2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Measurement of a Markov Jump Process: When is it too Costly?

Alan C. O'Connor

***Abstract*— The problem of scheduling measurements of a continuous-time finite-state Markov jump process is considered. This model is relevant to determining the optimal frequency for performing medical exams, inventories, polls, and for many other applications. Unlike the LQG case, where the Kalman filter gain is independent of the measured signals, for this system, the optimal measurement schedule depends on the actual measurements obtained. The feedback rule for generating a measurement schedule is optimized to trade-off between uncertainty about the state of the system and the cost of measurement. Analysis of this trade-off leads to the determination of a critical cost-per-observation above which measurements are too costly, and the best policy is never to observe the system.**

I. Introduction

A partially-observed Markov jump process is a versatile model that may be used to represent a system whose state transitions between a number of discrete configurations. Although the state of the underlying system is hidden, by means of observable outputs, that hidden state may be inferred. The knowledge about the hidden process is summarized by the conditional probability distribution of the underlying state, given the sequence of past observations. We will work with systems where the underlying hidden process evolves in continuous time, independent of any control, but where the process by which observations are obtained may be optimized.

This sort of model can be used to describe optimal search problems, where a target moves between a number of positions according to a known probabilistic rule, and limited resources mean that only a subset of the search space can be inspected at each stage. It may be used to model the progression of a chronic disease in a patient, where the schedule of tests and checkups must be optimized to trade-off between the cost and inconvenience of testing and the cost of uncertainty about the health of the patient. This model may also be used to address the question of how frequently to perform inventories, audits, polls, and many other applications.

A. Past work on control of observations of jump systems

Sondik's papers [1] and [2] presented the original work on optimal control of partially-observed Markov decision processes (POMDP) via dynamic programming. [3] gives a survey of more recent research on average-cost problems for discrete-time controllable Markov processes.

A discrete-time sensor selection problem where the system evolved according to a hidden Markov process was described in [4]. Policies for a finite horizon problems were found via approximate dynamic programming. In [5], the effectiveness of an observation policy for a discrete-time POMDP is discussed in terms of the estimation entropy. [6] and [7] both considered the problem of timing inspection and repair of machinery to maximize the economic gain. The latter paper permitted policies specifying random times-to-inspection, but proved that the best observation strategy is deterministic.

In the field of public health, the design of screening programs for disease involves trading off between the cost and inconvenience of periodic screening and the personal and societal costs of the disease. [8] considered the design of an optimal mass screening protocol. In that paper, the onset of disease was assumed to occur according to a Poisson process, and the objective was to minimize the average disutility associated with the detection delay. In [9], Kirch and Klein optimized the frequency of breast cancer screenings to trade-off between the detection delay and the number of tests adminstered. [10] also described a model for evaluating policies for repeated testing for a medical condition.

B. Contributions of this paper

We consider the problem of designing feedback policies to schedule measurements of a continuous-time Markov jump process. The trade-off between uncertainty and measurement cost is formalized as an infinite-horizon average cost minimization. We consider in detail policies for systems where each costly measurement unambiguously identifies the hidden state. We show that for such problems, the discrete-time process giving the measurement outcomes is a sufficient statistic for the average cost. Using this result, the continuous-time measurement optimization problem is reduced to the selection of a vector of inter-observation waiting times, for which we give a gradient-based optimization method. Surprisingly, there is a finite value of the measurement cost above which no frequency of measurement is beneficial. This feature has not been described previously, as previous investigations in this area have considered discrete-time problems where one observation must be chosen at each stage, or have optimized over finite horizons. Formulæ for the cut-off cost are derived for a number of cases.

This work was completed while the author was at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, and was supported in part by the U.S. Army Research Lab and the U.S. Army Research Office under contract/grant number W911NF-07-1-0376.

Contact: alan.c.oconnor@post.harvard.edu

II. Problem Formulation

A. Partially-observable Markov jump processes

We consider a system described by a continuous-time Markov jump process with state $x(t)$ taking on values in a finite state space. Let n be the number of different states (*i.e.* possible configurations) of the system. These states may be represented by the unit vectors in $\mathbb{R}^n$, which will be denoted by $\{e_1, e_2, \ldots, e_n\}$. The notation e_i' will be used to denote the transpose of a vector e_i.

Transitions or jumps between states may occur at arbitrary times, and the duration of the intervals between jumps are exponential random variables. For a stationary process, the transitions may be modeled by a number of Poisson counters N_{ij} with known, time-invariant rates. The state of the jump process is given by the Itô equation

$$dx = \sum_{i,j} G_{ij}\, x\, dN_{ij}\,. \tag{1}$$

$G_{ij} = (e_j - e_i)e_i'$ corresponds to jumps from e_i to e_j. By construction, the G_{ij} are infinitesimal generators, and if the rates of the counters dN_{ij} are $\lambda_{ij}dt$, then the total rate of transitions are given by the infinitesimal generator

$$A = \sum_{i,j} G_{ij}\lambda_{ij}\,. \tag{2}$$

The entries of the vector $p(t) := \mathcal{E}(x(t))$ give the conditional probability of the hidden process taking on each of the values $\{e_1, \ldots, e_n\}$, given the outcomes of past observations. Without observations, the conditional probabilities $p(t)$ obey the Kolmogorov forward equation:

$$\frac{d}{dt}p(t) = A\,p(t), \quad p(0) = \mathcal{E}(x(0)). \tag{3}$$

We think of $p(t)$ as taking values in the standard simplex Δ_{n-1}, which is the convex hull of the vectors $\{e_1, \ldots, e_n\}$. The vertices of Δ_{n-1} represent certain knowledge of the state $x(t)$. Interior points represent varying degrees of uncertainty. For a more complete exposition of this type of model, we recommend [11].

