2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Optimal Control of Stochastic Systems with Costly Observations—The General Markovian Model and the LQG Problem

Wei Wu
Electrical and Computer Engineering
The University of Texas at Austin
1 University Station C0803
Austin, TX 78712–0240
Email: wwu@ece.utexas.edu

Ari Arapostathis
Electrical and Computer Engineering
The University of Texas at Austin
1 University Station C0803
Austin, TX 78712–0240
Email: ari@ece.utexas.edu

Abstract— **In this paper, we examine a discrete-time stochastic control problem in which there are a number of observation options available to the controller, with varying associated costs. The observation costs are added to the running cost of the optimization criterion and the resulting optimal control problem is investigated. This problem is motivated by the wide deployment of networked control systems and data fusion. Since only part of the observation information is 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 (POMDP) model and then specialize to the stochastic LQG problem, where we show that the separation principle still holds. Moreover we show that the effect of the observation cost is manifested on the estimation strategy as follows: instead of a Kalman filter with gain determined by the algebraic Riccati equation, the optimal estimator includes, in addition, a query strategy which is characterized by a dynamic programming equation. The structure of the optimal query for a one-dimensional system is studied analytically and simulated with numerical examples.**

I. Introduction

Recently, much attention has been paid on network 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. Rather they are transmitted through communication channels, which might suffer delay, transmission errors, and as such this transmission carries a cost. Understanding the interaction between the control system and the communication system is very important as it plays a key role on the overall performance of NCS.

Stability is a basic requirement for a control system, especially for network control systems. This raises the question of how much information a feedback controller needs 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 [14], [15], Tatikonda and Mitter [12], [13], Elia and Mitter [6], Nair and Evans [11], Liberzon [10] and many others; stability of nonlinear control systems is further studied in [8] and [4].

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, generating more traffic in the network, 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 observation on the system state sent to the controller through a communication channel. The controller has the option of requesting different levels 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 requested 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 level of information requested, which is added to the running cost of the original control objective. As a result the observation space is not static, but rather changes dynamically as the controller issues different queries on the sensors.

There exists work in the literature studying the sensor scheduling problem [1], [9], in which there are a number of sensors with different levels of precision and operation costs and the controller accesses only one sensor at a time to obtain the observation and minimize the estimation error and operation cost. In [1] sensor scheduling is addressed for continuous-time linear systems while in [9] the dynamics correspond to a hidden Markov chain. In this work, we describe a general model in the context of POMDPs and then

0-7803-9098-9/05/$25.00 ©2005 AACC

specialize to the stochastic discrete-time linear quadratic Gaussian (LQG) problem. Our aim is to study optimization over the infinite horizon—both for the discounted cost (DC) and the long-term average cost (AC). We show that for the LQG problem, the separation principle of estimation and control still holds, and hence the optimal control can be fully decoupled into two subproblem: an optimal query/estimation problem and an optimal control problem with full observations. The estimation problem still reduces to a Kalman filter, with the gain computed by a discrete-Riccati equation. However, the optimal query is solved by a dynamic programming equation. We further specialize to the one-dimensional LQG problem where we obtain various analytical results: necessary and sufficient conditions for the query strategy to be dynamic, and also the existence of a solution to the long-term average dynamic programming equation. In particular, it is shown that optimal query is attained by a feedback policy which is a function of the error variance, and has a threshold structure for the one-dimensional problem.

II. 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_t 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 a space. 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

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

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

We suppose there are m different observation processes available but only one process 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 m 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 $\mathcal{P}(\boldsymbol{Y}^i) \times \boldsymbol{X}$, which may be interpreted as the conditional distribution of Y_t^i given X_t.

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 $\boldsymbol{Q} = \{1, \dots, m\}$. 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 and the data set $Y_{t+1}^q \in \boldsymbol{Y}^{q_{t+1}}$, corresponding to the queried sensors, is obtained.

