COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 10, No. 4, pp. 279-292, 2010

004

OPTIMAL OPEN LOOP MARKOV DECISION RULES MAY REQUIRE PARAMETRIC EXCITATION*

ROGER BROCKETT†

Abstract. We present here a general theory, and give a specific example, showing that there exist time invariant Markov decision problems, with no time variation in the model which, when optimized over an infinite interval, have optimal closed loop control laws that are time varying. Although similar behavior was observed much earlier for specific problems arising in chemical and aeronautical engineering, this work is not applicable to Markov decision problems because of the specific form of the constraints involving the action of the semigroup of stochastic matrices on the standard simplex and the bilinear structure that goes along with rate control for Markov processes. The results given here are especially interesting insofar as they are analogous to the optimal solutions of stochastic control problems associated with Carnot cycles. As in some earlier work, the conditions under which time varying controls are optimal are characterized in terms of the the second variation about a singular solution. In this case the second variation is expressible in terms of a kernel function and conditions under which the second variation is positive definite can be checked by determining if the transform of this kernel is positive real or not.

1. Introduction. A basic problem in Markov Decision theory is that of maximizing the expected value of a time averaged reward through the choice of an infinitesimal generator $A(t)$ subject to $A \in \mathcal{A}$. When the set $\mathcal{A}$ is described as

$$\mathcal{A} = \{A | A = \Big(A_0 + \sum u_i B_i\Big) \ ; \ u(t) \in U \subset \mathbb{R}^m\} \ ; \ U \text{ convex}$$

the problem my be stated as that of maximizing η where

$$\dot{p} = \Big(A + \sum u_i(t) B_i\Big) p \ ; \ \eta = \frac{1}{t_f} \int_0^{t_f} c^T p(t) \, dt \ ; \ u(t) \in U$$

General references and background material on Markov processes can be found in [1] and references [2-3] treat this particular control model. For a fixed value of t_f this problem bears some similarity to the bilinear optimal control problems considered in John Baillieul's 1975 thesis [4]. The important difference is that here the matrices A and B_i are not skew-symmetric, as they were in his work, but rather they are such that the entries of $A + \sum u_i B_i$ are nonnegative off the diagonal and the entries in each column sum to zero. This implies that if $p(0)$ lies in the so-called *standard simplex* consisting of vectors with nonnegative entries whose components sum to one then p remains in this set for all time. Put another way, in Baillieul's work the relevant geometric objects were the unit sphere and the action of the orthogonal group on it whereas here the corresponding objets are the standard simplex and the action of the semigroup of stochastic matrices on it.

*Dedicated to John Baillieul on the Occasion of His 65th Birthday.

†Harvard University, E-mail: brockett@seas.harvard.edu

We limit our attention to a special case of the situation considered in [1] and [2]; here there is no cost term associated with the choice of control, although we do put limits on the values $u(t)$ can take on. In such cases the control can be expected to consist of bang-bang segments and singular arcs. In the wider subject of optimal control, singular arcs only exist under rather special circumstances but what we see here is that the singular arcs associated with constant controls are not so exceptional. As we will show, looking for constant singular arcs comes down to looking for the solutions $p(u)$ of $(A + \sum u_i B_i)p(u) = 0$ which maximize $c^T p(u)$.

For the infinite time problem with

$$\eta = \lim_{t_f \to \infty} \frac{1}{t_f} \int_0^{t_f} c^T p(t)\, dt$$

the optimal control can not be unique; changes in $p(\cdot)$, limited to a finite interval, have no effect on the infinite time average. If we fix a value of $t_f >> 1$ and consider the problem of maximizing

$$\eta = \int_0^{t_f} c^T P\, dt$$

subject to the periodicity condition condition $p(0) = p(t_f)$, then one might expect that the optimal solution would be unique and that it would closely approximate a solution to the infinite time problem. However, there is no reason to expect that there is a period p which is optimal for the original problem, even though A and $B_1, B_2, ..., B_k$ are constant. (See reference [5].) It would be necessary to exclude the possibility that an optimal control is the sum of periodic terms with incommensurate frequencies, etc. We will see that it can happen that the optimal control, when expressed in feedback form, is time varying and we will give a general criterion for this to be the case. In a wider context there is a fairly large literature on the question of when periodic controls are, and are not, optimal for time invariant systems [6-8] but we are unaware of any such work directly applicable to MDP. Earlier work on the randomized controls [9] and our work on the second law of thermodynamics and the Carnot cycle [10] are, however, suggestive of what is seen here.