B. Observations

We assume that observations of the system are instantaneous but occur only occasionally, at times selected by a control policy to be designed. Although the framework we will introduce may be extended to problems with other kinds of observations, in this paper, we will restrict our attention to what we term "momentarily completely observed" processes. This is the case in which each costly measurement determines the state of the underlying jump process exactly.

An example system which is subject to momentary complete observation is the inventory of a warehouse. If the stock in a warehouse is depleted according to a stochastic process and replenished periocally, the uncertainty about the available stocks grows over time. Performing an inventory immediately eliminates any uncertainty about the stocks stored in the warehouse.

C. The measurement schedule

We find it convenient to specify the measurement schedule by a non-decreasing, integer-valued, right-continuous function $\mathcal{S} : [0, \infty) \to \mathbb{Z}^+$. For any $0 \leq t_1 < t_2$, $\mathcal{S}(t_2) - \mathcal{S}(t_1)$ is the number of times that a measurement has been performed on $(t_1, t_2]$. If the cost per measurement is c, the total measurement cost incurred on the time interval $(0, T]$ is $c\,\mathcal{S}(T)$. The time at which the kth measurement is made is given by $\sigma_k = \min\{\, t \;:\; \mathcal{S}(t) = k \,\}$. The time between observations $(\sigma_{k+1} - \sigma_k)$ will be called a "wait time." We also define the useful function $\sigma_-(t) = \min_\tau \{\mathcal{S}(\tau) = \mathcal{S}(t)\}$, which gives the time of the most recent measurement.

D. Feedback observation policies

A practical policy for determining an observation schedule may not depend on the outcome of future observations. If the timing of future observations is permitted to depend on the current value of the conditional probability, nothing further is gained by adding dependence on past observations.

Definition 1. A *Markov observation policy* depends only on the current conditional probability $p(t)$ and not on any past history of observations or past values of $p(t)$.

We will consider only deterministic Markov policies. It may be shown, as was done for a discrete-time replacement problem in [7], that stochastic measurement policies never outperform the best deterministic ones. A deterministic Markov observation policy may be given as a feedback rule by specifying a (possibly empty) subset of the simplex $U \subset \Delta_{n-1}$ as the measurement region. An observation is made whenever $p(t)$ enters U. An example measurement region is shown in Figure 1.

Between observations, the conditional probability obeys the deterministic equation $\dot{p} = Ap$. Thus, a policy given by a measurement region U may be translated into a function $T(p)$ giving, for each point of the simplex, the time to wait until the next measurement. The choice of a measurement region also determines the time of the first observation for any initial condition $p(t_0)$. However, the precise timing of the first observation (assuming it does occur) does not affect the infinite-horizon average cost which we will be considering. For a system with momentary complete observation, a deterministic policy is determined by n wait times, $\{T_i = T(e_i),\ i = 1 \ldots n\}$, where if the underlying system is observed to be in the state e_i, T_i is the time until the next observation.

The two means of specifying a policy–by measurement region or by wait times–emphasize different aspects of a policy. Wait times determine the amount of man-power or expense that will be required to implement a measurement policy. A measurement region more explicitly reveals what the distribution of outcomes of each test will be. The calculations in this paper will primarily involve the determination of optimal values for the wait times.

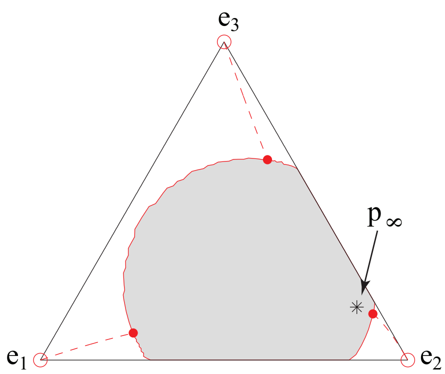

Fig. 1. The simplex Δ_2, showing a circular measurement region in grey. The corresponding wait times $\{T_1, T_2, T_3\}$ are the times required for the trajectories $p(t)$ starting from the vertices $\{e_1, e_2, e_3\}$ to reach U. The steady-state conditional probability that would be obtained without observation is indicated by p_∞.

E. Optimization criterion

In [4] and [12], the running cost function $h(p) = 1 - p'p$, that is 1 minus the norm squared of p, was proposed to penalize uncertainty regarding the state of the hidden system, but both eventually substituted a piecewise affine approximation to this quadratic cost. [5] advocated for minimizing the differential entropy $h(p) = -p' \log(p)$. Both of these functions impose zero cost when the underlying state is certain and maximum cost when $p(t)$ is a uniform distribution. We use the quadratic cost (without approximation) because it enjoys some advantages in terms of calculation.

In the infinite-horizon average cost optimal measurement problem, the feedback policy for generating the measurement schedule $\mathcal{S}$ is to be chosen so as to minimize the expected average cost

$$\eta = \lim_{T\to\infty} \mathcal{E}\left(\frac{1}{T}\int_0^T \Big(h\big(p(t)\big)\,dt + c\,d\mathcal{S}\Big)\right). \tag{4}$$

The expectation is with respect to the outcome of the measurements, which depend on the random underlying jump process.

III. Results

In this section, we state the results for the case of momentary complete observation, in which a costly observation identifies the state of the underlying system at any chosen instant.