Following the standard POMDP model formulation (as in [5]), we define the history spaces $\{\boldsymbol{H}_t\}$ by

$$\begin{aligned} \boldsymbol{H}_0 &= \mathcal{P}(\boldsymbol{X}) \\ \boldsymbol{H}_{t+1} &= \boldsymbol{H}_t \times \boldsymbol{U} \times \boldsymbol{Q} \times \boldsymbol{Y}\,, \quad t = 0, 1, \dots, \end{aligned}$$

where $\boldsymbol{Y} = \cup_{q \in \boldsymbol{Q}} \boldsymbol{Y}^q$. Thus, the generic element of the history space $\boldsymbol{H}_t$ is denoted by

$$h_t = (\mu, u_0, q_0, y_1, \dots, x_{t-1}, u_{t-1}, q_{t-1}, y_t) \in \boldsymbol{H}_t\,.$$

The information available for decision making at time t is the history $\mathcal{H}_t = \sigma\{H_t\}$, where $\{H_t\}$ stands for the history process.

An *admissible* control is a sequence $v = (v_0, v_1, \dots)$, where each v_t is a kernel on $\boldsymbol{U} \times \boldsymbol{Q}$ given $\boldsymbol{H}_t$. Specifying an admissible control v, obtains a unique probability measure $\mathbb{P}_\mu^v$ on the path space of the process. Markov controls and stationary controls are defined in the standard manner. We let $\mathcal{U}$ denote all admissible controls, and $\mathcal{U}_M$, $\mathcal{U}_S$ all Markov and 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 POMDPs, we obtain an equivalent completely observed model through the introduction of the conditional distribution $\varPsi_t$ of the state given the observations [2], [5], [7]. We let $\boldsymbol{\varPsi} := \mathcal{P}(\boldsymbol{X})$. 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. On the other hand, the query variable gives us the freedom to choose the nonlinear Bayesian filters to update the state estimates.

$$\begin{aligned} \tilde{P}(dx, y|\psi_t, u_t, q) &:= \int_{x' \in \boldsymbol{X}} \mathcal{K}^q(y|x) P(dx|x', u_t) \psi_t(dx'), \\ V(y, \psi_t, u_t, q) &:= \int_{x \in \boldsymbol{X}} \tilde{P}(dx, y|\psi_t, u_t, q), \end{aligned}$$

$$\begin{aligned} \psi_{t+1} &= T(\psi_t, y, u_t, q)(dx) \\ &= \begin{cases} \frac{\tilde{P}(dx, y|\psi_t, u_t, q)}{V(y, \psi_t, u_t, q)}, & \text{if } V(y, \psi_t, u_t, q) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{1}$$

Here T is the nonlinear filtering operator that has the intuitive interpretation as the a posteriori conditional distribution of the state, given that decision (u, q) was made, observation $y \in \boldsymbol{Y}^q$ obtained, and an a priori distribution ψ. Likewise, $V(dy, \psi, u, q)$ is interpreted as the one-step ahead conditional probability on the observation space $\boldsymbol{Y}^q$ given an a priori distribution ψ for the state, under decision u. With the option to select the query variable q at each step, we are able to choose the proper Bayesian filter to update the state estimates.

The model also 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 query 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 run average infinite-horizon criterion.

In other words we seek to minimize, over all admissible policies $v \in \mathcal{U}$,

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

We also consider the β-discounted criterion, $\beta \in (0,1)$,

$$J^{(v)}_\beta(x) := \mathbb{E}^v_x \left[\sum_{t=0}^{\infty} \beta^t g(X_t, U_t, Q_t) \right] . \qquad (3)$$

Define $J^* := \inf_{v\in\mathcal{U}} J^{(v)}$, $J^*_\beta(x) := \inf_{v\in\mathcal{U}} J^{(v)}_\beta(x)$.