Notation: We use e_i, for $i = 1, 2, ..., n$ to denote the standard basis vectors in $\mathbb{R}^n$ and let e denote the vector of all ones, $e = \sum e_i$. The $n-1$-dimensional manifold with boundary defined by the convex hull of the points $\{e_1, e_2, ..., e_n\}$ will be written as Δ^{n-1}, or simply Δ if the dimension is clear. That is, Δ is just the standard simplex defined above. If A is a square matrix whose the entries are nonnegative off the diagonal and whose columns sum to zero we will call it an *infinitesimal generator*. An infinitesimal generator is always singular; if its null space is one dimensional we will say that it is *irreducible*.

Remark: Notice that the the vector c which enters into the definition of the performance measure influences the optimal solution only insofar as it differs from a

multiple of e. Because $e^T p = 1$, regardless of the value of $p \in \Delta$, adding a multiple of e to c only adds a constant to η and does not change the optimal policy. We will say that c is in *reduced form* if $e^T c = 0$ and this condition can be assumed without loss of generality.

2. An Example. We begin with an example chosen to illustrate particular aspects of the type of problems we are investigating. Consider the differential equation and performance measure

$$\dot{p} = (A + uB)p \ ; \ \eta = \lim_{t_f \to \infty} \frac{1}{t_f} \int_0^{t_f} c^T p \, dt$$

with A and B given by

$$A = \begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \ ; \ B = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

and $c^T = [0,\ 1,\ 0\ 0]$. The objective is to maximize η. Clearly $A+uB$ is an infinitesimal generator of a continuous time Markov process for $u(t) \in [0, \infty)$ or any subinterval of $[0, \infty)$.

If M is an infinitesimal generator will refer to a probability vector satisfying

$$M\bar{p} = 0$$

as an *invariant distribution for* M. A short calculation shows that for the above choice of $A + uB$ the invariant distribution is a function of u given by

$$\bar{p}_i = \frac{(1+u)^{4-i}}{4 + 6u + 4u^2 + u^3} \ ; \ i = 1, 2, 3, 4$$

For $u = 0$ we see that $\bar{p} = [1/4, 1/4, 1/4, 1/4]^T$ and for $u >> 1$, $\bar{p} \approx [1, 0, 0, 0]^T$. Thus a short period over which the integral of u is large will drive p close to the vector e_1 and a lengthy period over which u is near zero causes p to approach $[1/4, 1/4, 1/4, 1/4]$. l. In this sense, a large value of u acts as a "reset", driving all the probability to state one whereas when $u = 0$ the distribution relaxes, to the uniform distribution.

The left-hand panel of figure 1 shows the graph of $\bar{p}_2$ as a function of u. An analysis of the function $p_2 = (1+u)^2/(4 + 6u + 4u^2 + u^3)$ shows that the maximum value of p_2 is about .277 which corresponds to $u \approx .52$. The steady state probability vector corresponding to this choice of u is $p_\infty \approx [.421, .277, .182, .120]^T$.

We will discuss in more detail the application of the maximum principle to this type of problem in section 5, but for now we observe that because there is no penalty on u and because u enters the hamiltonian linearly it is reasonable to expect that the optimal control for this problem is either constant or bang-bang. With this in

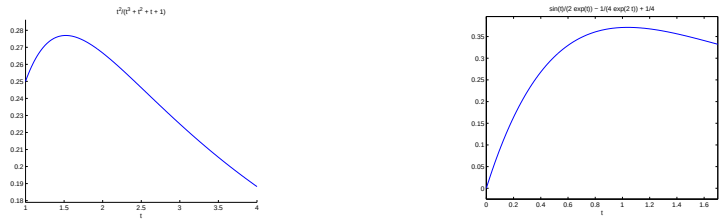

FIG. 1. *Left. A plot of the steady state value $\bar{p}_2$ as a function of $1+u$. Right. A plot of p_2 as a function of time for $\dot{p} = Ap$ and $p(0) = e_1$*

mind, we now construct for this example, a time varying control that gives a better performance than the best constant control. As we noted above, in a limiting sense, a large value of u will send p to e_1. On the other hand, a calculation shows that for $u = 0$ and initial condition $p(0) = e_1$ we have

$$p_2 = \frac{1}{2}e^{-t}\sin t + \frac{1}{4}(1 - e^{-2t})$$

The integral of p_2 from zero to t is

$$\int_0^t p_2(\sigma)\, d\sigma = \frac{1+t}{4} - \frac{1-e^{-2t}}{8} - \frac{e^{-t}}{4}(\cos t + \sin t)$$

The right-hand panel of figure one shows the graph of

$$\gamma(t_f) = \frac{1}{t_f}\int_0^{t_f} p_2(t)\, dt$$

as a function of t_f. It is the contribution to η that accrues if the $p(0) = e_1$ and the control $u = 0$ is used on $[0, t_f]$. The maximum value of $\gamma(t_f)$ is about .3, corresponding to $t_f \approx 2$. Now consider a periodic control policy consisting of an alternation between two types of segments. On one segment of length 2 we let $u = 0$. On the other segment, of very short duration, we let u be very large, effectively driving the state back to e_1. Repeating these steps infinitely often results in a performance which is about .30. Thus we see that there is a time varying control that gives a larger payoff than the best constant control.