A. Computing the measurement outcome probabilities

We now study how the probability distribution of measurement outcomes depends on the choice of measurement policy. The choice of wait times affects not only the running cost incurred on the interval following each observation, but also the relative probabilities of those observations occuring. This is somewhat counterintuitive, as the observed outputs are equal to the state of the underlying Markov jump process, which evolves independent of any control. However, consider the situation where the waiting time T_j after observation of a particular state e_j is very short. Then e_j will be observed many times, even if the underlying system spends an equal fraction of the time in each state.

Let $\mu(k)$ be a vector giving the probabilities of observing a given output at the kth measurement instant: $\mu_i(k) = \Pr\big(x(\sigma_k) = e_i\big)$. We now describe how the discrete-time stochastic process $\mu(k)$ is related to the choice of measurement policy.

By definition, $p_i(t) = e_i' p(t)$ gives the probability of observing $x(t) = e_i$ if the system is inspected at time t. The value of $p(t)$ given that the last observation revealed $x(\sigma_k) = e_j$, is $\Phi(t, \sigma_k)\, e_j = \Phi(t - \sigma_k, 0)\, e_j$. Therefore, the probability of observing $x(\sigma_{k+1}) = e_i$, given that $x(\sigma_k) = e_j$ and that the wait time is $T_j = (\sigma_{k+1} - \sigma_k)$, is $e_i'\,\Phi(T_j, 0)\, e_j$, where $\Phi(t, 0)$ is the solution to the matrix differential equation $\frac{d}{dt}\Phi = A\Phi,\ \ \Phi(0,0) = I$. The evolution of the probabilities of subsequent observations is therefore given by the difference equation

$$\begin{pmatrix}\mu_1(k+1)\\ \vdots \\ \mu_n(k+1)\end{pmatrix} = \underbrace{\begin{pmatrix}\Phi_{11}(T_1) & \cdots & \Phi_{1n}(T_n)\\ \vdots & \ddots & \vdots\\ \Phi_{n1}(T_1) & \cdots & \Phi_{nn}(T_n)\end{pmatrix}}_{\Psi(T_1,\ldots,T_n)} \begin{pmatrix}\mu_1(k)\\ \vdots \\ \mu_n(k)\end{pmatrix}. \tag{5}$$

That is $\Psi_{ij} = e_i' \exp(AT_j) e_j$.

Definition 2. A vector $\bar{\mu}$ giving the steady-state probability that a particular state of the underlying system will be observed will be called the *outcome probabilities.*

A measurement policy leads to set of steady-state outcome probabilities only if the discrete-time imbedded Markov chain given by $\mu(k+1) = \Psi \cdot \mu(k)$ in (5) is ergodic. We now connect ergodicity of the observation outcome process $\mu(k)$ with ergodicity of the underlying jump process $x(t)$.

Lemma 1. *If the jump process $x(t)$ is irreducible, $\Phi(t, 0) = \exp(At)$, the matrix exponential of the infinitesimal generator for $\mathcal{E}(x(t))$, is strictly positive $\forall t > 0$.*

Proof. By construction, the entries of $\Phi(t, 0)$ are probabilities and thus are non-negative. The assumption of irreducibility implies that $\forall i, j = 1 \ldots n$, $\exists T > 0$ such that $\Phi_{ij}(T, 0) > 0$. Suppose there is a time $0 < T_0 < T$ such that $\Phi_{ij}(T_0, 0) = 0$. Consider a sequence $\tau_k \in (0, T_0)$. By the composition rule for matrix solutions, $\forall k,\ \Phi(T_0, 0) = \Phi(T_0, \tau_k) \cdot \Phi(\tau_k, 0)$. Since the entries of both matrices are non-negative, the ijth entry of the left hand side can only be zero if either $\Phi_{ij}(T_0 - \tau_k, 0)$ or $\Phi_{ij}(\tau, 0)$ is also zero. Taking, if necessary, a subsequence of the τ_k, we find that the set of zeros of $\Phi_{ij}(\cdot, 0)$ has an accumulation point in $(0, T_0)$. Then, because the matrix exponential is analytic, $\Phi_{ij}(t, 0)$ must be identically zero for all $t > 0$, contradicting the assumption of irreducibility. □

Theorem 1 (Ergodicity of the observation outcome process)**.** *The matrix $\Psi(T_1, \ldots, T_n)$ defined in (5) is a stochastic matrix that defines a Markov chain $\mu(k+1) =$*

$\Psi(T_1,\ldots,T_n)\cdot\mu(k)$. If A defines an ergodic Markov jump process, then in the case of momentary complete observation, the imbedded process of observed states is ergodic if the vector $(T_1,\ldots,T_n)'$ is strictly positive.

Proof. By construction, each column of $\Psi(T_1,\ldots,T_n)$ is a probability vector, so Ψ is a stochastic matrix. If A defines an ergodic continuous-time Markov jump process, then by Lemma 1, $\exp(AT_i)\,e_i$ is strictly positive for $T_i>0$. Consequently, for $(T_1,\ldots,T_n)'>0$, Ψ is strictly positive as well. A strictly positive stochastic matrix defines a Markov chain that trivially satisfies the requirements for ergodicity. □

Corollary 1. *For a completely observed system with ergodic A, there is a unique $\bar\mu$ that solves $0=(\Psi-I)\bar\mu$.*

Thus, for momentarily completely observed systems, the specification of a feedback measurement policy uniquely determines the steady-state distribution of outcomes of those measurements.