Now let $\tilde{g}(\psi, u, q) = \int g(x, a, q)\psi(dx)$. It is straightforward to express the control objectives in (2) and (3) in the equivalent CO model. Due to its contraction properties, there always exists a stationary Markov policy for the objective in (3), satisfying the Hamilton-Jacobi-Bellman (HJB) equation,

$$\begin{aligned} J^*_\beta(\psi) = \min_{(u,q)\in \boldsymbol{U}\times\boldsymbol{Q}} \Big\{ &\tilde{g}(\psi, u, q) \\ &+ \beta \int_{\boldsymbol{Y}^q} V(dy, \psi, u, q) J^*_\beta\big(T(y, \psi, u, q)\big) \Big\} . \qquad (4) \end{aligned}$$

Minimizing the long-term average cost is accomplished, under certain conditions, by the HJB equation

$$\begin{aligned} J^* + h(\psi) = \min_{(u,q)\in \boldsymbol{U}\times\boldsymbol{Q}} \Big\{ &\tilde{g}(\psi, u, q) \\ &+ \int_{\boldsymbol{Y}^q} V(dy, \psi, u, q) h\big(T(y, \psi, u, q)\big) \Big\} , \qquad (5) \end{aligned}$$

provided of course that such a solution exists. In (5), J^* is the optimal average cost, and h is called the bias function.

III. Stochastic Linear Control with Observation Cost

In this section, we consider a stochastic linear system in discrete-time, with quadratic running penalty (LQG). First, in subsection III-A we derive the LQG control models for both discounted cost and long-term average cost. Then, the dynamic programming equation is further simplified and decoupled into two separate problems: the optimal estimation problem and the control problem, the latter being a standard LQG optimal control.

A. Linear quadratic Gaussian (LQG) control: the model

Consider a system evolving according to

$$x_{t+1} = Ax_t + Bu_t + \epsilon_t, \qquad (6)$$

where $x_t \in R^n$ is the state, $u_t \in R^l$ is the control, and the process noise ϵ_t is Gaussian, with zero mean and covariance matrix M. The state-process is observed by

$$y^q_t = C^q x_t + \eta^q_t, \qquad (7)$$

when the query action $q \in \boldsymbol{Q}$ is issued, where $y^q_t \in R^{n_q}$ is the measurement. The observation noise η^q_t is also assumed to be Gaussian, with zero mean and covariance matrix N^q. As usual, the family $\{(\epsilon_t, \eta^q_t), t = 0, \dots,\}$ is assumed independent, and also independent from the initial state x_0. Furthermore, the covariance of ϵ_t, η^q_t is given by

$$\operatorname{cov}\begin{pmatrix}\epsilon_t \\ \eta_t\end{pmatrix} = \begin{pmatrix} M & L^q \\ (L^q)^\mathsf{T} & N^q \end{pmatrix} .$$

Lastly, the running cost is assumed to be a quadratic function, i.e., $r(x, u) = \|Dx + Fu\|^2$, for some matrices D and F.

B. Stability

In the classical LQG control setup, it is well known that under the condition that (A, B) is *stabilizable* and (C, A) is *detectable*, the optimal control is stable, i.e., it results in a bounded variance for the state process. The following statement is straightforward to prove.

Theorem 1: Suppose (A, B) is stabilizable and there exists $q \in \boldsymbol{Q}$ such that (C^q, A) is detectable. Then the optimal control over the infinite horizon is stable.

It is interesting to note that it is possible for a feedback controller to stabilize the system even when none of the observation pairs (C^q, A) are detectable. Consider a system with the following parameters:

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} , \quad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad C^1 = (1, 0) , \quad C^2 = (0, 1) .$$

When $q = 1$, $q = 2$, the observer detects x_1, x_2, respectively. For any fixed q, no feedback controller can stabilize the system; on the other hand, under the query sequence $q = \{1, 2, 1, 2, \cdots\}$, sufficient information is obtained to design a stable controller.