Although this is not essential for the purpose of showing that a time varying policy can be preferable to the best constant policy, we point out that the result is not the consequence of some standard degeneracy. The linearized approximation of this system, obtained by linearizing about an equilibrium point, $(A + u_0 B)p_0 = 0$ is

$$\dot{\delta} = (A + u_0 B)\delta + vBp_0$$

This system is controllable for values of u such that

$$W(u) = [Bp_0, (A+uB)Bp_0, (A+uB)^2 Bp_0, (A+uB)^3 Bp_0]$$

is of rank three, so that the given vectors span the three dimensional tangent space of Δ. Expressing p_0 as $(ee^T + (A+uB)^T(A+uB))^{-1}e$ we see that the three-by-three principle minor is a rational function of u which is nonzero for $u = 0$ and hence non zero for all but a finite number of values of u. However, one may observe that the system is not generic in that typically two four-by-four Markov matrices will generate a 12-dimensional Lie algebra but the pair A, B only generate a five dimensional algebra. This algebra consists of the linear span of A, B, and $(e_1 - e_2)e^T, (e_2 - e_3)e^T$ and $(e_3 - e_4)e^T$. The situation with respect to observability based on observing $c^T p$ can be analyzed similarly with similar results.

By way of contrast, the performance of the three dimensional problem having the same basic reset-diffuse possibiliies as the problem just considered,

$$A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \; ; \; B = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

can not be improved using time variation.

3. Optimal Constant Policies. We begin by developing a pair of expressions for the invariant distribution associated with affine families of irreducible infinitesimal generators of the form $A + \sum u_i B_i$. Recall our definition, $e = \sum e_i$. If $Ap_0 = 0$ and $p_0 \in \Delta$ then

$$\begin{bmatrix} e^T \\ A \end{bmatrix} p_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Premultiplication by $[e \; A^T]$ gives $(ee^T + A^T A)p_0 = e$. Irreducibility implies that $(ee^T + A^T A)$ is invertible and so

$$p_0 = (ee^T + A^T A)^{-1} e$$

There is an alternative approach that is also useful because it avoids introducing second degree terms in A. If $Ap_0 = 0$ with $p_0 \in \Delta$ then we have

$$\begin{bmatrix} 0 & e^T \\ e & A \end{bmatrix} \begin{bmatrix} 0 \\ p_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \implies \begin{bmatrix} 0 \\ p_0 \end{bmatrix} = \begin{bmatrix} 0 & e^T \\ e & A \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

where, again, irreducibility implies that the indicated inverse exists.

One way to define the Moore-Penrose inverse of a matrix A is to define it as the limit

$$A^+ = \lim_{\epsilon \to 0} (\epsilon I + A^T A)^{-1} A^T$$

If A is an irreducible infinitesimal generator its range space contains all vectors whose components sum to zero; the null space of A consists of the multiples of the invariant

distribution. The null space of A^T is consists of the multiples of e. For any B such that $e^T B = 0$ we have

$$\begin{bmatrix} 0 & e^T \\ e & A \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ B \end{bmatrix} = \begin{bmatrix} 0 \\ A^+ B \end{bmatrix}$$

Now suppose that we are given k matrices, $B_1, B_2, ..., B_k$ all of whose columns sum to zero, and suppose we wish to determine the choice of u that maximizes $c^T p$ for $(A + \sum u_i B_i)p = 0$; $p \in \Delta^{n-1}$, and $u \in U$. Starting from the expression

$$\eta(u) = \begin{bmatrix} 0 & c^T \end{bmatrix} \begin{bmatrix} 0 & e^T \\ e & A + \sum u_i B_i \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and then differentiating with respect to u_i, we have

$$\frac{\partial \eta}{\partial u_i}\Big|_{u=0} = \begin{bmatrix} 0 & -c^T \end{bmatrix} \begin{bmatrix} 0 & e^T \\ e & \bar{A} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ 0 & B_i \end{bmatrix} \begin{bmatrix} 0 & e^T \\ e & \bar{A} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

where $\bar{A} = A + \sum u_i B_i$. This allows us to express the first order necessary conditions for $u = 0$ to be a stationary point as

$$\frac{\partial \eta}{\partial u_i}\Big|_{u=0} = \begin{bmatrix} 0 & -c^T \end{bmatrix} \begin{bmatrix} 0 & e^T \\ e & \bar{A} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ B_i p_0 \end{bmatrix} = 0$$

