}_{\mathcal{B}_s}(f^*_\beta) &\le \frac{(1-\beta^\kappa)f^*_\beta(\Sigma^*_\beta)+\varepsilon}{1-\beta^\kappa(1-\varepsilon)} \\ &= \frac{(1+\beta+\cdots+\beta^{\kappa-1})(1-\beta)f^*_\beta(\Sigma^*_\beta)+\varepsilon}{1-\beta^\kappa(1-\varepsilon)} \\ &\le \frac{\kappa}{\varepsilon}(1-\beta)f^*_\beta(\Sigma^*_\beta)+1\,.\end{aligned}$$

Since, by (28) $(1-\beta)f^*_\beta(\Sigma^*_\beta)$ is bounded, uniformly in $\beta \in (0,1)$, the same holds for $\operatorname{span}_{\mathcal{B}_s}(f^*_\beta)$. The result then follows by (38).

iii) Equicontinuity of $\{\bar{f}_\beta\}$ on bounded subsets of $\mathcal{M}^+_\varepsilon$, for any $\varepsilon > 0$, follows from the uniform boundedness and concavity of $\{\bar{f}_\beta\}$ [35]. Since, by (1), $\mathcal{T}_q(\Sigma) \ge \mathcal{T}_q(0)$, for any $\Sigma \in \mathcal{M}^+_0$, then by Lemma 3.5

$$\mathcal{T}_{q_\kappa} \circ \mathcal{T}_{q_{\kappa-1}} \circ \cdots \circ \mathcal{T}_{q_0}(\Sigma) \in \mathcal{M}^+_\varepsilon\,, \quad \forall \Sigma \in \mathcal{M}^+_0$$

for all $\{q_0,\ldots,q_\kappa\} \in \boldsymbol{Q}^{\kappa+1}$. Fix the initial condition Σ, and let $\{q_0,q_1,\ldots\}$ be a corresponding β-discounted optimal sequence of queries, i.e., selectors from the minimizer in (30). Define $\mathcal{T}_{q_k,\ldots,q_0} \triangleq \mathcal{T}_{q_k} \circ \mathcal{T}_{q_{k-1}} \circ \cdots \circ \mathcal{T}_{q_0}$. Using (32), for any $\Sigma' \in \mathcal{M}^+_0$

$$\begin{aligned}f^*_\beta(\Sigma') - f^*_\beta(\Sigma) \le\ & \operatorname{tr}(\tilde{\Pi}^*_\beta(\Sigma'-\Sigma)) \\ &+ \sum_{k=0}^{\kappa-1}\beta^{k+1}\operatorname{tr}\big(\tilde{\Pi}^*_\beta(\mathcal{T}_{q_k,\ldots,q_0}(\Sigma') - \mathcal{T}_{q_k,\ldots,q_0}(\Sigma))\big) \\ &+ \beta^{\kappa+1}\big[\bar{f}_\beta(\mathcal{T}_{q_\kappa,\ldots,q_0}(\Sigma')) - \bar{f}_\beta(\mathcal{T}_{q_\kappa,\ldots,q_0}(\Sigma))\big]\,. \quad (59)\end{aligned}$$

Thus, equicontinuity on every compact subset of $\mathcal{M}^+_0$ follows from (59), by exploiting the continuity of $\mathcal{T}_{q_\kappa,\ldots,q_0}$, the property $\mathcal{T}_{q_\kappa,\ldots,q_0}(\mathcal{M}^+_0) \subset \mathcal{M}^+_\varepsilon$, and the fact that and $\bar{f}_\beta$ is equicontinuous on bounded subsets of $\mathcal{M}^+_\varepsilon$.

References

[1] W. Wong and R. Brockett, "Systems with finite communication bandwidth constraints I: state estimation problems," *IEEE Trans. Automat. Control*, vol. 42, no. 9, pp. 1294–1299, Sep. 1997.

[2] W. Wong and R. Brockett, "Systems with finite communication bandwidth constraints II: Stabilization with limited information feedback," *IEEE Trans. Automat. Control*, vol. 44, no. 5, pp. 1049–1053, May 1999.

[3] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Automat. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.

[4] S. Tatikonda and S. Mitter, "Control under noisy channels," *IEEE Trans. Automat. Control*, vol. 49, no. 7, pp. 1196–1201, Jul. 2004.

[5] N. Elia and S. Mitter, "Stabilization of linear systems with limited information," *IEEE Trans. Automat. Control*, vol. 46, no. 9, pp. 1384–1400, Sep. 2001.

[6] G. Nair and R. Evans, "Stabilization with data-rate-limited feedback: Tightest attainable bounds," *Syst. Control Lett.*, vol. 41, pp. 49–76, 2000.

[7] D. Liberzon, "On stabilization of linear systems with limited information," *IEEE Trans. Automat. Control*, vol. 48, no. 2, pp. 304–307, Feb. 2003.

[8] G. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1585–1597, Sep. 2004.

[9] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Trans. Automat. Control*, vol. 50, no. 6, pp. 910–915, Jun. 2005.