C. Separation principle

In this section, we analyze the optimal control problem and show that the separation principle holds. First we consider the optimal control problem over a finite horizon, which reduces to minimizing the functional:

$$\begin{aligned} J_k = \mathbb{E}\Big[\sum_{t=0}^{k} c(q_t) + \|D_t x_t + F_t u_t\|^2 \Big] \\ + \mathbb{E}\big[x^\mathsf{T}_{k+1} G x_{k+1}\big] , \qquad (8) \end{aligned}$$

where the term $x^\mathsf{T}_{k+1} G x_{k+1}$ is the terminal penalty. The objective is to choose $\{q_0, u_0, q_1, u_1, \cdots, u_k\}$ so as to minimize J_k. According to the general POMDP model we have discussed in the last section, we can obtain an equivalent completely observed model through the conditional distribution $\psi_t = P(x_t \mid H_t)$. Let y^t and q^t denote the observation history and the query history up to time t, respectively. One important aspect of the standard LQG problem [3] is that the conditional distribution ψ_t of x_t given y^t is Gaussian. It follows along the same lines, that for the problem at hand, the conditional law of x_t given $h_t = (q^{t-1}, y^t)$ is also Gaussian, with mean $\bar{x}_t$ and variance W_t, satisfying

$$\bar{x}_{t+1} = A\bar{x}_t + Bu_t + K_t[y^{q_t}_{t+1} - C(A\bar{x}_t + Bu_t)], \quad (9a)$$

$$W_{t+1} = T_q(W_t) = AW_tA^\mathsf{T} + M - K_t\tilde{W}^\mathsf{T}_t, \qquad (9b)$$

where

$$\tilde{W}_t := L + MC^\top + AW_tA^\top C^\top$$
$$K_t := \tilde{W}_t(CAW_tA^\top C^\top + CMC^\top + N)^{-1}\,.$$

Here we denote $C^{q_t}, L^{q_t}, N^{q_t}$ as C, L, N respectively, to simplify the notation. Note that the form of (9a)–(9b) is the same as the classical Kalman filter [3]. Due to fact that the Gaussian distribution is uniquely determined by its mean and variance, the optimal policy in (8) can be computed via the dynamic programming equation

$$J_t(\bar{x}_t, W_t) = \min_{q_t,u_t}\Big\{c(q_t) + \mathbb{E}\,\|Dx_t + Fu_t\|^2 + \mathbb{E}\big[J_{t+1}(\bar{x}_{t+1}, W_{t+1}) \mid h_t\big]\Big\}. \quad (10)$$

By (9a)–(9b), we obtain

$$u_k = -(F^\top F + B^\top GB)^{-1}(F^\top D + B^\top GA)\bar{x}_k,$$

and the optimal q_k is a selector from

$$q_k \in \arg\min_q\Big\{c(q) + tr\big[DW_tD^\top\big] + tr\big[(AW_tA^\top + M - T_q(W_t))\Pi\big]\Big\},$$

where $tr[\cdot]$ denotes the trace of a matrix. Then we solve (10) for $t = k-1, k-2, \ldots$, to obtain

$$u_t = -(F^\top F + B^\top \Pi_{t+1}B)^{-1}(F^\top D + B^\top \Pi_{t+1}A)\bar{x}_t,$$

and after some algebra, we deduce that the optimal query, is obtained from

$$f_t(W_t) = \min_q\Big\{c(q) + tr\big[DW_tD^\top\big] + f_{t+1}(T_q(W_t)) + tr\big[(AW_tA^\top + M - T_q(W_t))\Pi_{t+1}\big]\Big\}, \quad (11)$$

where $f(W_t)$ is a function of the variance matrix W_t, and $T_q(\cdot)$ is the map defined in (9b). We refer to T_q as a *Riccati operator*. Note that (11) is in the form of a deterministic dynamic programming equation. As mentioned earlier, the optimal estimation is not provided by a Kalman filter alone, but it also involves the solution of the dynamic programming equation (11). In summary, we have following theorem:

Theorem 2 (Separation Principle): The optimal control (6)–(7), relative to the finite horizon criterion in (8) is given by

$$u_t = -(F^\top F + B^\top \Pi_{t+1}B)^{-1}(F^\top D + B^\top \Pi_{t+1}A)\bar{x}_t,$$

where $\Pi_{k+1} = G$, and for $0 \le t < k+1$,

$$\Pi_t = D^\top D + A^\top \Pi_{t+1}A - (D^\top F + A^\top \Pi_{t+1}B)\cdot (F^\top F + B^\top \Pi_{t+1}B)^{-1}(F^\top D + B^\top \Pi_{t+1}A), \quad (12)$$

and

$$f_t(W_t) = \min_q\Big\{c(q) + tr\big[DW_tD^\top\big] + f_{t+1}(T_q(W_t)) + tr\big[(AW_tA^\top + M - T_q(W_t))\Pi_{t+1}\big]\Big\}, \quad (13)$$

with $f_{k+1} = 0$.

We have derived the optimal control for finite horizon. Similar results hold for the infinite horizon problem, under suitable stability assumptions. First let us consider the infinite horizon discounted cost. The HJB equation of (4) can be expressed in the following form,

$$J_\beta(\bar{x}_t, W_t) = \min_{u,q}\Big\{\big[c(q) + \mathbb{E}\,\|Dx + Fu\|^2\,|h_t\big] + \beta\mathbb{E}[J_\beta(\bar{x}_{t+1}, W_{t+1})|h_t]\Big\}. \quad (14)$$

Using standard arguments, we show that the solution takes the form

$$J_\beta(\bar{x}, W) = \bar{x}^\top \Pi\bar{x} + f_\beta(W)\,.$$

where Π is a symmetric non-negative definite matrix. Then the optimal control u_t satisfies

$$u_t = -(F^\top F + \beta B^\top \Pi B)^{-1}(F^\top D + \beta B^\top \Pi A)\bar{x}_t \quad (15)$$

where Π is the solution of the algebraic Riccati equation,

$$\Pi = D^\top D + A^\top \Pi A - (D^\top F + \beta A^\top \Pi B)(F^\top F + \beta B^\top \Pi B)^{-1}(F^\top D + \beta B^\top \Pi A), \quad (16)$$

and q is the minimizer of

$$f_\beta(W) = \min_q\Big\{c(q) + tr\big[(AWA^\top + M - T_q(W))\Pi\big] + tr\big[DWD^\top\big] + \beta f_\beta(T_q(W))\Big\}. \quad (17)$$

Under the assumption that (A, B) is stabilizable and there exists $q \in \boldsymbol{Q}$ such that (C^q, A) is detectable, the discounted cost problem has a bounded solution for any $\beta < 1$.

Concerning the long-term average cost criterion the following result is stated without proof.

Theorem 3 (Ergodic control): Suppose (A, B) is stabilizable and there exists $q \in Q$ such that (A, C^q) is detectable, then there exists an optimal control which is stable, and q is the minimizer of the dynamic programming equation

$$\rho + f(W) = \min_q\Big\{c(q) + tr\big[(AWA^\top + M - T_q(W))\Pi\big] + tr\big[DWD^\top\big] + f(T_q(W))\Big\}, \quad (18)$$

while u is as in (15) and Π is the solution of the algebraic Riccati equation in (16), with $\beta = 1$.

One can show that if we fix the query variable q to be constant, (17) and (18) become equivalent to the Riccati equation for the Kalman filter.

In summary, the steps to compute the optimal controller are as follows:

1) First we solve for Π in the Riccati equation (16). The optimal control is given by the linear feedback controller in (15) based solely on the mean $\bar{x}$ with a constant gain.

2) With Π obtained in the previous step, we solve the dynamic programming equation (17) for the discounted cost, or (18) for the average cost, to obtain the optimal stationary policy for the query q.
3) At each step, the optimal q is determined by W_t, and the state estimates are updated according to (9a)–(9b).

IV. Optimal Estimation in 1-D

In this section, we concentrate on the optimal estimation problem and explore the properties and structure of the optimal estimator for a one-dimensional system.