If A is an irreducible infinitesimal generator then $e^T A$ is necessarily zero and $Ax = y$ can be solved uniquely for x subject to the conditions $e^T y = 0$ and $p_0^T x = 0$. The solution obtained under these circumstances coincides with $x = A^+ y$. Thus we can express the derivative more succinctly as

$$\frac{\partial \eta}{\partial u_i}\Big|_{u=0} = c\bar{A}^+ B_i (ee^T + A^T A)^{-1} e = -c\bar{A}^+ B_i p_0$$

Similarly, for the second derivative,

$$\frac{\partial^2 \eta}{\partial u_i u_j}\Big|_{u=0} = c(\bar{A}^+ B_i \bar{A}^+ B_j + \bar{A}^+ B_j \bar{A}^+ B_i) p_0 = q_{ij}.$$

The following elementary result serves to summarize this background material.

THEOREM 1. Let $U \subset \mathbb{R}^m$ be a a closed convex set. Assume that for all $u \in U$ the matrix $A + \sum u_i B_i$ is an irreducible infinitesimal generator and suppose that $c \neq 0$ is in reduced form. Then for each $u \in U$ there exists a unique $p(u) \in \Delta^{n-1}$ such that $(A + \sum u_i B_i)\, p(u) = 0$. If $u = \bar{u}$ lies in the interior of U and $\bar{A} = A + \sum \bar{u}_i B_i$, then the two conditions

$$c^T \bar{A}^+ B (ee^T + \bar{A}^T \bar{A})^{-1} e = 0$$

$$Q \leq 0$$

are necessary for $\bar{u}$ to maximize $c^T p(u)$, where

$$q_{ij} = c(\bar{A}^+ B_i \bar{A}^+ B_j + \bar{A}^+ B_j \bar{A}^+ B_i)(ee^T + \bar{A}^T \bar{A})^{-1} e$$

Proof. Because U is compact and $\eta(u)$ is continuous, the existence of a maximum is obvious. As we have seen, the invariant probability distribution is $(ee^T + A^T A)^{-1} e$. The maximizing value of u might or might not be unique and might or might not lie on the boundary of U. If it occurs in the interior, the first order necessary conditions imply that $c^T A^+ B_i p = 0$. The second order terms will not increase η if the Hessian is negative semidefinite.

There is a significant distinction between problems for which the optimal constant value of u lies on the boundary of the admissible set U, and those for which u lies in the interior of U. In section six we give a criterion which allows one to determine when a constant interior control can be improved by parametric excitation and in section seven we will compare the optimal closed loop performance with the optimal open loop performance. In that context we will need to understand the effect of a rank one perturbation on an invariant distribution.

Assume, initially, there is just one control u and that B is rank one. Consider the one parameter family of irreducible infinitesimal generators, $A + uB$ and the vector $p(u) \in \Delta$ that satisfies $(A + uB)p(u) = 0$. Writing B as an outer product, $B = gh^T$ we see that

$$(ee^T + A^T A)p + uA^T g h^T p = e$$

Thus,

$$p(u) + uA^+ g h^T p(u) = p(0)$$

Premultiply both sides by h^T and make the definitions $\alpha = h^T A^+ g$ and $h^T p = \beta$ to get $h^T p(u) + u\alpha h^T p(u) = \beta$. If $\beta = 0$ the invariant distribution associated with $A + ugh$ is independent of u. If $\beta \neq 0$ we must have $u\ \alpha \neq -1$ and

$$h^T p(u) = \frac{\beta}{1 + u\alpha}$$

Using this expression in the equation for p we get

$$p(u) = p(0) - \frac{\beta u}{1 + \alpha u} A^+ g$$

Finally, if we let $\gamma = c^T A^+ g$ we have

$$c^T p(u) = c^T p(0) - \frac{\beta \gamma u}{1 + \alpha u}$$

From this expression we see that $c^T p(u)$ is independent of u if $c^T A^+ g = 0$ and that if $\beta\gamma \neq 0$ the maximum is achieved with a u lying on the boundary of U.

The following theorem extends this analysis to the situation in which there are several u_i with the corresponding B_i of rank one.

THEOREM 2. Let $U \subset \mathbb{R}^m$ be a closed convex set of the form $\{u | a_i \leq u_i \leq b_i \ ; \ a_i < 0 < b_i; i = 1, 2, ..., m\}$ and let $B_1, B_2, ..., B_k$ be a set of rank one matrices. Suppose that $A + \sum u_i B_i$ is an irreducible infinitesimal generator for all u in U and that $c \neq 0$ is in reduced form. Then there is choice of u that maximizes $c^T p(u)$ and lies at a vertex of U.