[10] L. Meier, J. Peschon, and R. M. Dressler, "Optimal control of measurement subsystems," *IEEE Trans. Automat. Control*, vol. AC-12, no. 5, pp. 528–536, Oct. 1967.

[11] M. Athans, "On the determination of of optimal costly measurement strategies for linear stochastic systems," *Automatica.*, vol. 8, pp. 397–412, 1972.

[12] K. Herring and J. Melsa, "Optimum measurements for estimation," *IEEE Trans. Automat. Control*, vol. AC-19, no. 3, pp. 264–266, Jun. 1974.

[13] R. Mehra, "Optimization of measurement schedules and sensor designs for linear dynamic systems," *IEEE Trans. Automat. Control*, vol. AC-21, no. 1, pp. 55–64, Feb. 1976.

[14] E. Mori and J. J. DiStefano, "Optimal nonuniform sampling interval and test-input for identification of physiological system from very limited data," *IEEE Trans. Automat. Control*, vol. AC-24, no. 6, pp. 893–900, Dec. 1979.

[15] D. Avitzour and S. R. Rogers, "Optimal measurement scheduling for prediction and estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1733–1739, Oct. 1990.

[16] M. Shakeri, K. R. Pattipati, and D. L. Kleinman, "Optimal measurement scheduling for state estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 716–729, Apr. 1995.

[17] B. M. Miller and W. J. Runggaldier, "Optimization of observations: A stochastic control approach," *SIAM J. Control Optim.*, vol. 35, no. 5, pp. 1030–1052, 1997.

[18] A. V. Savkin, R. Evans, and E. Skafidas, "The problem of optimal robust sensor scheduling," *Syst. Control Lett.*, vol. 43, pp. 149–157, 2001.

[19] V. Krishnamurthy and R. J. Evans, "Hidden Markov model multiarmed bandits: A methodology for beam scheduling in multitarget tracking," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 2893–2908, Dec. 2001.

[20] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and dynamic sensor coverage," *Automatica.*, vol. 42, no. 2, pp. 251–260, 2006.

[21] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov Processes*. New York: Springer-Verlag, 1979, vol. 235.

[22] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus, "Discrete-time controlled Markov processes with average cost criterion: A survey," *SIAM J. Control Optim.*, vol. 31, no. 3, pp. 282–344, 1993.

[23] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models, Estimation and Control*. New York: Springer-Verlag, 1995, vol. 29.

[24] S.-P. Hsu, D. M. Chuang, and A. Arapostathis, "On the existence of stationary optimal policies for partially observed MDPs under the long-run average cost criterion," *Syst. Control Lett.*, vol. 55, pp. 165–173, 2006.

[25] D. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*. Englewood Cliffs, NJ: Prentice Hall, 1987.

[26] D. Liberzon and A. S. Morse, "Basic problems in stability and design of switched systems," *IEEE Contr. Syst. Mag.*, vol. 19, pp. 59–70, Oct. 1999.

[27] Z. Sun and D. Zheng, "On reachability and stabilization of switched linear systems," *IEEE Trans. Automat. Control*, vol. 46, no. 2, pp. 291–295, Feb. 2001.

[28] S. S. Ge, Z. Sun, and T. H. Lee, "Reachability and controllability of switched linear discrete-time systems," *IEEE Trans. Automat. Control*, vol. 46, no. 9, pp. 1437–1441, Sep. 2001.

[29] Z. Sun and S. S. Ge, "Analysis and synthesis of switched linear control systems," *Automatica*, vol. 41, no. 2, pp. 181–195, 2005.

[30] G. Xie and L. Wang, "Controllability and stabilizability of switched linear-systems," *Syst. Control Lett.*, vol. 48, no. 2, pp. 135–155, 2003.

[31] G. Xie and L. Wang, "Reachability realization and stabilizability of switched linear discrete-time systems," *J. Math. Anal. Appl.*, vol. 280, no. 2, pp. 209–220, 2003.

[32] G. Xie and L. Wang, "Stabilization of a class switched linear systems," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2003, vol. 2, pp. 1898–1903.

[33] L. Zhang and D. Hristu-Varsakelis, "LQG control under limited communication," in *Proc. 44th IEEE Conf. Decision Control Eur. Control Conf.*, Dec. 2005, pp. 185–190.

[34] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[35] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1946, vol. 28.

Wei Wu (S'01–M'07) received the B.S. degree in applied physics and the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1999 and 2002, respectively, and is currently pursuing the Ph.D. degree in electrical and computer engineering at the University of Texas at Austin.

His research interests include estimation and optimal control for stochastic systems, the heavy traffic analysis of communication networks, and feedback information theory.

Ari Arapostathis (F'07) received the B.S. degree from the Massachusetts Institute of Technology, Cambridge and the Ph.D. degree from the University of California, Berkeley.

He is currently with the University of Texas at Austin, where he is a Professor in the Department of Electrical and Computer Engineering. His research interests include stochastic and adaptive control theory, the application of differential geometric methods to the design and analysis of control systems, and hybrid systems.