B. Computing the uncertainty cost

Between succeeding observations the evolution of the conditional probability $p(t)$ is deterministic, so it is possible to compute ahead of time what cost will be incurred due to uncertainty. If $h(p)$, defined above, is the function used to penalize uncertainty, the integral

$$\mathcal{I}_i(t)=\int_0^t h(p(t))\,dt,\quad \text{where } p(t)=\exp(At)\,e_i,$$

is the uncertainty cost accumulated since the most recent observation if the underlying system was observed to be in the state $x=e_i$. Given a measurement schedule $\mathcal{S}$, the time interval $[0,\sigma_{N+1})$ may be broken up into the inter-observation periods $[0,\sigma_1)\cup[\sigma_1,\sigma_2)\cup\cdots\cup[\sigma_N,\sigma_{N+1})$. The cost function in (4) may then be rewritten in a form that puts these inter-observation wait times in evidence:

$$\eta=\lim_{N\to\infty}\mathcal{E}\left(\frac{\int_0^{\sigma_1}h(p)\,dt+\sum_{k=1}^N\left(c+\mathcal{I}_{i_k}(\sigma_{k+1}-\sigma_k)\right)}{\sigma_1+\sum_{k=1}^N(\sigma_{k+1}-\sigma_k)}\right).$$

Here $\mathcal{I}_{i_k}(\sigma_{k+1}-\sigma_k)$ is the cost due to uncertainty accumulated between the kth and $(k+1)$st observations. If $\mathcal{S}(t)\not\equiv 0$, we may neglect the finite time-interval before the first observation, and the cost simplifies to

$$\eta=\lim_{N\to\infty}\mathcal{E}\left(\frac{\sum_{k=1}^N\left(c+\mathcal{I}_{i_k}(\sigma_{k+1}-\sigma_k)\right)}{\sum_{k=1}^N(\sigma_{k+1}-\sigma_k)}\right). \quad (6)$$

If we group the subintervals in (6) according to which state was observed most recently, and let N_i be the number of times that e_i is observed out of a total of N observations, we obtain

$$\eta=\lim_{N\to\infty}\mathcal{E}\left(\frac{\sum_i\frac{N_i}{N}\left(c+\mathcal{I}_i(T_i)\right)}{\sum_i\frac{N_i}{N}(T_i)}\right). \quad (7)$$

Theorem 2. *If the underlying process given by A is ergodic, the cost in (7) is equal to*

$$\eta=\frac{\sum_i\bar\mu_i\left(c+\mathcal{I}_i(T_i)\right)}{\sum_i\bar\mu_i T_i}\quad\text{almost surely,} \quad (8)$$

where $\bar\mu(T_1,\ldots,T_n)$ is the vector of observation outcome probabilities given by the invariant measure of the Markov chain defined by (5).

Proof. This follows from Birkhoff's Ergodic Theorem ([14], Thm. 3.55) and Theorem 1. □

C. Solving the optimization via gradient descent

By construction, the columns of Ψ, defined in (5), are probability vectors, so $(\Psi-I)$ is an infinitesimal generator. From the relation $(\Psi-I)\bar\mu=0$, we can derive

$$\left(\Psi-I\right)\frac{\partial\bar\mu}{\partial T_i}=-\frac{\partial\Psi}{\partial T_i}\bar\mu=-A\Psi e_i\bar\mu_i\,. \quad (9)$$

In the completely observed case, if A is irreducible, then Ψ defines an ergodic Markov chain (Theorem 1), so the matrix $(\Psi-I)$ has rank $n-1$. Since $\bar\mu$ is constrained to take on a value in the probability simplex, we may solve for the $n-1$ independent entries of $\bar\mu$ and $\frac{\partial\bar\mu}{\partial T_i}$ using the singular value decomposition of $(\Psi-I)$.

Setting up for convenience the vectors $\bar T=(T_1,\ldots,T_n)$, and $\bar{\mathcal{I}}=(\mathcal{I}_1(T_1),\ldots,\mathcal{I}_n(T_n))$, we rewrite the cost function (8) as $\eta=(c+\langle\bar\mu,\bar{\mathcal{I}}\rangle)/\langle\bar\mu,\bar T\rangle$ and compute the partial derivative with respect to each waiting time T_i:

$$\frac{\partial\eta}{\partial T_i}=\frac{\bar\mu_i\,h(p(T_i))+\left\langle\frac{\partial\bar\mu}{\partial T_i},\bar{\mathcal{I}}\right\rangle}{\langle\bar\mu,\bar T\rangle} \quad (10)$$
$$-\frac{(c+\langle\bar\mu,\bar{\mathcal{I}}\rangle)(\bar\mu_i+\left\langle\frac{\partial\bar\mu}{\partial T_i},\bar T\right\rangle)}{\langle\bar\mu,\bar T\rangle^2}\,.$$

Algorithm 1. Start by selecting a threshold for the maximum wait time, T_{max}, and a stopping condition in terms of the difference between successive iterates, δ.

1) Set an initial value for the policy: $\bar T^{(0)}=(T_1,\ldots,T_n)$ with $T_i>0$.
2) Compute $\frac{\partial\bar\mu}{\partial T_i}$ using (9).
3) Compute $\nabla\eta$ using (10).
4) Let $\bar T^{(k+1)}=\bar T^{(k)}-\epsilon\nabla\eta$.
5) Stop if $\|\bar T^{(k+1)}-\bar T^{(k)}\|<\delta$ or if $\exists\, i$ *s.t.* $T_i^{(k+1)}>T_{max}$. Otherwise, repeat steps 2-5.