Proof. As noted above, for each $u \in U$ there is a unique $p(u) \in \Delta^{n-1}$ satisfying $(A + \sum u_i B_i)p = 0$. Suppose that $\hat{u}$ is an optimal value of u. Consider the dependence of $c^T p_0$ on u_1. From the analysis just given we see that

$$c^T p(u) = f + \frac{\beta_i \gamma_i u_1}{1 + \alpha_1 u_1}$$

where $f, \alpha_1, \beta_1, \gamma_1$ may depend on any of the inputs $u_2, u_3, ..., u_k$ but not u_1. As above, if $\beta_1 \gamma_1 = 0$ then $c^T p_0$ does not depend on u_1 and u_1 can be placed on the boundary without changing the value of η or any of the other components of u. (This is where we use the hypothesis that the boundaries of U are aligned with the coordinate axes.) Otherwise, $\beta_1 \gamma_1 \neq 0$ and there are no local maxima of $c^T p_0$ off the boundary of U. Repeating this argument for each component individually we see that there can be no local maxima in the interior, unless they are associated with a component of u which does not affect $c^T p_0$.

We note briefly the following result which provides additional insight about the conditions which lead to optimal solutions off the boundary. The hypothesis is both weaker (no longer rank one) but also stronger (the sign condition). It gives more insight about the kind of conflicting effects that must be present in B if one is to have an interior optimal solution.

THEOREM 3. Suppose that $A + uB$ is an irreducible infinitesimal generator for $a \leq u \leq b$ and that $c \neq 0$ is in reduced form. Suppose that all the eigenvalues of $A^+ B$ are real and that

$$A^+ B = \sum_{i=1}^{k} \xi_i \chi_i^T$$

with k being the rank of $A^+ B$. Then if $c^T A^+ \xi_i \chi_i^T p(0)$ have the same sign for all $i = 1, 2, ..., k$ there is a choice of u that maximizes $c^T p(u)$ and lies on the boundary of U.

Proof. From the developments given above we see that the performance measure has a partial fraction expansion of the form

$$c^T p(u) = c^T p(0) - \sum \frac{\beta_i \gamma_i u}{1 + \alpha_i u}$$

with

$$\alpha_i = \chi_i^T A^+ \xi_i \ \ ; \ \ \beta_i = \chi_i^T p_0 \ \ ; \ \ \gamma_i = c^T A^+ \xi_i$$

The hypothesis implies that residues all have the same sign and so the function has no relative maxima.

4. The Maximum Principle. The previous section can be viewed as a preamble to an investigation of what the maximum principle implies about the optimal solution to this class of problems. If instead of the optimal control problem defined in the introduction, we fix a value for T, the maximum principle asserts that there exists a nonzero pair (q, q_0) with q a vector and q_0 a nonnegative scalar, satisfying further conditions related to

$$h(p, q, , u) = q^T(A + \sum u_i B_i)p + q_0 c^T p$$

These conditions are as follows. In additional to p which satisfy the original equations of motion

$$\dot{p} = Ap + \sum u_i B_i p$$

the dual variable q must satisfy the costate equation

$$\dot{q} = -\left(A + \sum u_i B_i\right)^T q + q_0 c \ ; \ q_0 \geq 0$$

and the optimal $u_i(t)$ must maximize the Hamiltonian along optimal trajectories.

If we take U to be closed and convex then, because h is linear in u, we see that whenever $q^T B_i p$ is nonzero the control u_i takes on a value lying in the boundary of U. However, if there are solutions to the state-costate equations such that $q^T Bp$ vanishes over some interval then interior values of u may be part of an optimal control. A solution (p, q, u) defined on an interval over which $q^T B_i p \equiv 0$ for some i is called a *singular arc*.

We now reexamine the constant solutions described in Theorem 1 to determine when they are singular arcs. For p to be a constant solution we require $A+\sum u_i B_i)p = 0$ and for q to be a constant solution we require $(A + \sum u_i B)_i^T q = q_0 c$. However, because $(A + \sum u_i B_i)^T e = 0$ and, assuming that c is in reduced form, the equation for (p, q) can be expressed as

$$\begin{bmatrix} 0 & e^T \\ e & A + \sum u_i B_i \end{bmatrix} \begin{bmatrix} 0 \\ p_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & e^T \\ e & (A + \sum u_i B_i)^T \end{bmatrix} \begin{bmatrix} \alpha \\ q \end{bmatrix} = \begin{bmatrix} 0 \\ q_0 c \end{bmatrix}$$

These equations are decoupled and for a given value of u have a unique solution.

THEOREM 4. Let $\bar{A} = A + \sum u_i B_i$. Under the assumptions of Theorem one, the constant functions

$$p(t) = (ee^T + \bar{A}^T \bar{A})^{-1} e \ ; \ q(t) = (ee^T + \bar{A}^T \bar{A})^{-1} q_0 Ac$$

satisfy the state-costate equations. If t_f finite,

$$c^T \bar{A}^+ B(ee^T + A^T A)^{-1} e = 0$$

and u lies in the interior of U then such solutions define a singular arcs.