Consider the 1-D linear system,

$$\begin{aligned} x_{t+1} &= Ax_t + \epsilon_t \\ y_t &= C^q x_t + \eta_t^q \end{aligned}$$

The objective is to estimate the system state $\hat{x}_t$ in such a manner so as to minimize the infinite horizon criteria (DC and AC), with a running cost given as the sum of the estimation error variance w_t and the observation cost $c(q_t)$, at each time step t. Let $v(q) := N^q/(C^q)^2$, which may be viewed as the normalized noise. Then, $T_q(w)$ in (9b) can be written as

$$T_q(w) = \frac{v(q)(A^2w + m)}{A^2w + m + v(q)},$$

and the HJB equation of (17) takes the form

$$f_\beta(w) = \min_q \bigl\{c(q) + w + \beta f_\beta(T_q(w))\bigr\}.$$

If we let $g_\beta(w) = f_\beta(w) - w$, we have

$$g_\beta(w) = \min_q \bigl\{c(q) + \beta T_q(w) + \beta g_\beta(T_q(w))\bigr\}. \qquad (19)$$

Now suppose there are two query options, $q \in \boldsymbol{Q} = \{1, 2\}$, with observation costs $c(1)$, $c(2)$, and corresponding normalized noise $v(1)$, $v(2)$, where $c(1) < c(2)$ and $v(1) > v(2)$. The analysis of this dynamic programming equation even for the one-dimensional case is rather involved and the results stated are given without proofs due to lack of space.

Let $\hat{w}_1$, $\hat{w}_2$ denote the unique fixed points of T_1, T_2 respectively. Since $v_1 > v_2$, we have $\hat{w}_1 > \hat{w}_2$. Let q_β^* denote the minimizer of (19) (i.e., an optimal β-discounted Markov policy), and q^* denotes the minimizer of the ergodic-cost optimality equation. For example, $q^*(w) = 2$ denotes that the optimal policy is to use the query $q = 2$, when the estimation error variance is w. It is clear that we may restrict our attention to the set $[\hat{w}_2, \hat{w}_1]$, since for $w > \hat{w}_1$, $q_\beta^*(w) = q_\beta^*(\hat{w}_1)$ and $q^*(w) = q^*(\hat{w}_1)$, while for $w < \hat{w}_2$, $q_\beta^*(w) = q_\beta^*(\hat{w}_2)$ and $q^*(w) = q^*(\hat{w}_2)$. Once $w \in [\hat{w}_2, \hat{w}_1]$, applying $q(w) = 1$ will result in a variance $T_1(w) > w$, but incur the smallest penalty $c(1)$, while applying $q(w) = 2$, results in $T_2(w) < w$, but the higher penalty $c(2)$ is paid. We investigate under what conditions the optimal stationary query policy q^* is dynamic, i.e., not a constant. For this, we need to state some properties of T_q.

Lemma 4: It holds, for all w, w',

$$T_q(w) - T_q(w') = T_q(w)T_q(w')\frac{A^2(w - w')}{(A^2w + m)(A^2w' + m)}\,.$$

Lemma 5: Define

$$S_q(w) := \frac{T_q(w)}{(A^2T_q(w) + m)}\,.$$

Let $(q_1, q_2, \dots, q_k)$ be any finite sequence with elements in $\{1, 2\}$. Then, using Lemma 4, we obtain by induction on k that

$$\begin{aligned} &T_{q_1} \circ \cdots \circ T_{q_k}(w) - T_{q_1} \circ \cdots \circ T_{q_k}(w') = \\ &A^{2k}T_{q_1}(w)T_{q_1}(w')\Gamma(w, w')\frac{(w - w')}{(A^2w + m)(A^2w' + m)}\,, \end{aligned}$$