Improved method: The most computational intensive part of each iteration is computing the singular value decomposition of $(\Psi-I)$. Alternatively, $\bar\mu$ and $\frac{\partial\mu}{\partial T_i}$ may be computed iteratively, using $\bar\mu_{k+1}=\Psi\cdot\bar\mu_k$ and $\left(\frac{\partial\mu}{\partial T_i}\right)_{k+1}=\Psi\cdot\left(\frac{\partial\mu}{\partial T_i}\right)_k+\left(\frac{\partial\Psi}{\partial T_i}\right)\cdot\bar\mu$, respectively. These iterations both converge to the desired value because each has a unique solution, and by construction, the eigenvalues of Ψ all have absolute value less than or equal to one.

D. Systems with symmetric dynamics

The cases with symmetric dynamics are easiest to analyze. First, we distinguish two types of symmetry.

Definition 3 (Circulant dynamics)**.** The dynamics of an n-state system are called *circulant* if A is nonzero and successive columns are successive cyclic permutations of one another.

Lemma 2. *If a system with circulant dynamics is ergodic, $\mathbf{e}/N$ is the unique fixed point of $\dot{p} = Ap$.*

Proof. Because the system is circulant, the rows of A sum to zero. Thus $\mathbf{e}$ is an eigenvector with eigenvalue zero. By ergodicity, there is only one zero eigenvalue. $\square$

Definition 4 (Completely symmetric dynamics)**.** We call the dynamics of an n-state system *completely symmetric* if all the off-diagonal entries of A are equal:

$$A = \begin{pmatrix} -(n-1)\lambda & \lambda & \cdots \\ \lambda & -(n-1)\lambda & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix},\ \lambda \in \mathbb{R}^+ \qquad (11)$$

Lemma 3. *The nonzero eigenvalues of a completely symmetric system are identical and equal $n\lambda$.*

Proof. Omitted.

Clearly, a completely symmetric system is circulant, but not *vice versa*.

Lemma 4. *For an ergodic circulant system, the integral of the cost with respect to the waiting time, $\mathcal{I}(t) = \int_0^t h(p(\tau))\,d\tau$, is a strictly concave function of t.*

Proof. By Lemma 2, the fixed point of $\dot{p} = Ap$ is also the maximum of $h(\cdot)$ and $p(t)$ converges exponentially to the fixed point. Thus, $h(p(t))$ is strictly monotonic increasing on any time interval. Thus the integral $\mathcal{I}$ is a strictly concave function of the waiting time. $\square$

Theorem 3. *The optimal observation schedule for the average cost problem for a momentarily completely observed system with circulant dynamics is periodic, that is, it must have identical waiting times for each state: $\forall i,\ T_i \equiv \tau$.*

Proof. Since the states are identical, the cost (6) can be simplified to

$$\eta = \lim_{N\to\infty} \mathcal{E}\left(\frac{c}{\frac{1}{N}\sum_{j=1}^N \tau^j} + \frac{\sum_{j=1}^N \mathcal{I}(\tau^j)}{\sum_{j=1}^N \tau^j}\right). \qquad (12)$$

Suppose $\mathcal{S}$ is an arbitrary, not necessarily periodic, observation schedule. The same observation cost is maintained if $\mathcal{S}$ is replaced by a periodic schedule with the same mean time between observations $\bar{\tau} = \frac{1}{N}\sum \tau^j$. However, by the concavity of $\mathcal{I}$ (Lemma 4), the periodic schedule has a lower or equal uncertainty cost. Thus, the optimal schedule is periodic. $\square$

A similar approach was used in [8] to prove the optimality of periodic schedules when screening for disease.

Thus, the optimal feedback observation policy for systems with circulant dynamics is specified by a single parameter, the observation period τ. A first order necessary condition for optimality of a given observation period is that

$$0 = \left.\frac{\partial \eta}{\partial \tau}\right|_\tau = \frac{-1}{\tau^2}\big(c + \mathcal{I}(\tau)\big) + \frac{1}{\tau}h(p(\tau)). \qquad (13)$$

There can be at most one τ satisfying this condition:

Theorem 4. *For systems with circulant dynamics, depending on the value of c, the cost of observation, the cost function $\eta(\tau)$ has either one or no critical point on $\tau \in (0,\infty)$. If τ is a critical point, it minimizes the cost. Thus in the circulant case, the optimizing policy is unique.*

Proof. Rewrite the necessary condition (13) as:

$$\tau \cdot h(p(\tau)) - \mathcal{I}(\tau) = c\,. \qquad (14)$$

The left-hand side of (14) takes on the value 0 for $\tau = 0$. Its derivative is

$$\frac{\partial}{\partial \tau}\big(\tau \cdot h(p(\tau)) - \mathcal{I}(\tau)\big) = \tau\Big\langle \frac{\partial h}{\partial p}, \dot{p}\Big\rangle > 0\,.$$

This derivative is always postive for finite τ, but goes exponentially to zero as $p(t) \to p_\infty$, for which the quadratic penalty function is maximized. Thus the left-hand side of (14) has a horizontal asymptote and is monotonic increasing, which implies that the necessary condition (14) has zero or one solutions, as claimed.