Proof. Because we are assuming that $\bar{A}$ is irreducible, $ee^T + \bar{A}^T\bar{A}$ is nonsingular and $\bar{A}(ee^T + \bar{A}^T\bar{A})^{-1}e = 0$. To show that the constant solution $q(t) = q_0(ee^T - \bar{A}^T\hat{A})^{-1}Ac$ satisfies the costate equation we set $\dot{q} = 0$ in the costate equation and premultiply by $(A = \sum u_i B_i)$. Noting that $e^T c = 0$ we get the result claimed. The solution then satisfies all the conditions of the maximum principle if for $i = 1, 2, ..., k$

$$c^T A^T(ee^T + \bar{A}^T\bar{A})^{-1}B_i(ee^T + \bar{A}^T\bar{A})^{-1}e = c(\bar{A}^+)^T B_i p_0 = 0$$

5. The Second Order Analysis. We have seen that a constant control corresponding to a value of u in the interior of U necessarily generates a singular arc if it is a local optimizer in the class of constant controls. By computing the second variation about this solution we can hope to determine wether or not it is also a local maximum in the wider class of time varying controls. In section 4 we computed the second derivative with respect to constant changes in u. Here we carry out the second analysis with respect to more general perturbations. As it happens, the results are more transparent for the infinite time problem and we limit our analysis to that situation.

For the sake of readability, we first analyze the situation in which u is one dimensional. Starting with the equilibrium solution (p_0, u_0) corresponding to $(A + uB)p_0(u) = 0$, introduce $\delta = p - p_0(u)$. After integrating both sides of $\dot{\delta} = A\delta + vBp_0(u) + vB\delta$ we get

$$\delta(t) = A\int_0^t \delta(\sigma)d\sigma + Bp_0(u)\int_0^t v(\sigma)d\sigma + \int_0^t v(\sigma)B\delta(\sigma)d\sigma$$

Of course δ is bounded and so

$$\lim_{t_f\to\infty}\frac{\delta(t_f) - \delta(0)}{t_f} = 0$$

Thus

$$0 = \lim_{t_f\to\infty}\frac{1}{t_f}\left(A\int_0^{t_f}\delta(\sigma)d\sigma + Bp_0(u)\int_0^{t_f}v(\sigma)d\sigma + \int_0^{t_f}v(\sigma)B\delta(\sigma)d\sigma\right)dt$$

Using an over-bar to denote time averages this can be written as

$$0 = A\bar{\delta} + Bp_0(u)\bar{v} + \lim_{T\to\infty}\frac{1}{t_f}\int_0^{t_f}v(\sigma)B\delta(\sigma)d\sigma dt$$

Because δ is first order in v the last term is second order in v. Thus, to first order in v we have $c^T\bar{\delta} = c^T A^+ Bp_0(u)\bar{v}$ and, from the first order necessary conditions, this term vanishes.

Assuming that η is stationary with respect to first order changes, we can drop the terms that are linear in v and restrict our attention to

$$\delta(t) = A\int_0^t \delta(\sigma)d\sigma + \int_0^t v(\sigma)B\delta(\sigma)d\sigma$$

Replacing δ in the nonlinear term by its first order approximation gives

$$\delta(t) = A\int_0^t \delta(\sigma)d\sigma + \int_0^t v(\sigma)B\int_0^\sigma e^{A(\sigma-\tau)}Bp_0v(\tau)\ d\tau\ d\sigma + \epsilon(t)$$

where ϵ is order three or higher.

As is easily seen, If $\bar{u}$ is the average value of u then the average value of the solution of an asymptotically stable linear system $\dot{x} = Ax + bu$ is given by $\bar{x} = A^{-1}\bar{u}$. Slightly modifying this to fit our present circumstances, we note that if A has a simple eigenvalue at 0 with all other eigenvalues having negative real parts and if b lies in the range space of A, then we have $\bar{x} = -A^{+}b\bar{u}$. In terms of the definition

$$r(t) = \int_0^t v(\sigma)B\int_0^\sigma e^{A(\sigma-\tau)}Bp_0v(\tau)\ d\tau\ d\sigma$$

we have

$$c^T\bar{\delta} = c^TA^{+}\bar{r}$$

provided that the indicated average exists. This places in evidence the role of the kernel function

$$w(\sigma,\nu) = c^TA^{+}Be^{A(\sigma-\nu)}Bp_0$$

Quadratic functionals of this form play an important role in various areas of mathematics and certainly in system theory. Its definiteness properties are conveniently characterized in terms of its Laplace transform and well known tests for the positive realness. In concrete terms, if the Laplace transform of $w(t,0)$ is $\phi(s)$ and if at some frequency ω we have $\mathcal{R}e\ \phi(i\omega) > 0$ then by letting $v(t) = \epsilon\sin\omega t$ with $|\epsilon|$ small we will improve on c^Tp_0. On the other hand, u_0 is maximizing if this function is negative definite. If it is nonzero and neither negative definite nor negative semidefinite then u_0 is not maximizing.