where

$$\Gamma(w, w') = \prod_{i=2}^{k} S_{q_i}(w) \times S_{q_i}(w')\,.$$

Lemma 6: The following identity holds

$$T_1(w) - T_2(w) = A^2\frac{v_1 - v_2}{v_1v_2}\,T_1(w)T_2(w)\,.$$

Indeed write

$$\begin{aligned} \hat{w}_1 + \hat{w}_2 - T_2(\hat{w}_1) - T_1(\hat{w}_2) &= \bigl(T_1(\hat{w}_1) - T_2(\hat{w}_1)\bigr) \\ &\quad - \bigl(T_1(\hat{w}_2) - T_2(\hat{w}_2)\bigr)\,, \end{aligned}$$

and the result follows from the fact that $w \to T_1(w) - T_2(w)$ is strictly monotone increasing.

Combining Lemmas 4–6 and noting that T_q and S_q are strictly monotone increasing, we have the following theorem:

Theorem 7: Let $(q_1, q_2, \dots, q_k)$ be any finite sequence with elements in $\{1, 2\}$. Then the map

$$w \mapsto T_{q_1} \circ \cdots \circ T_{q_k}\bigl(T_1(w)\bigr) - T_{q_1} \circ \cdots \circ T_{q_k}\bigl(T_2(w)\bigr)$$

is strictly monotone increasing on $[\hat{w}_2, \hat{w}_1]$.

The following theorem provides a necessary and sufficient condition for the query policy to be non-constant.

Theorem 8: In order for $q_\beta^*(\hat{w}_2) = 1$, it is necessary and sufficient that

$$c_2 + \sum_{k=0}^{\infty} \beta^k\bigl(\hat{w}_2 - T_2^k \circ T_1(\hat{w}_2)\bigr) > c_1\,, \qquad (20)$$

while in order for $q_\beta^*(\hat{w}_1) = 2$, it is necessary and sufficient that

$$c_2 < c_1 + \sum_{k=0}^{\infty} \beta^k\bigl(\hat{w}_1 - T_1^k \circ T_2(\hat{w}_1)\bigr)\,. \qquad (21)$$

Under the conditions (20)–(21), one can show that the span of the discounted cost function g_β in (19) is bounded, and passing to the limit as $\beta \to 1$, the standard approach obtains the existence of an stationary optimal policy for the AC criterion, which is characterized as in the following theorem.

Theorem 9: If (20)–(21) are satisfied, there exists an optimal stationary query policy minimizing the long term average cost for the estimation problem, which is the minimizer of the dynamic programming equation

$$g(w) = \min_q \{ c(q) + T_q(w) + g(T_q(w)) \} .$$

The solution $g : [\hat{w}_2, \hat{w}_1] \to \mathbb{R}$ is a concave function.

It turns out that the optimal query policy for both the discounted cost and average cost is a threshold policy based on the estimation error variance, namely, the optimal q^* is in the form

$$q^* = \begin{cases} 1, & w < w^* \\ 2, & w \geq w^* \end{cases}$$

In Figure 1, we first compare the optimal policies for the discounted cost and the average cost. As β approaches 1, the optimal thresholds for the discounted cost converge to the threshold of the optimal policy for the average cost. Furthermore, the optimal threshold is decreasing as β approaches 1, which agrees with intuition, namely 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. Figure 2, shows how the difference in cost between the two queries affects the optimal policy. As the price difference increases, the threshold point of the optimal policy is also increasing. Once the price difference reaches 0.45, then the optimal policy is constant: the cost for the better observation is high enough that the controller chooses to use the least costly observation all the time.