Consider the second derivative of the cost

$$\frac{\partial^2 \eta}{\partial \tau^2} = \frac{2}{\tau^3}(c + \mathcal{I}(\tau)) - \frac{2}{\tau^2}h(p(\tau)) + \frac{1}{\tau}\Big\langle \frac{\partial h}{\partial p}, \dot{p}\Big\rangle.$$

The first two terms cancel at a critical point leaving

$$\frac{\partial^2 \eta}{\partial \tau^2} = \frac{1}{\tau}\Big\langle \frac{\partial h}{\partial p}, \dot{p}\Big\rangle > 0$$

so that τ is a minimum. $\square$

1) Bound on measurement cost for symmetric dynamics: If there is no solution to (14), then the best policy is evidently to permit the system to reach steady state without ever observing. For completely symmetric dynamics, we have the following result, which gives a critical cost per measurement above which the optimal policy is to never observe the system.

Theorem 5. *Optimal policies for completely symmetric n-state systems with transition probability λ, as given in (11), include no observations if $c \geq (n-1)/(2n^2\lambda)$.*

Proof. By Lemma 2, the steady-state conditional probability without observation is $p_\infty = \mathbf{e}/n$. By Theorem 3, the optimal observation policy is periodic, and by symmetry such a policy will observe every state equally often, so $\mu = \mathbf{e}/n$ as well.

Now we compute $\mathcal{I}(T)$, the running cost incurred in the interval between observations. By Lemma 3, all of the nonzero eigenvalues are equal to $n\lambda$. The conditional probability vector evolves according to $p_i(t) = \exp(At)\,e_i = \mathbf{e}/n + \exp(-n\lambda t)(e_i - \mathbf{e}/n)$. Then, $\forall i$, the running cost in the interval following $p(\sigma) = e_i$ is

$$\begin{aligned}\mathcal{I}(T) = \int_0^T &\Big(1 - \big(\mathbf{e}/n + e^{-n\lambda t}(e_i - \mathbf{e}/n)\big)' \\ &\quad \cdot \big(\mathbf{e}/n + e^{-n\lambda t}(e_i - \mathbf{e}/n)\big)\Big)\,dt \\ = &\Big(\frac{n-1}{n}\Big)\cdot T \;+\; \frac{1}{2n\lambda}\big(e^{-2n\lambda T} - 1\big)\cdot\Big(\frac{n-1}{n}\Big)\,.\end{aligned}$$

By (8), the infinite-horizon average cost is almost surely

$$\begin{aligned}\eta &= \frac{c + \langle \mu, \bar{\mathcal{I}} \rangle}{\langle \mu, \bar{T} \rangle} \\ &= \frac{c + \langle \mathbf{e}/n, ((n-1)/n)\big(\bar{T} + (e^{-2n\lambda \bar{T}} - 1)/(2\lambda n)\big)\rangle}{\langle \mathbf{e}/n, \bar{T} \rangle} \\ &= \frac{n-1}{n} + \frac{c + ((n-1)/n)\big(e^{-2n\lambda T} - 1\big)/(2\lambda n)}{T}.\end{aligned}$$

The cost $(n-1)/n$ can be achieved by letting $T \to \infty$, (*i.e.* never observing). For observation to be of benefit, the numerator of the second term must be negative for some range of T, so it is necessary that $c < \frac{n-1}{2\lambda n^2}$. $\square$

2) Cost bound for circulant dynamics: A similar bound holds for systems with circulant dynamics. In this case, the spectrum of A has a much wider range of behaviors. Nevertheless, we obtain the following necessary condition for existence of optimal policies with nonzero observation in terms of the least stable eigenvalue of A.

Theorem 6. *If A is circulant and if $c \geq (n-1)/(2n\nu_{\min})$, where $\nu_{\min}$ is the smallest real part of a nonzero eigenvalue of $-A$, an optimal policy makes no observations.*

Proof. The proof is similar to that of Theorem 5, and is given in [15].

3) General dynamics: For systems with general dynamics, p_∞ may be anywhere in the simplex. The results for symmetric systems given above depended on comparing the cost of observation policies to the steady-state cost without observation $\eta = h(\mathbf{e}/n) = (n-1)/n$. In the next section, we derive a cost bound for an arbitrary two-state system using a slightly different approach. Similar results could, in principle, be obtained for systems with $n \leq 5$, for which closed-form expressions for p_∞ exist.

E. Two-state system

In this section, we specialize some of the results derived previously to the case where the underlying system has two states. Despite its simplicity, this system exhibits interesting behavior.

$$\frac{d}{dt}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} -\lambda_1 & \lambda_2 \\ \lambda_1 & -\lambda_2 \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \tag{15}$$

For the two-state system with unequal transition rates in (15), the stationary feedback policies may be parametrized by two wait times, T_1 and T_2, the durations of the interval between observations following an observation in which $x = e_1$ or $x = e_2$. By Theorem 2, the cost to be minimized (8) equals

$$\eta(T_1, T_2) = \frac{c \ + \ \mu_1 \cdot \mathcal{I}_1(T_1) + \mu_2 \cdot \mathcal{I}_2(T_2)}{\mu_1 \cdot T_1 + \mu_2 \cdot T_2}, \tag{16}$$

where $\mu_1 = \mathcal{E}(N_1/N)$ and $\mu_2 = \mathcal{E}(N_2/N)$ are the steady-state outcome probabilities. As noted previously, the outcome probabilities depend on the choice of the two wait times T_1 and T_2. Parametrize the matrix $\Psi(T_1, T_2)$ in (5) by $\left(\begin{smallmatrix} 1-k_1 & 1-k_2 \\ k_1 & k_2 \end{smallmatrix}\right)$. Then the steady-state outcome probabilities are

$$\mu_1 = \frac{1-k_2}{(1-k_2)+k_1}, \ \ \mu_2 = \frac{k_1}{(1-k_2)+k_1},$$

The overall cost to be minimized, is therefore

$$\eta(T_1, T_2) = \frac{c\big(1-k_2+k_1\big) + (1-k_2) \cdot \mathcal{I}_1(T_1) + k_1 \cdot \mathcal{I}_2(T_2)}{(1-k_2) \cdot T_1 + k_1 \cdot T_2}. \tag{17}$$