With a little additional effort we can generalize this to the multiple input situation.

THEOREM 5. Let A, B_i, U and c be as in Theorem 1, and consider the optimization problem defined there. Suppose $\bar{u} \in U$ is constant and that for $\bar{A} = A + \sum \bar{u}_iB_i$ we have $c^T\bar{A}^{+}Bp(\bar{u}) = 0$. Consider the matrix W whose ij^{th} entry is

$$w_{ij}(t) = -c^T\left(B_ie^{\bar{A}t}B_j + B_je^{\bar{A}t}B_i\right)p_0$$

If $\bar{u}$ belongs to the interior of U then $\bar{u}$ is optimal in the class of time varying open loop controls if the matrix $W(t-\sigma)$ is positive definite in the sense that a^TWa defines a positive definite functional for all constant vectors a. It is not optimal if for some choice of a the kernel a^TWa is neither positive definite nor positive semidefinite.

Proof. The multivariable version of the second order change in η that results from changing u_0 to $u_0 + v$ depends on

$$r_{ij}(t) = \int_0^t v_i(\sigma)B \int_0^\sigma e^{A(\sigma-\tau)}Bp_0v_j(\tau)\ d\tau\ d\sigma$$

As above, this leads to

$$c^T\bar{\delta} = c^TA^+\bar{r}_{ij}$$

Define the matrix

$$w_{ij}(t-\sigma) = c^TA^+B_ie^{A(t-\sigma)}B_jp_0$$

where p_0 is the invariant distribution corresponding to u_0. We can then express the effect of a change in input as

$$\delta\eta = \lim_{t_f\to\infty}\frac{1}{t_f}\int_0^{t_f}\int_0^t \sum_{i,j} v_i(t)w_{ij}(t-\sigma)v_j(\sigma)\ d\sigma\ dt$$

If all of these are negative definite then u_0 indeed represents a local maximum. If any fail to be at least negative semidefinite then the solution u_0 can not be optimal.

6. Optimal Closed Loop Solutions. In our papers [2-3] we have described the the optimal closed loop control for the class of Markov decision problems treated here, with a more general class of performance measures. It is of interest to compare the open loop control, which makes no use of any observation of the state, and the closed loop control with perfect observation found there. Suppose that the states are labeled $x_1, x_2, ..., x_n$ and that $p_i(t)$ is the probability that the system is in state i at time t. Recall the special role played by rank one matrices in Theorem 2. If each of the B_i is of the form $B_i = g_ie_j^T$ then there is no difference between the performance of open loop and closed loop control provided that U is of the form $U = \{u|a_i \le u_i \le b_i\}$. In this situation the value of u_i does not influence the transition rates except when the system is in state x_j. Because the control only influences the evolution of the state when it is in one particular state it can be considered to be an open loop control or a closed loop control and Theorem 2 makes it clear that the optimal control will lie on the boundary.

7. Diffusion Processes. We have focused our attention here on finite state Markov processes but the basic ideas are also applicable to a class of stochastic control problems involving diffusion processes. In principle, for these problems the operators

are infinite dimensional but there are important classes of problems for which the analysis can be reduced to a finite dimensional situation. In order to convey a flavor of how this happens we give an example involving a diffusion process in $\mathbb{R}^2$.

Consider the Itô equation with a multiplicative control

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \left(\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} + u \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix} dt + \begin{bmatrix} dw \\ 0 \end{bmatrix}$$

In this case the probability density satisfies the u-dependent Fokker-Plank equation

$$\frac{\partial \rho(t,x,y)}{\partial t} = \frac{\partial (x - ux - uy)\rho(t,x,y)}{\partial x} + \frac{\partial (y + uy + ux)\rho(t,x,y)}{\partial y} - \frac{1}{2}\frac{\partial^2 \rho(t,x)}{\partial x^2}$$

Suppose we wish to minimize the average value, of the variance over $[0,\infty)$,

$$\eta = \lim_{t_f \to \infty} \frac{1}{t_f} \int_0^{t_f} \int_{-\infty}^{\infty} x^2 \rho(t,x) dx$$

subject to the constraint $|u(t)| \leq 2$

If u is identically zero then ρ approaches a zero mean Gaussian with variance $1/2$. On the other hand, if $u = -1$ the steady state variance is $1/4$ and this is the smallest value of σ_{11} that can be obtained using a constant control. The obvious question is then, can we do better using a time varying control?