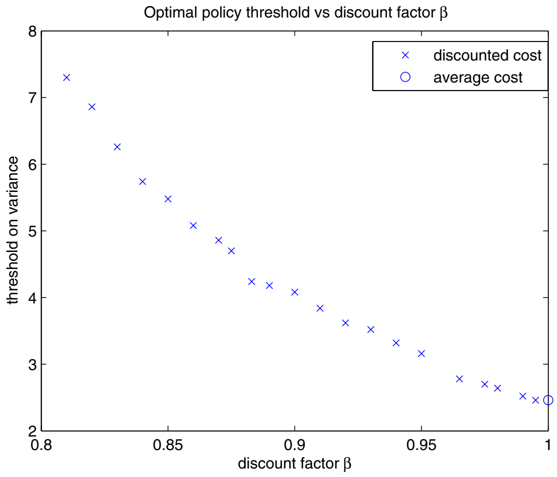

Fig. 1. The optimal policy threshold vs the discounted factor *beta*

V. Acknowledgement

This research was supported in part by the National Science Foundation under Grants ECS-0218207, ECS-0225448 and ECS-0424169, and in part by the Office of Naval Research through the Electric Ship Research and Development Consortium.

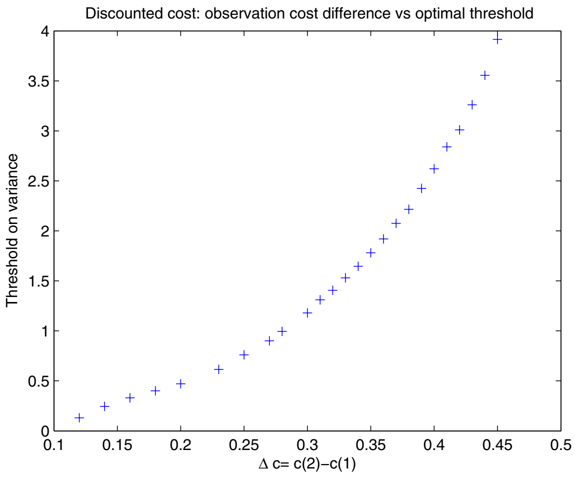

Fig. 2. The cost difference vs the optimal threshold

References

[1] R. Evans A. V. Savkin and E. Skafidas, *The problem of optimal robust sensor scheduling*, Systems and control letters **43** (2001), 149–157.

[2] 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 and Optimization **31** (1993), no. 3, 282–344.

[3] D. Bertsekas, *Dynamic programming: Deterministic and stochastic models*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1987.

[4] J. Hespanha D. Liberzon, *Stabilization of nonlinear systems with limited information feedback*, submitted for review (2004).

[5] E. B. Dynkin and A. A. Yushkevich, *Controlled Markov processes*, Grundlehren der mathematischen Wissenschaften, vol. 235, Springer-Verlag, New York, 1979.

[6] N. Elia and S. Mitter, *Stabilization of linear systems with limit information*, IEEE Trans. on Automatic Control **46** (2001), no. 9, 1384–1400.

[7] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov models, estimation and control*, Applications of Mathematics, vol. 29, Springer-Verlag, New York, 1995.

[8] I. Mareels G. Nair, R. Evans and W. Moran, *Topological feedback entropy and nonlinear stabilization*, IEEE Trans. on Automatic Control **49** (2004), no. 9, 1585–1597.

[9] V. Krishnamurthy and R. J. Evans, *Hidden Markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking*, IEEE trans. on signal processing **49** (2001), no. 12, 2893–2908.

[10] D. Liberzon, *On the stabilization of linear systems with limited information*, IEEE Trans. on Automatic Control **48** (2003), no. 2, 304–307.

[11] G. Nair and R. Evans, *Stabilization with data-rate-limited feedback: Tightest attainable bounds*, Systems and Control Letters **41** (2000), 49–76.

[12] S. Tatikonda and S. Mitter, *Control under communication constraints*, IEEE Trans. on Automatic Control **49** (2004), no. 7, 1056– 1068.

[13] ______, *Control under noisy channels*, IEEE Trans. on Automatic Control **49** (2004), no. 7, 1196– 1201.

[14] W. Wong and R. Brockett, *Systems with finite communication bandwidth constraints I: State estimation problems*, IEEE Trans. on Automatic Control **42** (1997), no. 9, 1294–1299.

[15] ______, *Systems with finite communication bandwidth constraints II: Stabilization with limited information feedback*, IEEE Trans. on Automatic Control **44** (1999), no. 5, 1049–1053.