If we parametrize the simplex Δ_1 by the second entry of p, the quadratic running cost $h(p) = 1 - p'p$ may be written $h(p_2) = 2p_2(1-p_2)$, and the integrals $\mathcal{I}_1(T_1)$ and $\mathcal{I}_2(T_2)$ can be evaluated explicitly.

The steady-state probability vector for the unequal transition rate system is

$$p_\infty = \begin{pmatrix} \frac{\lambda_2}{\lambda_1+\lambda_2} \\ \frac{\lambda_1}{\lambda_1+\lambda_2} \end{pmatrix} =: \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}. \tag{18}$$

The solution to $\dot{p} = Ap$ starting from $p(0) = e_1$ is

$$p(t) = \begin{pmatrix} s_2 e^{-(\lambda_1+\lambda_2)t} + s_1 \\ -s_2 e^{-(\lambda_1+\lambda_2)t} + s_2 \end{pmatrix}.$$

The integral of the running cost up to time T_1 is

$$\begin{aligned}\mathcal{I}_1(T_1) = 2s_2(1-s_2)\, T_1 &+ \frac{s_2^2(e^{-2(\lambda_1+\lambda_2)T_1} - 1)}{\lambda_1+\lambda_2} \\ &- \frac{2(2s_2^2 - s_2)(e^{-(\lambda_1+\lambda_2)T_1} - 1)}{\lambda_1+\lambda_2}.\end{aligned}$$

The solution starting from $p(0) = e_2$ is

$$p(t) = \begin{pmatrix} -s_1 e^{-(\lambda_1+\lambda_2)t} + s_1 \\ s_1 e^{-(\lambda_1+\lambda_2)t} + s_2 \end{pmatrix},$$

which gives a running cost of

$$\begin{aligned}\mathcal{I}_2(T_2) = 2s_2(1-s_2)\, T_2 &+ \frac{s_1^2(e^{-2(\lambda_1+\lambda_2)T_2} - 1)}{\lambda_1+\lambda_2} \\ &- \frac{2(s_1 - 2s_1 s_2)(e^{-(\lambda_1+\lambda_2)T_2} - 1)}{\lambda_1+\lambda_2}.\end{aligned}$$

Now, k_1 and k_2 are in fact the edges of the measurement region if it is parametrized by p_2 and they are related to the wait times by $T_1 = \frac{1}{\lambda_1+\lambda_2}\log\left(\frac{s_2}{s_2-k_1}\right)$ and $T_2 = \frac{1}{\lambda_1+\lambda_2}\log\left(\frac{1-s_2}{k_2-s_2}\right)$. Substituting $\mathcal{I}_1(T_1)$ and $\mathcal{I}_2(T_2)$ and the formulæ for T_1 and T_2 into (17), we find that the cost to be minimized is

$$\eta = \eta_0 + \frac{k_1(1-k_2)(k_1-k_2-1) + c\,(\lambda_1+\lambda_2)(1-k_2+k_1)}{(1-k_2)\log\left(\frac{s_2}{s_2-k_1}\right) + k_1 \log\left(\frac{1-s_2}{k_2-s_2}\right)} \tag{19}$$

where $\eta_0 = 2s_2(1-s_2)$ is the fixed part of the cost.

Notice from (19) that the cost η_0 can be achieved by letting either threshold k_1 or k_2 equal the steady-state s_2, in which case the system is never observed (excepting an initial transient). So, a constant observation policy is only better than never observing when the second term of (19) is negative. The denominator is always positive, so we can restrict attention to the numerator. Observation can reduce the cost only if, for the given values of c, λ_1, and λ_2, the numerator $k_1(1-k_2)(k_1-k_2-1) + c(\lambda_1+\lambda_2)(1-k_2+k_1)$ takes on negative values somewhere in the rectangle $(k_1, k_2) \in (0, s_2) \times (s_2, 1)$.

Let $\tilde{c} = (\lambda_1 + \lambda_2)c$. The numerator is equal to zero if

$$k_2 = \frac{\tilde{c} + k_1^2 \pm \sqrt{\tilde{c}^2 + (k_1-2)^2 k_1^2 - 2\tilde{c}k_1(k_1+2)}}{2k_1}. \quad (20)$$

The smallest value of c for which the discriminant in (20) vanishes gives the supremum of the costs for which observations can reduce the average cost below η_0. The discriminant is non-positive for $\tilde{c} \geq 2s_2 - 2\sqrt{2}s_2^{3/2} + s_2^2$, which means we have a necessary condition for any constant observation policy to be beneficial:

Theorem 7. *For the completely observed two-state problem, if the cost of observation satisfies the inequality*

$$c \geq \frac{2s_2 - 2\sqrt{2}s_2^{3/2} + s_2^2}{\lambda_1 + \lambda_2}, \quad (21)$$

then the optimal constant observation policy is never to observe. The conditional probability will converge to the steady-state, $p_\infty = (s_1, s_2)'$, *so the cost of this policy is* $h(p_\infty) = 1 - s_1^2 - s_2^2$.