It is convenient to work directly with the given stochastic equation rather than the equation for the density. For the linear stochastic equation

$$dx = (A + uB)x dt + R dw \;\; ; \;\; \eta = \lim_{t_f \to \infty} \frac{1}{t_f} \int_0^{t_f} c^T \Sigma(\sigma) c \, d\sigma$$

we have

$$\dot{\Sigma} = (A + uB)\Sigma + \Sigma(A + uB)^T + RR^T$$

If we wish to minimize the average value of some linear functional of the variance then here again, the control enters the hamiltonian linearly and the situation is much the same as we had above.

Continuing with the example, the variance equation can be written as

$$\frac{d}{dt} \begin{bmatrix} \sigma_{11} \\ 2\sigma_{12} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} -2 + 2u & u & 0 \\ -2u & -2 & 2u \\ 0 & -u & -2 - 2u \end{bmatrix} \begin{bmatrix} \sigma_{11} \\ 2\sigma_{12} \\ \sigma_{22} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

From this we see that in steady state we have

$$\frac{1}{4} \begin{bmatrix} 2 + 2u + u^2 & 2u + 2u^2 & u^2 \\ -2u - 2u^2 & 4 - 4u^2 & -2 + 2u^2 \\ u^2 & -2u + 2u^2 & 2 - u + u^2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{22} \end{bmatrix}$$

Thus

$$\sigma_{11} = (2 + 2u + u^2)/4$$

which, as we claimed above, takes on a minimum value of $1/4$ when $u = -1$. At $u = -1$ we have

$$\frac{d}{dt}\begin{bmatrix} \sigma_{11} \\ 2\sigma_{12} \\ \sigma_{22} \end{bmatrix} = \begin{bmatrix} -4+2v & -1+v & 0 \\ 2-2v & -2 & -2+2v \\ 0 & 1-v & -2v \end{bmatrix}\begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ 2\sigma_{22} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and $\sigma_{11} = 1$; $2\sigma_{12} = 0$; $\sigma_{22} = 1$. In terms of the definitions

$$\bar{A} = \begin{bmatrix} -4 & -1 & 0 \\ 2 & -2 & -2 \\ 0 & 1 & 0 \end{bmatrix} \; ; \; B = \begin{bmatrix} 2 & 1 & 0 \\ -2 & 0 & 2 \\ 0 & 0-1 & -2 \end{bmatrix}$$

the kernel function defining the second variation is

$$w(t-\sigma) = e_1^T B e^{\bar{A}(t-\sigma)} B(e_1 + e_2)$$

8. Acknowledgements. Some of these results were presented orally at the 2009 CDC in Shanghai. Helpful comments by Sean Meyn and Jeff Shamma are acknowledged.

This material is based upon work supported by, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under contract/grant number W911NF-07-1-0376.

REFERENCES

[1] Sean Meyn, Richard L. Tweedie, Peter W. Glynn, *Markov Chains and Stochastic Stability*, Cambridge Mathematical Library, 2008.

[2] R. W. Brockett, *Optimal Control of Observable Continuous Time Markov Chains*, Proceedings, IEEE CDC, (2008) pp. 4269-4274.

[3] R. W. Brockett, *Asymptotic Properties of Markov Decision Processes*, Proceedings, IEEE CDC , (2009) pp. 3587-3591.

[4] John Baillieul, *Some Optimization Problems in Geometric Control Theory*, Ph. D. Thesis, Division of Applied Sciences, Harvard University, 1975.

[5] Fritz Colonius and Wolfgang Kliemann, *Infinite time optimal control and periodicity*, Applied Mathematics and Optimization, 20(1989), pp. 113-130.

[6] Elmer G. Gilbert, *Optimal Periodic Control: A General Theory of Necessary Conditions*, SIAM J. Control Optimization, 15:5(1970), pp. 717-746

[7] Bittanti, G. Fronza, and G. Guardabassi,*Periodic Control: A Frequency Domain Approach*, IEEE Trans. on Automatic Control, AC-18:1 (1973), pp. 33-38.

[8] J. Speyer and R. Evans, *A second variational theory for optimal periodic processes*, IEEE Transactions on Automatic Control, vol. AC 18, (1973).

[9] Keith W. Ross, *Randomized and Past-Dependent Policies for Markov Decision Processes with Multiple Constraints*, Operations Research, 37:3(1989), pp. 474-477

[10] R. W. Brockett, *Control of Stochastic Ensembles*, Åström Symposium on Control, (B. Wittenmark, A. Rantzer, eds) Studentlitteratur, Lund Sweden, 1999, pp. 199-216.