The bound given by (21) is maximized for the completely symmetric system, $\lambda_1 = \lambda_2$. In that case, (21) implies optimal policies make no observations if $c \geq \frac{1}{8\lambda}$, which agrees with the bound in Theorem 5.

Example: Let $\lambda_1 = 1/5$, $\lambda_2 = 1/4$. Then $s_2 = \frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{4}{9}$. Theorem 7 indicates that an optimizing policy will include observations only if the cost per observation satisfies $c < \frac{160}{729}(11 - 6\sqrt{2}) \approx 0.552$. If the cost of observation exceeds this threshold, then the best policy is not to observe, and $\eta = r = 2s_2(1-s_2) = \frac{40}{81} \approx 0.494$. These results are summarized in Figure 2.

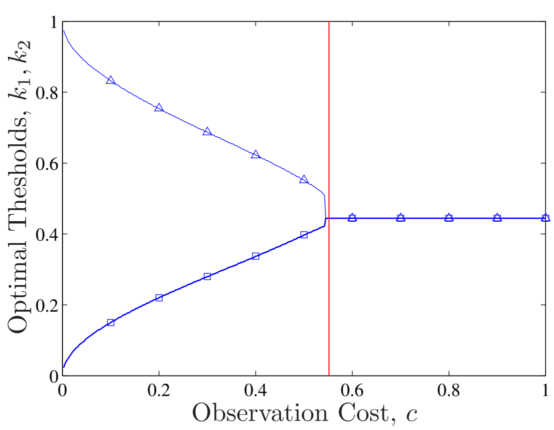

Fig. 2. The plot shows the dependence of the optimal thresholds k_1 (squares), and k_2 (triangles) for when observation of the two-state system with $\lambda_1 = 1/4$, $\lambda_2 = 1/5$ should take place. The optimal constant policy is: 'observe if $prob(x = s_2) \in [k_1, k_2]$'. The thresholds collapse to the steady-state value of p_2 when the observation cost exceeds the critical value ($c \approx 0.552$) marked by the vertical line, indicating that when the cost of observation is too great, the best policy is to allow the conditional probability to reach steady state without ever making an observation.

IV. Discussion and Conclusions

Scheduling costly measurements of a Markov jump process is an optimization problem relevant to real-world applications in the medical field, in search problems, and in scheduling communication between parts of a distributed system. The question of how frequently to observe a system involves a trade-off between the cost of measurement and the cost of uncertainty.

In this paper, we considered systems where each costly observation momentarily observes the full state of the underlying system. Stationary feedback policies for the infinite horizon average cost problem were described completely in terms of the inter-observation wait times, and a fast gradient descent method for optimizing these wait times was given. The global convexity of the optimization problem was proven for the case of circulant dynamics. Our numerical investigations suggest the general problem also has a unique optimizer.

We demonstrated the existence of a critical cost per measurement above which no frequency of observation is beneficial. This critical value depends on the cost of uncertainty and the dynamics of the underlying system.

References

[1] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21(5), pp. 1071–1088, 1973.

[2] E. Sondik, "The optimal control of partially observable Markov processes over the infinite horizon: discounted costs," *Operations Research*, vol. 26(2), pp. 282–304, 1978.

[3] A. Arapostathis, V. Borkar, E. Fernández-Gaucherand, M. Ghosh, and S. Marcus, *Discrete-time controlled Markov processes with average cost criterion: a survey.* Tech Report: University of Maryland & Harvard University, 1991.

[4] V. Krishnamurthy, "Algorithms for optimal scheduling and management of hidden Markov model sensors," *IEEE Trans. Signal Processing*, vol. 50(6), pp. 1382–1397, 2002.

[5] M. Rezaeian, "Sensor scheduling for optimal observability using estimation entropy," *IEEE Intl. Conf. on Pervasive Computing and Communications*, 2007.

[6] I. Savage, "Surveillance problems," *Naval Research Logistics Quart.*, vol. 9(3), pp. 187–209, 1962.

[7] S. Chitgopekar, "A note on the costly surveillance of a stochastic system," *Naval Research Logistics Quart.*, vol. 21(2), pp. 365–371, 1974.

[8] W. Pierskalla and J. Voelker, *A model for optimal mass screening and the case of perfect test reliability.* Tech Report: Northwestern University, 1976.

[9] R. Kirch and M. Klein, "Surveillance schedules for medical examinations," *Management Science*, vol. 20(10), pp. 1403–1409, 1974.

[10] D. Eddy, "A mathematical model for timing repeated medical tests," *Medical Decision Making*, vol. 3(1), pp. 45–62, 1983.

[11] R. Brockett, "Optimal control of observable continuous time Markov chains," in *IEEE Conf. Decision and Control*, pp. 4269–4274, 2008.

[12] B. la Scala, M. Rezaeian, and B. Moran, "Optimal adaptive waveform selection for target tracking," in *Intl. Conf. on Information Fusion*, pp. 552–557, 2005.

[13] L. Zhang and E. Bukas, "Stability and stabilization of Markovian jump linear systems with partly unknown transition probabilities," *Automatica*, vol. 45, pp. 463–468, 2009.

[14] W. Woess, *Denumerable Markov Chains.* European Mathematical Society, 2009.

[15] A. O'Connor, *Optimal Control of Active Sensing Systems*, Ph.D. Thesis, Harvard University, 2011.